CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 113, No. 2 February 2003

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 113, NO. 2, FEBRUARY 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 113, NO. 2, FEBRUARY 2003

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 113, NO. 2, FEBRUARY 2003

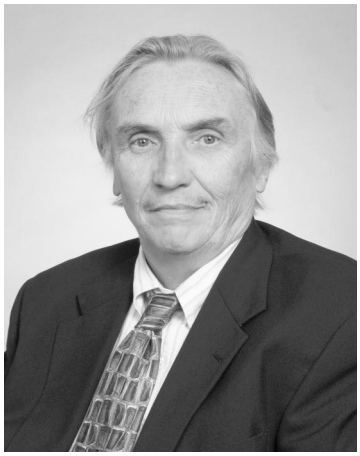
CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

Editor's Note: *Readers of this Journal are encouraged to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news items and notices are 2 months prior to publication.*

New Fellows of the Acoustical Society of America



Richard H. Campbell—For contributions to measurement techniques for listening rooms.

Laurel H. Carney—For contributions to an integrated understanding of the physiology and psychophysics of hearing.

James V. Candy receives IEEE Award

James Candy, Chief Scientist for Engineering at the Lawrence Livermore National Laboratory and an Adjunct Professor at the University of California, Santa Barbara, received the Institute of Electrical and Electronics Engineers (IEEE) Distinguished Technical Achievement Award at the IEEE OCEANS '02 Conference in Biloxi, MS with the following citation ''For the Development of Model-Based Signal Processing in Ocean Acoustics.'' He received a plaque and engraved gold watch.

This award is open to all IEEE members and is awarded to honor outstanding technical contributions to oceanic engineering in either the fundamental or applied areas. The award is for either a single major invention or scientific contribution or for a distinguished series of contributions over a long period of time.

James Candy is the Chair of the ASA Technical Committee on Signal Processing in Acoustics.

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.

2003

13–15 March	American Auditory Society Annual meeting, Scottsdale, AZ [American Auditory Society, 352 Sundial Ridge Cir., Dammeron Valley, UT 84783; Tel.: 435-574-0062; Fax: 435-574-0063; E-mail: amaudsoc@aol.com; WWW: www.amauditorysoc.org].
28 April–2 May	145th Meeting of the Acoustical Society of America, Nashville, TN [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].
5–8 May	SAE Noise & Vibration Conference & Exhibition, Traverse City, MI [P. Kreh, SAE International, 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084; Fax: 724-776-1830; WWW: http://www.sae.org].
23–25 June	NOISE-CON 2003, Cleveland, OH [INCE Business Office, lowa State Univ., 212 Marston Hall, Ames, IA 50011-2153; Fax: 515-294-3528; E-mail: ibo@ince.org].
5–8 Oct.	IEEE International Ultrasonics Symposium, Honolulu, HI [W. D. O'Brien, Jr., Bioacoustics Research Lab., Univ. of Illinois, Urbana, IL 61801-2991; Fax: 217-244-0105; WWW: www.ieee-uffc.org].
10–14 Nov.	146th Meeting of the Acoustical Society of America, Austin, TX [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].

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
3–7 Aug.	8th International Conference of Music Perception and Cognition, Evanston, IL [School of Music, Northwestern Univ., Evanston, IL 60201; WWW: www.icmpc.org/conferences.html].

15–19 Nov. 148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

ACOUSTICAL NEWS—INTERNATIONAL

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

16th International Symposium on Nonlinear Acoustics, Moscow, 2002

The 16th International Symposium on Nonlinear Acoustics was convened in Moscow, Russia, August 19–23, 2002. Just like ISNA 6 in 1975, the venue for ISNA 16 was the Department of Physics at Moscow State University. MSU is well known for its research in nonlinear acoustics, due initially in large part to the pioneering work in this field by Professor Vladmir Krasilnikov and Academician Rem Khokhlov during the 1950s and 1960s. ISNA 16 was organized by the following generation in this school of nonlinear acoustics. Professor Oleg Rudenko was chair of the symposium, which was coordinated jointly by MSU, the Acoustical Institute, and the General Physics Institute of the Russian Academy of Sciences. By any measure, the scientific content and overall organization of the symposium were outstanding.

Participants resided in the main building of MSU and at hotels in downtown Moscow. The registration room for the symposium, coordinated by Professor Valeriy Andreev, provided Internet access and considerable assistance in general throughout the symposium.

The local organizing committee reported that there were 300 registered participants, 130 of whom were from outside Russia. The program listed 214 oral presentations, delivered over four days in three parallel sessions, and there were 117 posters. 52 of the speakers were invited, representing 11 countries. Overall there appeared to be few no-shows. There were also two satellite conferences. One was on modern group analysis, organized as a special session of ISNA. The other, held at the Russian State University of Oil and Gas, was on elastic wave effects on fluids in porous media.

In the opening ceremony, the participants were treated by Oleg Rudenko to a pictorial overview of the previous ISNA held at MSU, including a retrospective of the rich tradition of nonlinear acoustics in Moscow. Professors Oleg Sapozhnikov and Vera Khokhlova, organizers of the technical and social events, described these activities for the coming week. A special moment during the opening ceremony was the award of an honorary professorship in the MSU Department of Physics to Larry Crum of the USA.

The topical organization of the symposium consisted of 11 areas: general theory—mathematics and numerical methods; solid state physics; structures and nondestructive testing; fluids, multi-phase media, and cavitation; medicine and biology; atmospheric, oceanic and seismic acoustics; flows, instabilities, turbulence, and thermoacoustics; noise and chaos; acoustics and optics; time reversal phenomena; and devices and industrial applications. An effort was thus made to represent not only the traditional areas of nonlinear acoustics, but also to include emerging subjects of interest. The proceedings of the symposium, containing all contributed and invited papers (4 and 8 pages in length, respectively), will be available in spring 2003.

The venue for this ISNA was particularly conducive to scientific interactions outside the formal presentations. Set atop hills overlooking the city and the Moscow River, MSU's campus offered a distinctive academic environment for engaging in discussions. Social events included a lavish banquet at the university (complete with vodka and caviar, of course), an evening boat tour with dinner, drinks and dancing on the Moscow River, and a guided tour of the Diamond Treasury in the Kremlin. An assortment of images of symposium participants in both scientific and social settings may be viewed in the extensive photo gallery on the ISNA 16 website at htpp://acs366b.phys.msu.su/isna16/.

In the closing ceremony it was announced that the next ISNA will be held in 2005 at Pennsylvania State University, chaired by Professor Anthony Atchley.

MARK F. HAMILTON
Department of Mechanical Engineering
The University of Texas at Austin
Austin, Texas 78712-1063

International Meetings Calendar

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

March 2003

17–20 **German Acoustical Society Meeting (DAGA2003)**, Aachen, Germany. (Fax: +49 441 798 3698; e-mail: dega@aku-physik.uni-oldenburg.de)

18–20 **Spring Meeting of the Acoustical Society of Japan**, Tokyo, Japan. (Fax: +81 3 5256 1022; Web: www.soc.nii.ac.jp/asj/index-e.html)

24–26 **27th International Acoustical Imaging Symposium**, Saarbrücken, Germany. (Fax: +49 6819032 5903; Web: www.izfp.fhg.de)

April 2003

6–10 **IEEE International Conference on Acoustics, Speech, and Signal Processing**, Hong Kong, Hong Kong. (Web: www.en.polyu.edu.hk/%7Ecassp03)

7–9 **WESPAC8**, Melbourne, Australia. (Web: www.wespac8.com)

May 2003

19–21 **5th European Conference on Noise Control** (Euronoise 2003), Naples, Italy. (Fax: +39 81 239 0364; Web: www.euronoise2003.it)

22–23 ***2nd International Styrian Noise, Vibration & Harshness Congress**, Graz, Austria. (Gesellschaft für Akustikforschung, ACC Akustikkompetenzzentrum, Inffeldgasse 25, 8010 Graz, Austria; Fax: +43 316 873 4002; Web: www.acgraz. com)

June 2003

8–13 **XVIII International Evoked Response Audiometry Study Group Symposium**, Puerto de la Cruz, Tenerife, Spain. (Web: www.ierasg-2003.org)

16–18 **Acoustics 2003—Modeling & Experimental Measurements**, Cadiz, Spain. (Fax: +44 238 029 2853; Web: www.wessex.ac.uk/conference/2003/acoustics/index.html)

29–3 **8th Conference on Noise as a Public Health Problem**, Amsterdam-Rotterdam, The Netherlands. (Fax: +31 24 360 1159; e-mail: ofice.nw@prompt.nl)

30–3 ***Ultrasonics International (UI'03)**, Granada, Spain. (T. Collier, UI'03 Secretariat, 7 Gibbs Road, Banbury OX16 3HJ, UK; Fax: +44 1295 253 334; Web: www.ccmr.cornell.edu/~ui03/ or www.ui03.com)

July 2003

7–11 **10th International Congress on Sound and Vibration**, Stockholm, Sweden. (Fax: +46 88 661 9125; Web: www.congex.com/icsv10)

14–16 **8th International Conference on Recent Advances in Structural Dynamics**, Southampton, UK. (Web: www.isvr.soton.ac.uk/sd2003)

August 2003

6–9 **Stockholm Music Acoustics Conference 2003 (SMAC03)**, Stockholm, Sweden. (Web: www.speech.kth.se/music/smac03)

25–27 **Inter-Noise 2003**, Jeju Island, Korea. (Fax: +82 42 869 8220; Web: www.icjeju.co.kr)

 0001-4966/2003/113(2)/671/2/$19.00

25–29 ***XIII Session of the Russian Acoustical Society**, Moscow, Russia. (Fax: +7 095 126 0100; Web: www.akin.ru)

September 2003

1–4 **Eurospeech 2003**, Geneva, Switzerland. (Web: www.symporg.ch/eurospeech2003)

7–10 **World Congress on Ultrasonics**, Paris, France. (Fax: +33 1 46 33 56 73; Web: www.sfa.asso.fr/wcu2003)

16–19 **Autumn Meeting of the Acoustical Society of Japan**, Nagoya, Japan. (Fax: +81 3 5256 1022; Web: www.soc.nii.ac.jp/asj/index-e.html)

18–19 ***Surface Acoustics 2003**, Salford University, Manchester, UK. (Web: www.ioa.org.uk/salford2003)

23–25 **2nd International Symposium on Fan Noise**, Senlis, France. (Fax: +33 4 72 44 49 99; Web: www.fannoise2003.org)

October 2003

15–17 **34th Spanish Congress on Acoustics**, Bilbao, Spain. (Fax: +34 91 411 7651; Web: www.ia.csic.es/sea/index.html)

15–17 ***Acoustics Week in Canada**, Edmonton, Alberta, Canada. (Fax: +1 780 414 6376; Web: caa-aca.ca/edmonton-2003.html)

December 2003

10–12 **3rd International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications**, Firenze, Italy. (Fax: +39 55 479 6767; Web: www.maveba.org)

March 2004

17–19 **Spring Meeting of the Acoustical Society of Japan**, Atsugi, Japan. (Fax: +81 3 5256 1022; Web: www.soc.nii.ac.jp/asj/index-e.html)

22–26 **Joint Congress of the French and German Acoustical Societies (SFA-DEGA)**, Strasbourg, France. (Fax: +49 441 798 3698; e-mail: sfa4@wanadoo.fr)

31–3 ***International Symposium on Musical Acoustics (ISMA2004)**, Nara, Japan. (Fax: +81 774 95 2647; Web: www2.crl.go.jp/jt/a132/isma2004)

April 2004

5–9 **18th International Congress on Acoustics (ICA2004)**, Kyoto, Japan. (Web: www.ica2004.or.jp)

11–13 ***International Symposium on Room Acoustics (ICA2004 Satellite Meeting)**, Hyogo, Japan. (Fax: +81 78 803 6043; Web: rad04.iis.u-tokyo.ac.jp)

June 2004

8–10 ***Joint Baltic–Nordic Acoustical Meeting**, Mariehamn, Åland, Finland. (Acoustical Society of Finland, Helsinki University of Technology, Laboratory of Acoustics and Signal Processing, P. O. Box 3000, 0215 TKK, Finland; Fax: +358 09 460 224; e-mail: asf@acoustics.hut.fi)

July 2004

5–8 ***7th European Conference on Underwater Acoustics (ECUA 2004)**, Delft, The Netherlands. (D. Middendorp, D'Launch Communications, Forellendaal 141, 2553 JE The Hague, The Netherlands; Fax: +31 70 322 9901; Web: www.ecua2004.tno.nl)

11–16 ***12th International Symposium on Acoustic Remote Sensing (ISARS)**, Cambridge, UK. (S. Bradley, School of Acoustics and Electronic Engineering, Brindley Building, Room 301, University of Salford, Salford M5 4WT, UK; Fax: +44 161 295 3815; Web: www.isars.org.uk)

August 2004

23–27 **2004 IEEE International Ultrasonics, Ferroelectrics, and Frequency Control 50th Anniversary Conference**, Montreal, Canada. (Fax: +1 978 927 4099; Web: www.ieee-uffc.org/index2-asp)

24–27 **Inter-noise 2004**, Prague, Czech Republic. (Fax: +1 765 494 0787; Web: www.i-ince.org)

September 2004

13–17 **4th Iberoamerican Congress on Acoustics, 4th Iberian Congress on Acoustics, 35th Spanish Congress on Acoustics**, Guimarães, Portugal. (Fax: +351 21 844 3028; e-mail: dsilva@lnec.pt)

August 2005

28–2 ***Forum Acusticum Budapest 2005**, Budapest, Hungary. (e-mail: sea@fresno.csic.es)

News from India

The Acoustical Foundation of India awarded its first Silver Medal in Noise to Professor B. V. Aswathanaryana Rao for his contribution to noise reduction of products through design. Professor Rao had been active in research and teaching at many institutions in Europe, USA, and India before he became Pro-Chancellor at Vellore Institute of Technology in 2001.

Professor B. V. A. Rao will be taking over as President of the Acoustical Society of India at the Acoustic Conference in Aligarth, India, in October 2002. He will be President for a period of two years.

New Web Links

The following organizations have a new Web address:
German Acoustical Society: http://www.dega-akustik.de
International Commission for Acoustics: http://www.icacommission.org

Congratulations!

This year Russian acousticians celebrate the 250th anniversary of Lomonosov Moscow State University and the 50th anniversary of Moscow's N. N. Adreyev Acoustics Institute. The Russian Acoustical Society has devoted its August 2003 meeting to these two events. Congratulations!

OBITUARIES

Richard Henry Bolt • 1911–2002

Richard Henry Bolt, a Fellow, a former President, and a Gold Medal awardee of the Acoustical Society, died on January 13, 2002, at the age of 90. He was a delightful companion to his colleagues and was noted for his incisive intellect. He had a broad range of interests, an exceptional ability to communicate, and was effective and dedicated to helping colleagues and students. His leadership and contributions to the field of acoustics, to the government, and, in particular, to the Acoustical Society of America, were extensive and highly significant.

Richard Bolt was born on 22 April 1911 in Peking, China to medical missionary parents, Richard Arthur Bolt and Beatrice French Bolt. The family returned to the United States in 1916 and Richard completed his schooling at Berkeley High School. Bolt entered the University of California at Berkeley in 1928, initially expecting that he would major in either music or graphic design, but eventually choosing to major in Architecture, graduating with a Bachelor of Arts degree in 1933. His only contact with physics up to that time was a low-level course for architects and pre-med students. Apparently, he developed an interest in acoustics by reading architectural journals—this would have been a natural interest for him because acoustics is a subject that has obvious ties to both music and to the design professions.

He married his wife, Katherine Mary Smith, immediately after graduation and they honeymooned in Europe. Somehow, he was introduced to Professor Erwin Meyer and Professor Hermann Biehle, both of whom were teaching acoustics in Berlin. Dick decided that he wanted to learn acoustics from those masters, even though at the time he did not know German. However, he had an amazing talent of mastering complex subjects by intensely applying himself to study. In 6 weeks' time, he mastered enough German to enroll in that fall's classes: Meyer's at the Heinrich Hertz Institute and Biehle's in his own institute. His wife Kay made it financially possible. She wrote a play for Berlin shortwave radio, acted in it, and earned enough to continue the honeymoon and finance Dick's education during the next 10 months.

Braver still, after returning to Berkeley in 1934, Dick disregarded the doubts of the physicists in the university and enrolled in Physics A and B in summer school and qualified for the graduate physics program. But, there were financial constraints. The 1933–36 period was in the depth of our country's greatest depression, and his parents could not help. Kay returned to her teaching career and taught English and dramatics in a junior high school for the next 3 years. For one term, Dick worked from 4 p.m. to midnight at a pharmaceutical company. Scholarships filled out the difference. He passed the qualifying examinations for entry into the school's Ph.D. program in 1937. From then to 1939, because of his outstanding grade record, the Bolts' finances were assisted by academic fellowships.

At that time, Berkeley had no research facilities or faculty in acoustics, so Bolt arranged to do his doctoral research under Professor Vern O. Knudsen, the Acoustical Society's third president, at the University of California's Los Angeles campus. In June 1939, he received the Ph.D. from the University of California in Berkeley–his thesis delivery coinciding with the day that their first child was born. His next academic year was spent at the Massachusetts Institute of Technology (MIT) doing research and publishing papers jointly with Philip M. Morse, Albert Clogston, and Herman Feshbach on several aspects of sound in various-shaped rooms, financed by a National Research Council postdoctoral fellowship in Physics.

In 1940, after a few months on the faculty at the University of Illinois, the impending war beckoned and he returned to MIT where, for 2 years, he directed MIT's Underwater Sound Laboratory. In 1943, Bolt was named Scientific Liaison Officer in Subsurface Warfare to the Office of Scientific Research and Development in London.

In 1945, at the close of World War II, he was appointed Director of a newly conceived Acoustics Laboratory at MIT, with faculty supervisors from the fields of physics, electrical engineering, architecture, mechanical and aeronautical engineering, psychology, and the arts. In February 1947, he and Philip Morse enticed me from my faculty position at Harvard to serve as Associate Professor in Communication Engineering at MIT and Technical Director of the Acoustics Laboratory. With our offices located across the corridor from each other, Dick and I built the USA's largest acoustics laboratory. At its peak we employed more than 80 persons and, in the next 12 years, 108 graduate theses were completed.

During the period immediately following World War II, one of Bolt's activities was the teaching of courses in architectural acoustics, with the assistance of Robert Newman. William Cavanaugh, who was then an undergraduate in the MIT School of Architecture and City Planning and who was later to become a long-term colleague of Bolt and myself, took both the introductory course and a special course with the title ''Advanced Seminar in Architectural Acoustics,'' and recalls that the first course was definitely not a ''watered down'' version for architects, but rather was largely grounded in physical principles.

Before my arrival at MIT, the university had received a request for a proposal that would cover consultation on the acoustics of the United Nations Permanent Headquarters buildings in New York. Dick was asked by the President to respond personally, and in the fall of 1948 he received a contract to act as consultant to the renowned New York architectural firm, Harrison and Abramovitz, coordinating architects for a consortium of internationally acclaimed architects designing this prestigious project. My own independent consulting practice was already well developed when the drawings for the UN Complex, 16 feet long and piled 8 inches high, arrived in Dick's office across from mine. In awe at the extent, complexity, and tight time schedule for the project, Dick asked me to join him in forming a consulting partnership, with MIT's permission, to handle our combined projects. Thus, in November 1948, the partnership of Bolt and Beranek was formed, with two part-time graduate students hired to help us. We began by renting two rooms from MIT. A year later, Robert B. Newman, now on the faculty of the MIT School of Architecture and Planning, joined as a partner and, in 1951, Bolt Beranek and Newman (BBN) was incorporated with Bolt as Chairman of the Board, with myself as President and Chief Executive Officer, with Newman and Samuel Labate as Vice Presidents and with Jordan Baruch as Treasurer. By 1950 we had located our offices in Harvard Square, first sharing space in a building housing the architectural offices of MIT Professor Carl Koch and later, as BBN grew, moving to 16 Eliot Street. In 1956 we moved to 50 Moulton Street which, when augmented by a two-story building designed largely by Bolt, served the needs of BBN 's Cambridge headquarters offices and laboratories for the next several decades.

In 1957, The National Institutes of Health appointed Bolt as principal consultant in Biophysics and to work with a new study section in that field. Dick was now at his best, building a scientific approach to a new field and answering questions amenable to group assault. By the summer of 1958 he had organized a resoundingly successful international conference that took place in Boulder, Colorado, to explore further directions for biophysical science. Attended by 117 people, the conference stimulated a step function of activity in the biological sciences. Ninety of those present received collaborative research contracts and no fewer than six of the participants later reaped a Nobel Prize.

This conference brought Richard Bolt into national visibility. In 1960 he was named Associate Director of the National Science Foundation, where he served in Washington, DC for 3 years. Next, he spent a year as a Fellow of the Center for Advanced Study in Behavioral Sciences at Stanford University. On his return to MIT, he served for several years as a Lecturer in the Department of Political Science. In the next years, Dick aided agency after agency and committee on committee in organizing their deliberations and overseeing their published proceedings.

In 1973, Bolt was made chairman of a committee of six experts to investigate the 18-minute gap on a tape made in President Nixon's office 3 days after the Watergate break-in. It was purported that this gap contained a mention of the Watergate affair in a conversation between the President and H. R. Haldeman. The committee was unable to discover who may have erased the tape, but Bolt stated that the erasure was no accident because the

erasure was started over during the gap at least five times, maybe nine. This project made him a household name in this nation.

The Acoustical Society awarded Bolt its first Biennial Award (later called the R. Bruce Lindsay Award) in 1942 and its Gold Medal in 1979. The citation for the latter read: ''For outstanding contributions to acoustics through research, teaching, and professional leadership, and for distinguished administrative and advisory service to science, engineering and government.'' He was elected to the National Academy of Engineering in 1978 for ''Contributions to acoustics and leadership in engineering enterprises and public service.'' Over the years he published over 50 papers in acoustics and coauthored, with Theodore F. Hueter, a book titled *Sonics: Techniques for the Use of Sound and Ultrasound in Engineering and Science* (Wiley, 1955 republished by the Acoustical Society of America in 2002).

During his service as a member of the Executive Council from 1944–1947, he undertook with Cyril Harris and John Steinberg to rewrite the Society's Constitution. The following year he became the first person to serve as President-Elect of the Society under one of its newly written by-laws. In 1949, he succeeded to a term as the Society's President. He also was an active founding member of the Institute of Noise Control Engineering, and an active fellow of several other prestigious professional societies including the American Physical Society, the American Institute of Physics, and the American Academy of Arts and Sciences. When the International Commission on Acoustics (ICA) was founded in 1951, he was chosen to be its first President. When the Gold Medal Award of the Acoustical Society was established in 1954, he worked as consultant on graphic design to the sculptor of the medal. In 1979, at the Fiftieth Anniversary of the Society's founding at MIT in Cambridge, he served as meeting organizer and General Chair. (He also received the Gold Medal at this meeting.)

When Robert Newman died in 1983, Dick enthusiastically joined a group of Bob's friends and colleagues to form the Robert Bradford Newman Student Award Fund to promote the teaching of architectural acoustics in schools of architecture. Dick designed the Newman Medal which has been awarded to many outstanding graduating students in the field.

Dick served as Chairman of the BBN Board of Directors from 1953 to 1976 and continued to serve on the Board until 1981. He maintained an office at BBN until well into the 1990's and acted as consultant on a number of projects. His spouse, Katherine, died in 1991 after an extended illness. They had three children: Dick Bolt, Bea Schribner, and Deborah Bolt-Zieses. There were seven grandchildren and, at the time of this writing, eight great-grandchildren.

LEO L. BERANEK

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

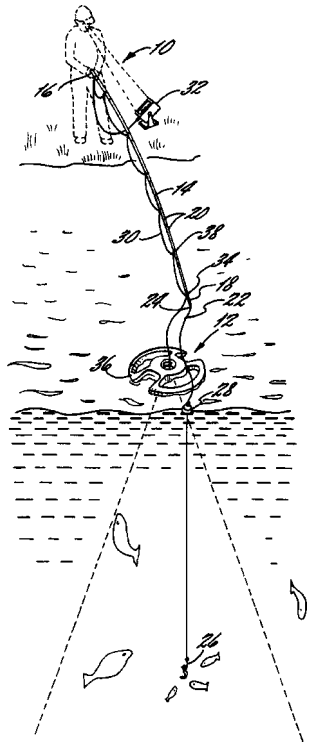
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DAVID PREVES, *Micro-Tech Hearing Instruments, 3500 Holly Lane No., Suite 10, Plymouth, Minnesota 55447*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, Univ. of Rochester, Rochester, New York 14627*

5,771,205

43.30.Sf BUOYANT TRANSDUCER ASSEMBLY FOR ASSISTING AN ANGLER

Jon David Currier and Kenton Sterling Jopling, assignors to Lowrance Electronics, Incorporated
23 June 1998 (Class 367/107); filed 18 April 1996

Hollow buoyant disk-shaped float **36** is formed from plastic by injection molding and supports a fish-locating sonar transducer (not shown). Electrical cable **30** runs along fishing pole **14** in parallel with fishing line **22** and is supported at various points along the pole with appropriate fasteners.



The cable **30** is connected to battery-powered display unit **32**. The fisherman casts both his/her line and the float assembly **36** from the shore, dock, or boat. Alternatively, the float assembly can be disconnected from the fishing pole and cast separately like a Frisbee.—WT

6,430,104

43.30.Xm SONAR SYSTEM PERFORMANCE METHOD

Michael A. Rosario, assignor to the United States of America as represented by the Secretary of the Navy
6 August 2002 (Class 367/13); filed 27 February 2001

A new metric for rating sonar systems is discussed. A performance rating factor is formulated based upon measurements of the detection range of a given target as the positions of source, receiver, and target are varied over wide ranges.—WT

5,768,216

43.30.Yj FLEXITENSIONAL TRANSDUCER HAVING A STRAIN COMPENSATOR

Hidenori Obata *et al.*, assignors to Oki Electric Industry Company, Limited
16 June 1998 (Class 367/172); filed in Japan 28 June 1995

In a class IV flextensional transducer, one end of the stack of piezoceramic actuators is rigidly fixed to the oval shell while the other end is attached to the shell via a piston in a fluid cylinder arrangement which is aligned with the stack of actuators along the major axis of the shell. This assembly relieves the changing hydrostatic force that would be applied to the stack as the shell is submerged in the water. However, the dynamic forces generated by the stack, at the frequencies at which it vibrates, are readily communicated to the shell through the piston/cylinder assembly, causing the shell to vibrate. Static precompression of the piezoceramic must be accomplished, in this case, by conventional stress rods.—WT

5,781,506

43.30.Yj METHOD AND APPARATUS FOR FREQUENCY FILTERING USING AN ELASTIC FLUID-FILLED CYLINDER

Mark S. Peloquin, assignor to the United States of America as represented by the Secretary of the Navy
14 July 1998 (Class 367/135); filed 29 May 1997

By suitable choice of the density and dilatational velocity of the fill fluid, as well as the wall thickness of the enclosing boot, the first radial

 0001-4966/2003/113(2)/679/19/$19.00

resonance of the boot of a towed array can be adjusted, thus causing it to act as a low-pass mechanical filter against sound waves in the surrounding medium.—WT

6,421,299

43.30.Yj SINGLE-TRANSMIT, DUAL-RECEIVE SONAR

David A. Betts and Louis Loving, assignors to Techsonic Industries, Incorporated
16 July 2002 (Class 367/105); filed 28 June 2000

A fish-finder/depth-finder type of sonar features a transducer that transmits with a wide beam but, in the receive mode, can listen to returns with either a wide or a narrow beam. The time duration that the system listens with the narrow beam is a function of the water depth, which was determined during a prior transmit/receive cycle.—WT

6,421,301

43.30.Yj TRANSDUCER SHIELD

William J. Scanlon, McComb, Mississippi
16 July 2002 (Class 367/173); filed 4 January 2001

A simple boxlike shield is described that surrounds and protects a fish-finder type of transducer from impact with underwater objects when that transducer is mounted, for example, on the bottom of a trolling motor housing. The shield, which may feature some basic hydrodynamic shaping, is fashioned from aluminum or other material and is open at its bottom surface so as not to interfere with the sonar transmissions and receptions.—WT

6,424,451

43.35.Sx PHASE ARRAY ACOUSTO-OPTIC TUNABLE FILTER BASED ON BIREFRINGENT DIFFRACTION

I-Cheng Chang, Sunnyvale, California
23 July 2002 (Class 359/308); filed 22 May 1999

This invention relates to the field of electronically tunable optical filters utilizing acoustooptic diffraction. An acoustooptic tunable filter (AOTF) utilizes phased array transducers for use as a dynamically reconfigurable wavelength division multiplexer. This type of AOTF has the capability of simultaneous and independent selection of multi-wavelength signals and separation into multiple output ports. Preferred embodiments of the AOTF are described that are said to provide increased resolution, narrow channel spacing, lower drive power, and reduced coherent crosstalk.—DRR

6,424,857

43.35.Sx USE OF ACOUSTO-OPTICAL AND SONOLUMINESCENT CONTRAST AGENTS

Paul Mark Henrichs and Henry Raphael Wolfe, assignors to Amersham Health AS
23 July 2002 (Class 600/431); filed in the United Kingdom 16 June 1997

This patent relates to a method of diagnostic imaging of a human or animal subject, in particular, a method in which the image consists of light generated by or characteristically affected by ultrasound irradiation of the subject. The method of generating information from an animate body entails administering to the body a physiologically tolerable material capable of absorbing, scattering, or emitting light at a wavelength in the range 300 to 1300 nm, subjecting at least a portion of the body to ultrasound irradiation, detecting light in the wavelength range 300 to 1300 nm from that portion of the body, and manipulating the detected light to generate information. A considerable number of candidate detection materials are listed, perhaps rendering the patent a bit too general in scope.—DRR

6,424,864

43.35.Wa METHOD AND APPARATUS FOR WAVE THERAPY

Masayuki Matsuura, Hamamatsu-shi, Shizuoka-ken, Japan
23 July 2002 (Class 607/3); filed in Japan 28 November 1997

This patent describes a wave therapeutic system intended for treating definitive diseases by applying at least one of a low-frequency current, an electromagnetic wave, and an acoustic wave. A low-frequency oscillator generates multiple frequencies selected according to the disease type. These frequencies are delivered from the lowest value in an ascending order at intervals of a specified time and are introduced into a human body at the same time or separately as low-frequency electric currents from a therapeutic electrode and an inactive electrode, as well as electromagnetic waves from oscillating coils and acoustic waves from an oscillator or body sonic device.—DRR

6,419,632

43.35.Yb HIGH RESOLUTION FLOW IMAGING FOR ULTRASOUND DIAGNOSIS

Eiichi Shiki and Yoshitaka Mine, assignors to Kabushiki Kaisha Toshiba
16 July 2002 (Class 600/443); filed in Japan 30 March 1999

This apparatus provides an ultrasound image of blood flow or perfusion in a high-resolution flow mode, thus enabling observation of the presence of fine blood vessels. An ultrasound pulse having a wide-band frequency characteristic is transmitted at least two times in the same direction within an object. A cross section of the object to be imaged is scanned to obtain an echo signal at each time of transmission. High-pass filtering or differential processing is performed with rows of data in the time axis direction so that signals from blood flow are extracted. Luminance and power information extracted from the signals is displayed as a high-resolution color or grayscale flow image indicative of blood flow or perfusion.—DRR

6,423,006

43.35.Yb METHOD AND APPARATUS FOR AUTOMATIC VESSEL TRACKING IN ULTRASOUND SYSTEMS

Zoran Banjanin, assignor to Siemens Medical Solutions USA, Incorporated
23 July 2002 (Class 600/453); filed 21 January 2000

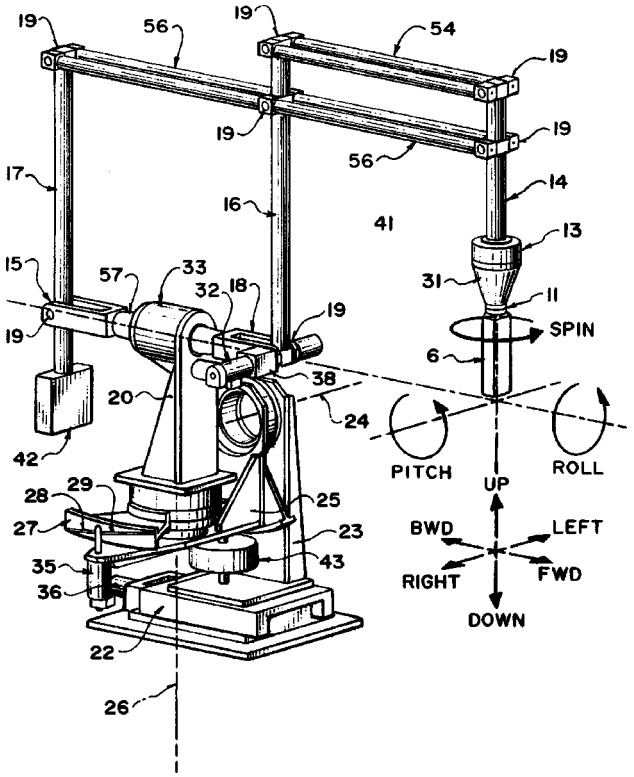
The method covered in this patent is one that automatically places a range gate over a moving blood vessel during ultrasound imaging. Doppler data received from a number of depths in the tissue are analyzed in order to calculate the average velocity at each depth. A search is performed of the average velocities in order to determine a maximum velocity. The maximum velocity is associated with a blood vessel and a range gate is placed at a depth corresponding to the maximum velocity. In a presently preferred embodiment of the invention, the average velocity is calculated by performing a first lag autocorrelation of the echo data received from each depth in response to a series of Doppler pulses.—DRR

6,425,865

43.35.Yb ROBOTICALLY ASSISTED MEDICAL ULTRASOUND

Septimiu E. Salcudean *et al.*, assignors to The University of British Columbia
30 July 2002 (Class 600/437); filed 11 June 1999

A system is presented for controlling medical ultrasound in which an ultrasound probe is positioned by a robot arm under the shared control of the ultrasound operator and a computer. The system consists of a robot arm suitable for diagnostic ultrasound, a passive or active hand-controller, and at least one computer system to coordinate the motion and forces of the robot and hand-controller as a function of operator input, sensed parameters, and



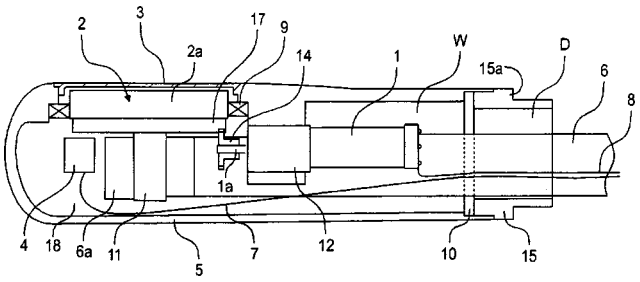
ultrasound images. The motions of the robot arm and the hand controller are based on measured positions and forces, acquired ultrasound images, and/or taught position and force trajectories. Several modes of control are presented, including the control of the transducer using ultrasound image tracking.—DRR

6,425,870

43.35.Yb METHOD AND APPARATUS FOR A MOTORIZED MULTI-PLANE TRANSDUCER TIP

Aimé Flesch, assignor to Vermon
30 July 2002 (Class 600/459); filed 11 July 2000

An ultrasonic phased-array imaging transducer device is provided which includes a flexible sealing membrane **10** disposed within a housing so as to divide the housing into a wet chamber **W** and a dry chamber **D**. The wet chamber contains a fluid and includes a phased array transducer disposed in the tip end of the housing and oriented so as to provide a sound path extending perpendicular to the longitudinal axis of the housing. A motor provides rotation of the transducer while an encoder supplies positional



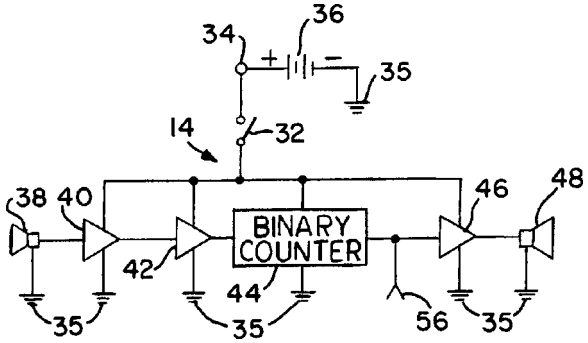
information with respect to the transducer. A flexible electronic circuit **6** is connected to the transducer and is coiled relative to the transducer so as to permit transducer rotation of more than 180 degrees. The flexible circuit extends from the wet chamber through the flexible sealing membrane to the dry chamber. A torque limitation device limits the torque transmitted to the transducer.—DRR

6,426,919

43.35.Yb PORTABLE AND HAND-HELD DEVICE FOR MAKING HUMANLY AUDIBLE SOUNDS RESPONSIVE TO THE DETECTING OF ULTRASONIC SOUNDS

William A. Gerosa, Pleasantville, New York
30 July 2002 (Class 367/132); filed 4 January 2001

Considering that handheld ultrasonic to audio converters have been used by the bat research community for many years, it is not clear what the patent offers. The circuit is certainly interesting: the amplified audio input is



used as clock input to a 7-bit counter. The output of the counter is also directly amplified and therefore no A/D or D/A conversion is done. The output is surely not as useful as straightforward downconversion.—MK

6,388,406

43.35.Zc METHOD AND SYSTEM FOR DETECTING AN OBJECT IN THE PATH OF A VEHICLE POWER WINDOW SYSTEM USING ACOUSTIC SIGNALS

Christos Kyrtsos, assignor to Maritor Light Vehicle Systems, Incorporated
14 May 2002 (Class 318/286); filed 30 July 1999

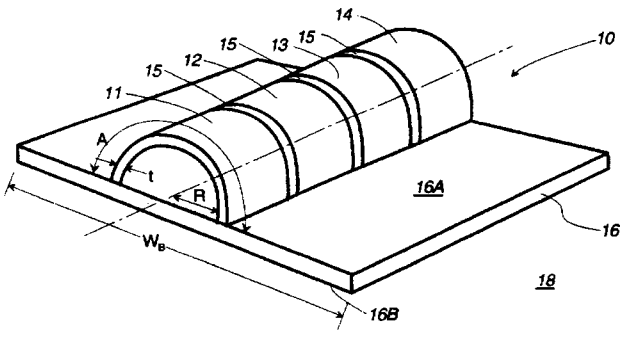
A method to detect the presence of an object in the path of a closing window consists of an acoustic sensor and a control circuit to interrupt power to the drive motor when the detected signal differs from the baseline acoustic signal. An alternative scheme includes an actuator attached to the window that generates signals in the 40-kHz range. The aforementioned sensor again detects differences from an expected baseline as an indication that an object is present.—KPS

5,781,509

43.38.Fx WIDE BEAM ARRAY WITH SHARP CUTOFF

William J. Zehner, assignor to the United States of America as represented by the Secretary of the Navy
14 July 1998 (Class 367/159); filed 28 May 1996

A coaxial array of half cylinders (or even smaller fractions of cylinders) of piezoelectric material **11**, **12**, ... is mounted on a baffle **16** which may be absorbent or reflective. The baffle dimension $\mathbf{W_B}$ is to be at least three or four times the radius **R**. It is alleged that the directivity pattern in a



plane perpendicular to the cylinder axis is characterized by a rather uniform beam in the half space where the piezoelectric cylinders exist but the response rapidly decreases with angle to a low value in the half of that plane which lies in the opposite half space from the cylinders.—WT

6,432,068

43.38.Fx HIGH OUTPUT THERAPEUTIC ULTRASOUND TRANSDUCER

Paul D. Corl *et al.*, assignors to Pharmasonics, Incorporated
13 August 2002 (Class 601/2); filed 20 March 2000

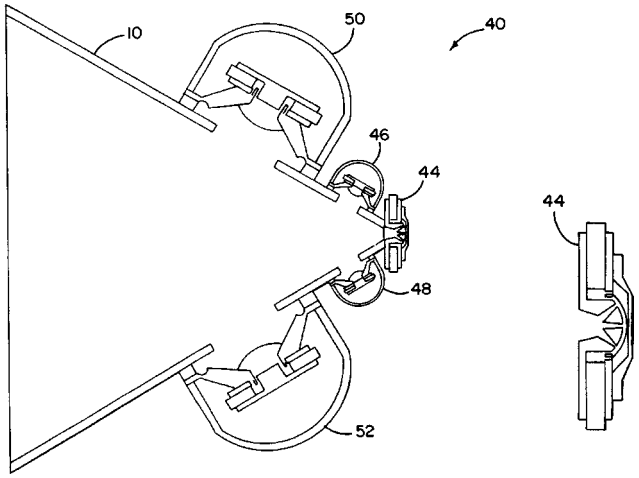
This is a therapeutic ultrasound energy delivery system that includes a probe incorporating a vibrational transducer. A restraint is deployed about the transducer in order to exert a compressive prestress on the transducer. The restraint inhibits tensile failure of the vibrational transducer at high acoustic output. The underlying principle has also been applied to prestressed concrete structural members in buildings.—DRR

6,411,718

43.38.Ja SOUND REPRODUCTION EMPLOYING UNITY SUMMATION APERTURE LOUDSPEAKERS

Thomas J. Danley and Bradford J. Skuran, assignors to Sound Physics Labs, Incorporated
25 June 2002 (Class 381/342); filed 28 April 1999

High-frequency driver **44** is coupled to a ''multiply segmented'' horn on which additional lower frequency assemblies **46**, **48**, **50**, and **52** are



mounted. In 1944 this reviewer mounted a radio loudspeaker on one wall of an acoustic phonograph horn; does that count as prior art?—GLA

6,411,719

43.38.Ja LOUDSPEAKER ASSEMBLY

Erik Möster *et al.*, assignors to Telefonaktiebolaget LM Ericsson (publ)
25 June 2002 (Class 381/345); filed in Sweden 18 May 1999

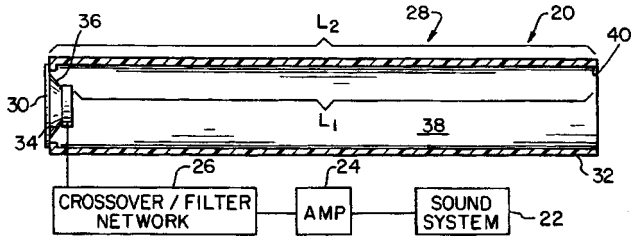
The loudspeaker of a small, hand-held telephone is normally mounted to the underside of the front cover. The patent suggests that by mounting the speaker on an intermediate panel between the front and rear covers, a number of practical advantages can be realized, including improved audio performance.—GLA

6,411,721

43.38.Ja AUDIO SPEAKER WITH HARMONIC ENCLOSURE

William E. Spindler, Fort Wayne, Indiana
25 June 2002 (Class 381/349); filed 19 December 1997

Loudspeaker **30** drives a simple, undamped pipe **38**. If desired, the pipe can be folded to fit within a rectangular enclosure. This is all well-known prior art going back to the 1950s. However, the patent teaches that the arrangement shown is novel because its effective length is one-eighth of



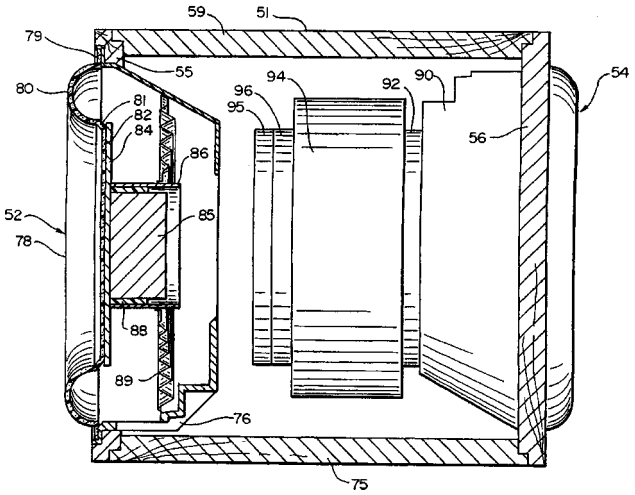
a wavelength at the lowest frequency to be reproduced, maintaining useful summation of cone and pipe outputs down to that frequency. Whether this assertion justifies a patent is immaterial because it is wrong. The invention is based on a fallacy, as would have been apparent from even one actual test.—GLA

6,411,723

43.38.Ja LOUDSPEAKERS

Christopher Colin Lock *et al.*, assignors to Slab Technology Limited
25 June 2002 (Class 381/431); filed in New Zealand 22 June 1998

During the past three years we have witnessed a flood of patents dealing with distributed-mode, panel-type loudspeakers. Most of these were assigned to the same company. Somewhat earlier, a steady stream of patents dealing with distributed-mode, panel-type loudspeakers were acquired by a different company. The patent at hand is assigned to a third company which has also been producing distributed-mode, panel-type loudspeakers for a number of years. In this case the preferred embodiment makes use of a special panel material which is said to provide ease of manufacture, good performance at low cost, and compatibility with a wide range of diaphragm shapes and sizes.—GLA



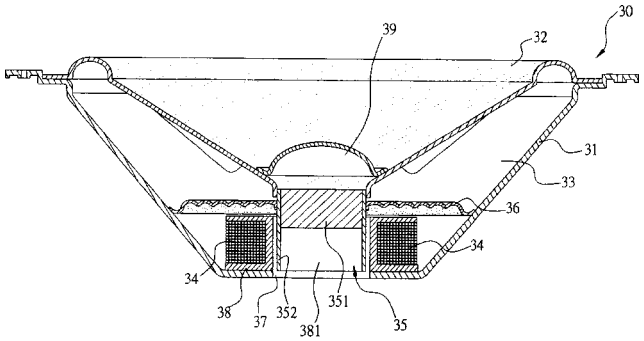
Sunfire subwoofer. It also includes mind-numbing extraneous details, such as the semantic equivalence of nails and screws in relation to cabinet joinery.—GLA

6,415,037

43.38.Ja SPEAKER AND THE MANUFACTURING METHOD THEREOF

Chao-Lang Wang, assignor to Elecinic Corporation
2 July 2002 (Class 381/418); filed 20 October 2000

Neodymium magnets used in contemporary microphones and loudspeakers are lightweight and small. More than one loudspeaker manufacturer has toyed with the idea of a fixed coil, moving magnet design. The patent describes such a device which includes moving magnet **351**, fixed coil **34**, resonant space **33**, and through-hole **35**.—GLA



6,418,231

43.38.Ja HIGH BACK EMF, HIGH PRESSURE SUBWOOFER HAVING SMALL VOLUME CABINET, LOW FREQUENCY CUTOFF AND PRESSURE RESISTANT SURROUND

Robert W. Carver, Snohomish, Washington
9 July 2002 (Class 381/395); filed 9 August 1999

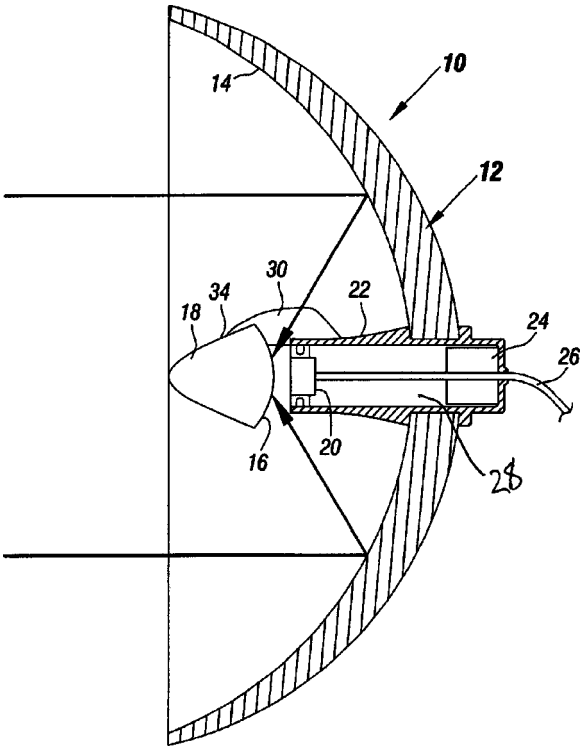
This patent is the latest in a series of continuations going back to 1997. The patent document is 25 pages long and contains 149 claims. It provides a lot of interesting information about the inventor's commercially successful

6,408,080

43.38.Kb BOUNDARY LAYER MICROPHONE

David H. Koizumi, assignor to Intel Corporation
18 June 2002 (Class 381/160); filed 29 November 1999

Microphone **20** is located at the focus of concave reflector **14** to form a familiar directional sound pickup device. In this case, a second reflecting surface **16** provides a boundary layer to increase the sensitivity of the microphone. According to the patent text, the main reflector might have a diameter of 8 to 10 in. for speech pickup. If the illustration is drawn to scale, then reflector **16** is less than 2 in. in diameter, implying that its usefulness will be limited to frequencies above 5 kHz or so.—GLA



6,416,483

43.38.Kb SENSOR AND METHOD FOR DETECTING VERY LOW FREQUENCY ACOUSTIC SIGNALS

Michael E. Halleck *et al.*, assignors to iLife Systems, Incorporated
9 July 2002 (Class 600/561); filed 24 March 2000

Information about a patient's heartbeat and respiration can be obtained from acoustic signals, eliminating the need for physically attached sensors. The inventors state that unwanted noise can be mostly avoided by limiting pickup to frequencies between 0.1 and 30 Hz. The preferred embodiment of a suitable sensor consists of a pressure microphone mounted inside a sealed chamber whose walls can flex in response to very-low-frequency acoustic energy.—GLA

6,418,229

43.38.Kb DIRECTIONAL MICROPHONE, IN PARTICULAR HAVING SYMMETRICAL DIRECTIVITY

Raimund Staat, assignor to Sennheiser electronic GmbH & Company KG
9 July 2002 (Class 381/357); filed in Germany 17 February 1997

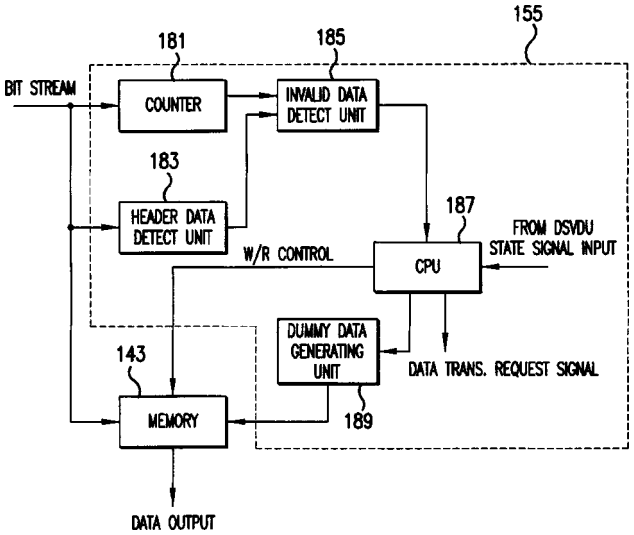
A small, highly directional microphone can be made by adding an interference tube to the front of a cardioid microphone capsule. The patent describes a number of improvements to this basic geometry. A symmetrical pickup pattern is obtained by locating sound inlets in a controlled, nonsymmetrical fashion.—GLA

6,408,040

43.38.Md METHOD AND APPARATUS FOR COMPENSATING REPRODUCED AUDIO SIGNALS OF AN OPTICAL DISK

Jae Ryong Cho, assignor to LG Electronics Incorporated
18 June 2002 (Class 375/377); filed in the Republic of Korea 4 February 1997

If a DVD audio block is too short, it must be extended. If it's too long, it must be truncated. Likewise, noisy blocks should be eliminated. But while



the inventor is able to identify the problem, the patent doesn't disclose any of the methods needed to avoid perceptual discontinuities.—MK

6,427,370

43.38.Md PICTURE FRAME WITH SOUND AND MOTION PRODUCING MEANS

Nathan Smith, Los Angeles, California
6 August 2002 (Class 40/717); filed 14 September 2001

This patent duplicates the work of United States Patent 6,393,401 [reviewed in J. Acoust. Soc. Am. **112**(6), 2515 (2002)]. The only new wrinkle is the addition of a mechanical winding mechanism.—MK

6,408,081

43.38.Si BONE CONDUCTION VOICE TRANSMISSION APPARATUS AND SYSTEM

Peter V. Boesen, Des Moines, Iowa
18 June 2002 (Class 381/312); filed 5 June 2000

An air conduction sensor (microphone) and a bone conduction sensor are both contained in an in-the-ear assembly. Speech processing circuitry compares signals from the two transducers. ''In comparing the sampled output, the speech processor is able to filter noise and select a pure voice signal for transmission.'' Exactly how this is accomplished is not explained.—GLA

6,415,034

43.38.Si EARPHONE UNIT AND A TERMINAL DEVICE

Jarmo Hietanen, assignor to Nokia Mobile Phones Limited
2 July 2002 (Class 381/71.6); filed in Finland 13 August 1996

An earphone assembly incorporates both a microphone and a sound reproducing transducer. A separate error microphone provides noise canceling capabilities. ''In order to improve the quality of speech and prevent problems caused by double-talk, signals are processed digitally utilizing, for example, band limitation and prediction of missing bands.''—GLA

6,418,226

43.38.Vk METHOD OF POSITIONING SOUND IMAGE WITH DISTANCE ADJUSTMENT

Masahiro Mukojima, assignor to Yamaha Corporation
9 July 2002 (Class 381/17); filed in Japan 12 December 1996

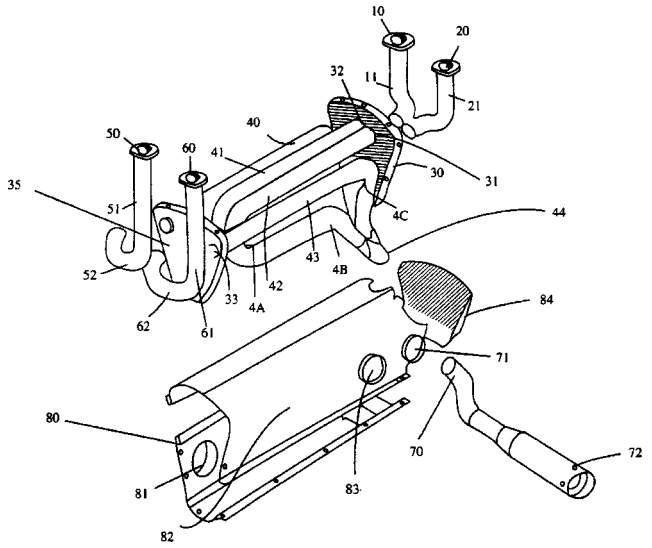
Virtual reality systems can produce phantom sound images at almost any location by making use of head-related transfer functions. However, it is obviously not practical to store enough such functions to match every location that might be utilized. Because of this limitation, some variations, particularly near versus far sources, can be miscalculated. This invention computes more accurate HRTFs by making use of more than one reference point in relation to a given image.—GLA

6,374,599

43.50.Gf COMPACT TUNED EXHAUST SYSTEM FOR AIRCRAFT WITH RECIPROCATING ENGINES

Robin G. Thomas, assignor to Power Flow Systems, Incorporated
23 April 2002 (Class 60/312); filed 21 July 2000

A muffler intended for piston-engined aircraft is described. The four-cylinder version shown is designed to fit within the small available space and has the key characteristic that all four exhaust tubes are of the same



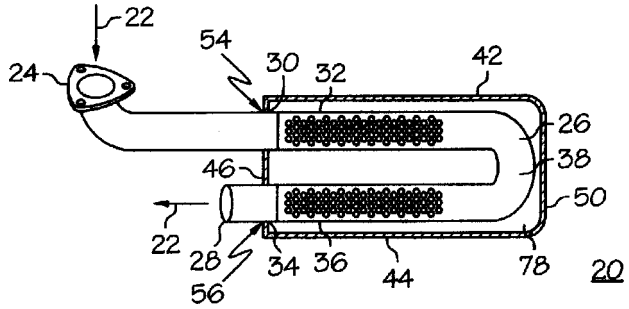
length. For the example cited, a Lycoming engine in a Cessna 172, the 40-in. tuned exhaust tubes result in a significant improvement in horsepower and fuel efficiency. An arrangement for a six-cylinder engine is also described.—KPS

6,382,347

43.50.Gf EXHAUST MUFFLER FOR AN INTERNAL COMBUSTION ENGINE

Brian H. Gerber, assignor to GHL Motorsports, L.L.C
7 May 2002 (Class 181/227); filed 8 May 2001

A high performance muffler for a Porsche designed to produce a ''deep throaty sound'' consists of perforated stainless steel pipes **32**, **36** surrounded by sound-absorbing material (steel wool), not shown. This single-input



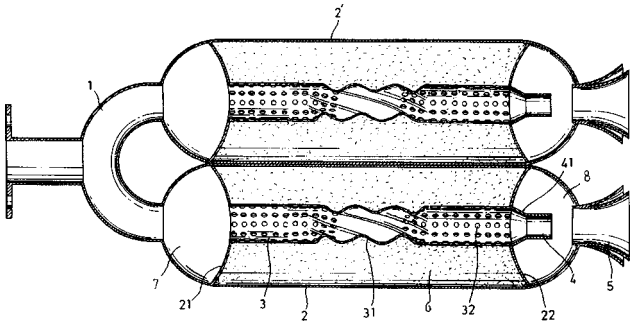
single-output design is modified to accommodate dual-input, single-output and single-input, dual-output schemes with similar design characteristics.—KPS

6,382,348

43.50.Gf TWIN MUFFLER

Shun-Lai Chen and Yuanlin Chen, Changhwa Hsien, Taiwan, Province of China
7 May 2002 (Class 181/239); filed 9 February 2001

Exhaust gases flow through **1** into two parallel mufflers consisting of perforated pipes surrounded by sound-absorbing material. Chambers **7** and **8**



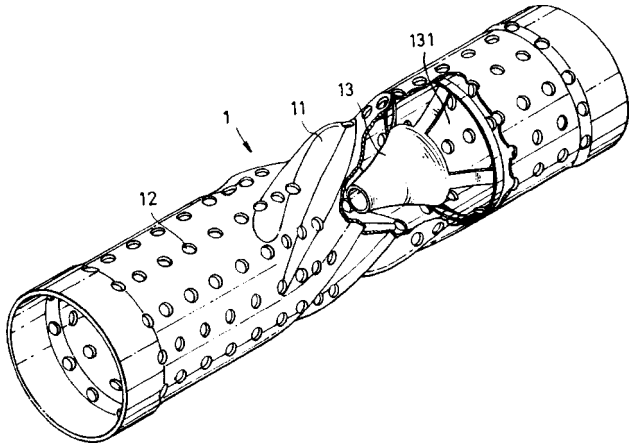
are placed at either end. The center section of each pipe is twisted, causing the flow to spiral and accelerate, thus pulling in gases through the perforations.—KPS

6,385,967

43.50.Gf EXHAUST PIPE FOR MOTOR VEHICLE MUFFLER

Shun-Lai Chen, Changhwa Hsien, Taiwan, Province of China
14 May 2002 (Class 60/312); filed 31 May 2000

Exhaust gases flow through conical section **13** and through spiral section **11** that serves to accelerate the flow and pull exhaust gases in through holes **12**. This arrangement is aimed at preventing ambient air being sucked



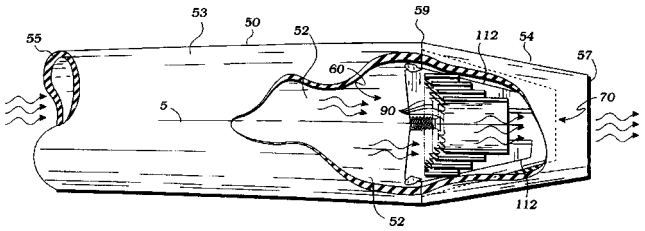
into the muffler that causes vibration when the accelerator pedal is suddenly released.—KPS

6,415,887

43.50.Gf REFRACTIVE WAVE MUFFLER

Clay Moran and Rodney Lee Asher, assignors to CR Patents, Incorporated
9 July 2002 (Class 181/264); filed 14 November 2000

This muffler consists of multiple concentric cylinders, graduated in length. The inlets of each cylinder are of a sawtooth design. This arrange-



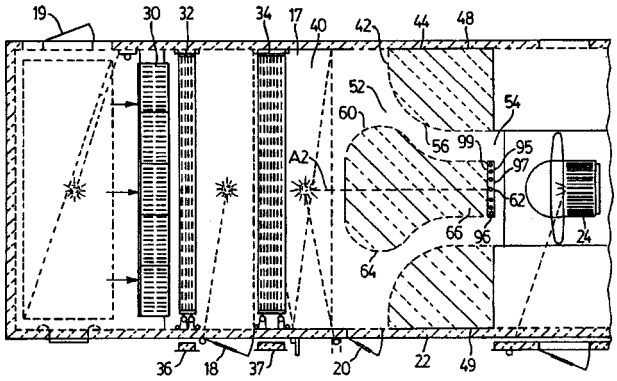
ment supposedly reduces reflected acoustic energy, thus minimizing backpressure and providing sound attenuation.—KPS

6,419,576

43.50.Gf SOUND ATTENUATING INLET SILENCER FOR AIR SUPPLYING FAN

Ming Hui Han, assignor to Air Handling Engineering Limited
16 July 2002 (Class 454/338); filed 22 March 2001

This silencer, intended for a building or other large structure, has a housing with interior walls that define an annular airflow passageway **52** extending between an air inlet located at one end and a circular air outlet adapted for connection to a fan unit. The outlet defines a central axis extending through the center of the outlet and perpendicular to a plane in which the outlet lies. A central airflow-defining member **60** has a central axis **A2** substantially coaxial with the outlet central axis. This airflow defining



member is substantially circular in transverse cross-section along its length and has a relatively wide bulblike end section at its outer end and a narrow end section at the inner end **66**. In a preferred embodiment, the central airflow-defining member projects outwardly from the air inlet. Also, the airflow passageway is tapered and the airflow-defining member is curved as seen in an axial plane extending through the second central axis so that the airflow speed increases smoothly to maximum velocity at the air outlet when the fan unit is operating.—DRR

6,422,083

43.50.Gf TUNED ENERGY REDISTRIBUTION SYSTEM FOR VIBRATING SYSTEMS

Gregg K. Hobbs, Westminster, Colorado
23 July 2002 (Class 73/663); filed 24 March 2000

This system is used to precisely control the amplitude of vibrations in a vibrating system. The vibrating system typically has a resonant frequency or other vibration frequency mode that is either undesirable or of excessive amplitude. The present system functions to redistribute the vibrational energy from these undesirable frequencies into other selected frequencies, such as by spreading them out over a wide band. The tuned energy redistribution consists of vibration input elements and/or vibration shaping elements that collectively function to enable the user to program the frequency and magnitude of the vibrations that are produced by the vibrating system. This energy redistribution is typically accomplished by providing tuned absorbers, consisting of a vibrating mass and a vibration stop, which shape the frequency response of the vibrating system by responding to frequencies mostly near their natural frequency of vibration and then, when the response displacement is sufficient, provide a vibration input by impacting the element in the vibrating system to which they are connected via the vibration stop, causing broad band vibration to be generated by the series of impacts.—DRR

6,416,016

43.50.Ki ACTUATOR FOR AN ACTIVE TRANSMISSION MOUNT ISOLATION SYSTEM

William A. Welsh, assignor to Sikorsky Aircraft Corporation
9 July 2002 (Class 244/54); filed 29 November 2000

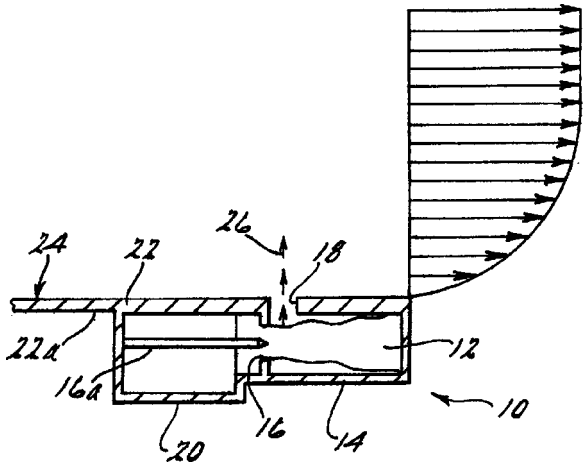
An actuator is described for use in an active control system applicable to helicopter interior noise and vibration. This piston actuator includes a housing which attaches to the aircraft fuselage and controls transmission of vibration from the feet of the gearbox. A clever arrangement of bearings allows the transmission of axial, moment, and shear loads. This arrangement is said to be superior to others which incorporate a strut in parallel with the cylinder and which allows transmission of unwanted high-frequency vibration.—KPS

6,375,118

43.50.Nm HIGH FREQUENCY EXCITATION APPARATUS AND METHOD FOR REDUCING JET AND CAVITY NOISE

Valdis Kibens and Ganesh Raman, assignors to The Boeing Company
23 April 2002 (Class 244/53 R); filed 30 August 2000

Reduction of jet noise for high-performance military engines is achieved by a circumferential distribution of resonant devices that energize the shear layer in the exhaust flow. Each resonator is designed to generate high-frequency (5 kHz) pulsed excitation of the airstream. High-pressure (100 psi) air from **20** drives resonator **14**, a Hartmann resonator that pro-



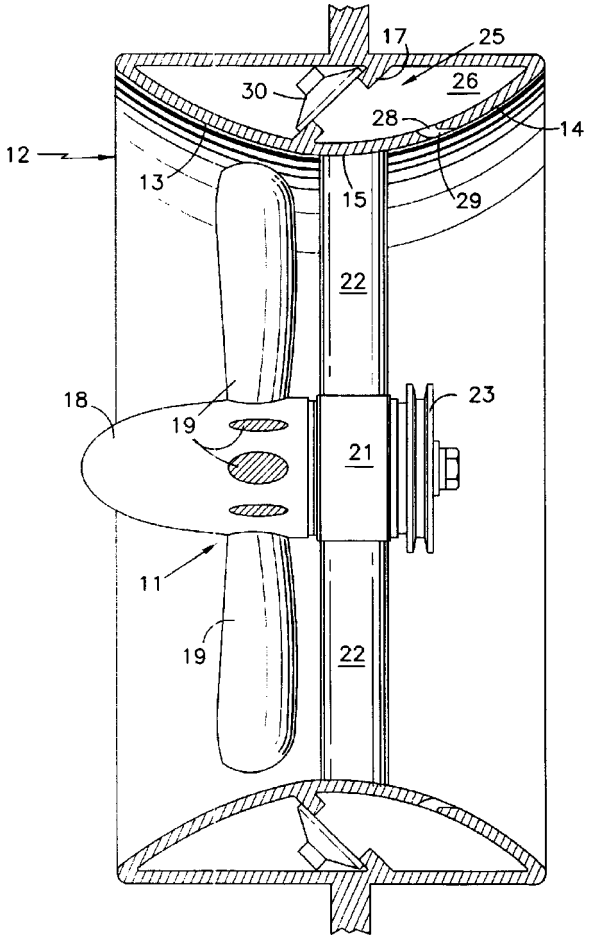
vides high-frequency, supersonic, airstream pulses. Another application of these resonators is proposed in which they are arranged near the leading edge of a cavity such as a weapons bay on a military aircraft. They serve to reduce the flow-induced cavity resonances by energizing the shear layer.—KPS

6,390,418

43.50.Nm TANGENTIALLY DIRECTED ACOUSTIC JET CONTROLLING BOUNDARY LAYER

Duane C. McCormick and Daniel L. Gysling, assignors to United Technologies Corporation
21 May 2002 (Class 244/204); filed 25 February 1999

Boundary layer separation is controlled through acoustic sources that are placed below the surface and that are directed tangential to the flow. In the figure, an example application shows a loudspeaker **30** controlling the boundary layer via aperture **28**. This Helmholtz resonator arrangement enables the generation of high velocities at the aperture. Other illustrated examples include flow over airfoils, engine inlets, and centrifugal fans. Extensive results that illustrate the effectiveness of the method are included.—KPS

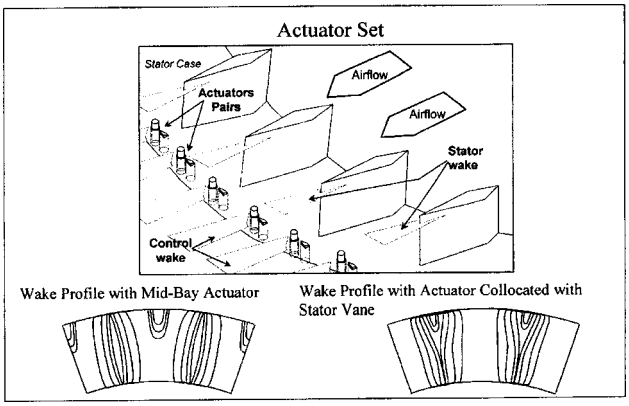


6,409,465

43.50.Nm BLADE VIBRATION CONTROL IN TURBO-MACHINERY

Andreas von Flotow and George Zipfel, assignors to Hood Technology Corporation
25 June 2002 (Class 415/1); filed 29 August 2000

A method to reduce blade vibration in turbomachinery involves the placement of flow obstructions or gas injectors upstream of the blades. These only need be employed when the rotor speed corresponds to resonant excitation. In contrast to simply filling the wake from the upstream stators, this approach modifies the spatial distribution in such a way as to reduce modal excitation, thus requiring less input power.—KPS



6,427,537

43.50.Yw MEASURING EQUIPMENT

Nils Christer Svensson, Tullinge, Sweden
6 August 2002 (Class 73/660); filed in Sweden 8 April 1998

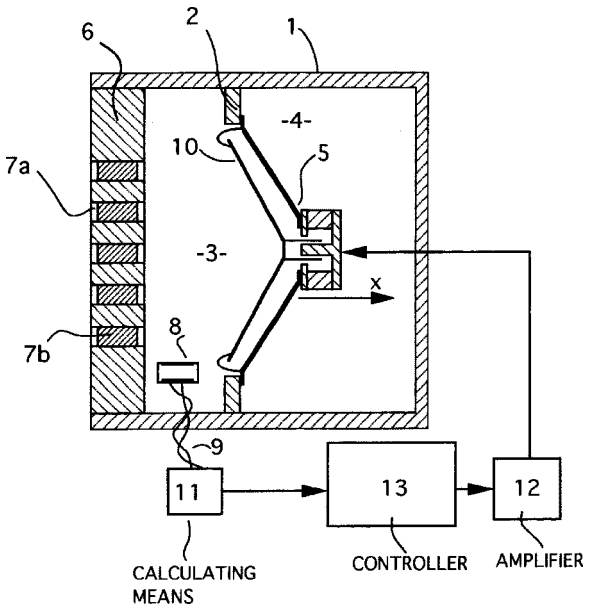
This rather generically named device includes at least two sensors and a measuring card by which the sensors are coupled to a computer unit. At least one of the sensors measures prevailing frequencies and at least one of the sensors evaluates prevailing tacho-pulses of an object to be evaluated. The measuring card circuits evaluate prevailing sound and vibration generated by the rotating object by dividing the time interval between two tacho-pulses into a number of subsections and generating a circuit-internal tacho-signal for each subsection.—DRR

6,408,078

43.55.Lb ACTIVE REACTIVE ACOUSTICAL ELEMENTS

Maximilian Hobelsberger, Wuerenlingen, Switzerland
18 June 2002 (Class 381/96); filed 12 April 2000

In the field of architectural acoustics, achieving sufficient low-frequency absorption can be difficult, especially in small rooms. Low-Q Helmholtz resonators are sometimes used, but to be very effective they must



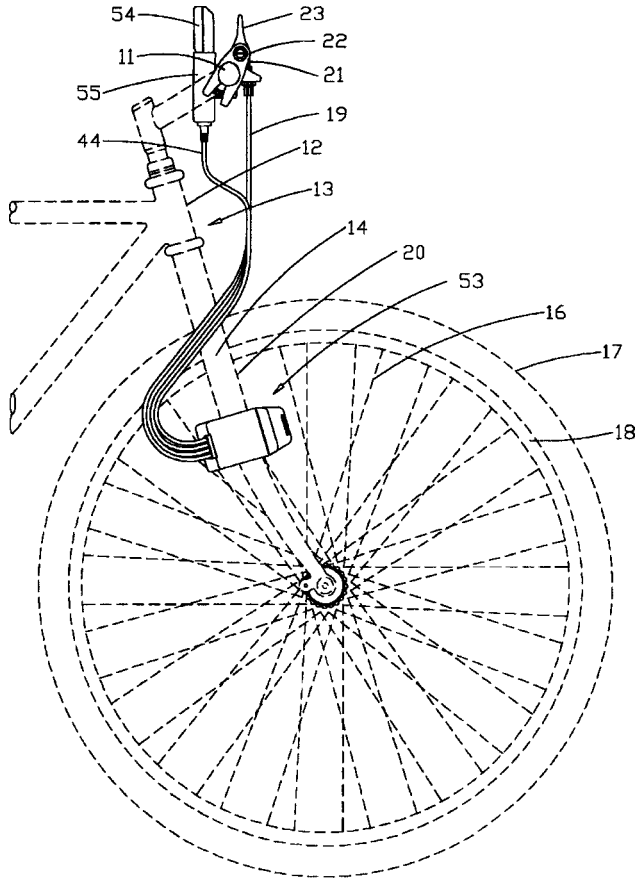
also be very large. In the active device shown, the effective volume of Helmholtz chamber **3** is dynamically controlled by movements of diaphragm **10** in response to pressure sensor **8**.—GLA

6,406,049

43.58.Wc AMUSEMENT AND ALERT ACCESSORY FOR BICYCLES

James W. Jimison, Palo Alto, and Ronald L. Coleman, Oakland, both of California
18 June 2002 (Class 280/288.4); filed 18 May 1999

Instead of a card that is hit by rotating bicycle spokes, imagine a ''striker rod'' that is used to sense wheel revolution and generate an acoustic



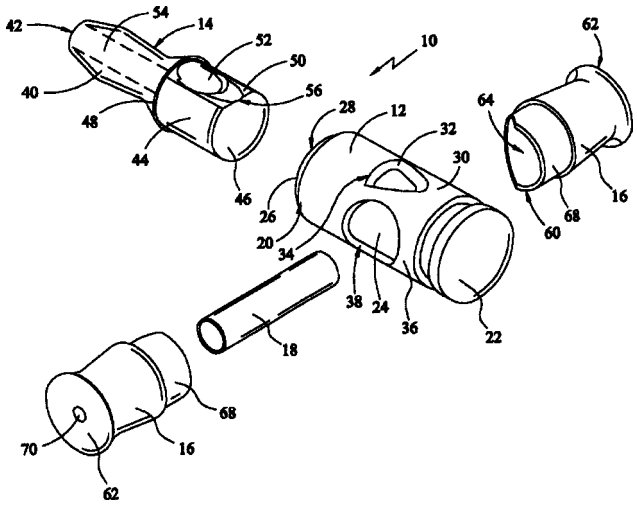
signal. Nowhere is the effect of the striker on the retardation of speed discussed.—MK

6,413,139

43.58.Wc WHISTLE-TYPE DUCK WHISTLE

Wesley E. Douglas, Los Banos, California
2 July 2002 (Class 446/204); filed 14 April 2000

Duck calls often take a whistling form. The problem is how to generate a suitably loud whistle to attract flying ducks close enough to the blind to



become a meal. Rather than use a vibrating membrane, the inventor proposes a whistle with features.—MK

6,413,140

43.58.Wc MODULAR GAME CALL SYSTEM

Wilbur R. Primos, assignor to Primos, Incorporated
2 July 2002 (Class 446/207); filed 29 July 1997

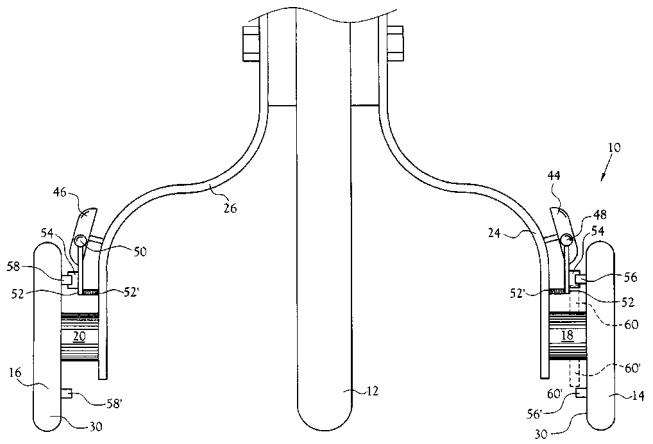
Different wild mammalian game have different vocal tracts. The inventor proposes a modular system with differing attachments for differing mammals.—MK

6,419,256

43.58.Wc NOISE FEEDBACK FOR TRAINING WHEELS

Richard R. Clark, Mooresburg, Tennessee
16 July 2002 (Class 280/288.4); filed 20 April 2001

Learning to ride a bicycle is often aided and abetted by the addition of training wheels **30**. The inventor proposes the addition of sound generators



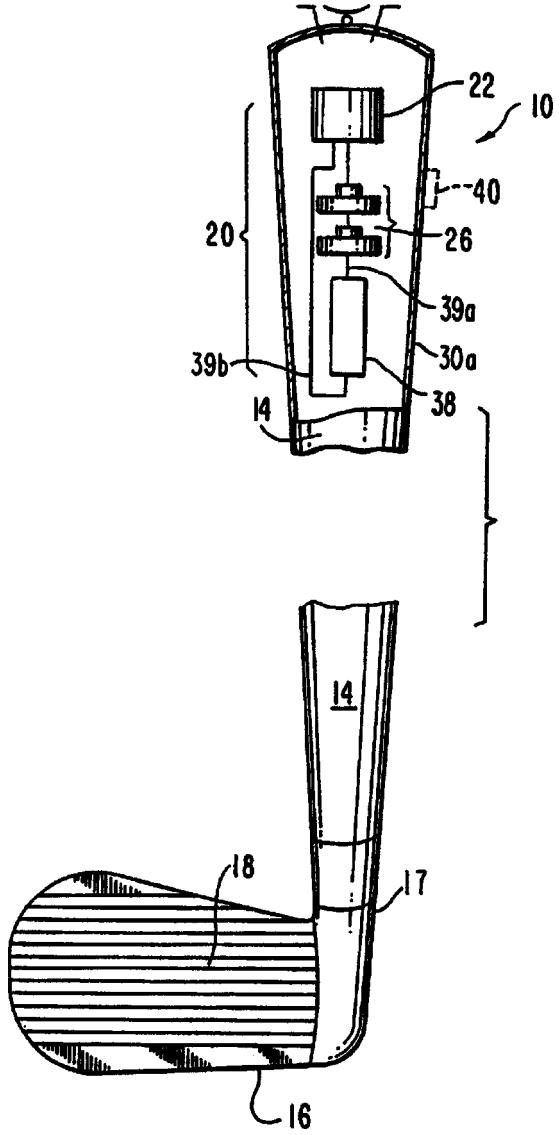
(such as bells **44**, **46**) that sound off when the training wheels touch the ground. This has the added benefit that the child will be easily found—until they learn how to balance.—MK

6,413,167

43.58.Wc GOLF OVERSWING ALERTING MECHANISM AND GOLF CLUB WITH OVERSWING ALERTING MECHANISM

Thomas J. Burke, Whitehouse Station, New Jersey
2 July 2002 (Class 473/224); filed 9 November 1999

Consider adding an accelerometer **38** to a golf club to measure shaft



speed as well as detecting ''overswing.'' The club generates an audio signal when it thinks the duffer is exhibiting bad form. Golf pro in a club?—MK

6,420,986

43.60.Bf DIGITAL SPEECH PROCESSING SYSTEM

Mark Shahaf *et al.*, assignors to Motorola, Incorporated
16 July 2002 (Class 341/139); filed in the United Kingdom 20 October 1999

In this speech processing system having a finite range of audio levels, the system receives an incoming audio signal and amplifies the signal with an audio gain factor. A method is provided for decreasing the gain factor when clipping of the amplified signal is detected. The gain factor is maintained for a hold time period and is increased when the incoming sound level amplification is lower than the highest level of the finite range of audio levels.—DRR

6,429,191

43.64.Gz TREATMENT OF HEARING IMPAIRMENTS

Wei-Qiang Gao, assignor to Genentech, Incorporated
6 August 2002 (Class 514/2); filed 30 March 2001

Compositions and methods are provided for prophylactic or therapeutic treatment of a mammal for hearing impairments involving neuronal damage, loss, or degeneration, preferably of spinal ganglion neurons, by administration of a therapeutically effective amount of a trkB or trkC agonist, particularly a neurotrophin, more preferably NT-4/5. Also provided are improved compositions and methods for treatments requiring administration of a pharmaceutical having an ototoxic side-effect, wherein the improvement includes administering a therapeutically effective amount of a trkB or trkC agonist to treat the ototoxicity.—DRR

6,428,484

43.64.Ri METHOD AND APPARATUS FOR PICKING UP AUDITORY EVOKED POTENTIALS

Rolf Dietter Battmer, Hannover, The Netherlands *et al.*
6 August 2002 (Class 600/554); filed in France 5 January 2000

A device is described for picking up auditory evoked potentials generated by acoustic, electrical, and/or mechanical stimulation of the cochlea. The implantable device includes at least two extracochlear pickup electrodes connected to the inputs of a differential amplifier.—DRR

6,430,443

43.64.Yp METHOD AND APPARATUS FOR TREATING AUDITORY HALLUCINATIONS

Manuel L. Karell, San Francisco, California
6 August 2002 (Class 607/55); filed 21 March 2000

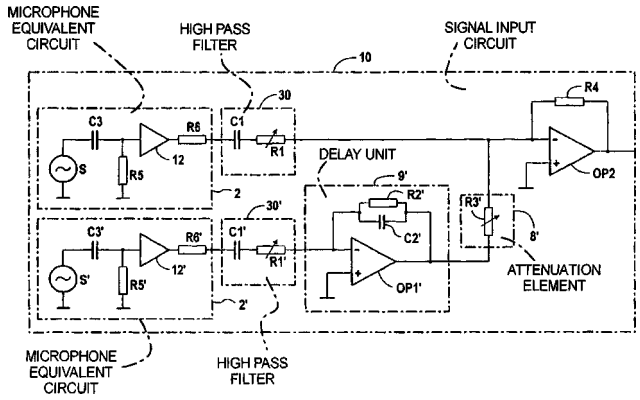
It is argued here that stimulating one or more vestibulocochlear nerves or the cochlea or cochlear regions will treat, prevent, and control auditory hallucinations. The claim is made that an abnormal brain activation inducing auditory hallucinations may be blocked by applying modulating electrical signal stimulation of a vestibulocochlear cranial nerve or cochlea or cochlear region. An abnormal brain activation inducing auditory hallucinations may be blocked by applying sound, audible or inaudible, with or without bone conduction, to an ear.—DRR

6,421,448

43.66.Ts HEARING AID WITH A DIRECTIONAL MICROPHONE CHARACTERISTIC AND METHOD FOR PRODUCING SAME

Georg-Erwin Arndt *et al.*, assignors to Siemens Audiologische Technik GmbH
16 July 2002 (Class 381/312); filed in Germany 26 April 1999

A directional hearing aid is typically made from two omnidirectional microphones of the same type, one located in each of two parallel signal paths. Initially these microphones have differing amplitude and phase characteristics. Rather than matching the transfer functions of the microphones themselves, which is costly and time consuming, a high-pass filter with adjustable corner frequency is placed in series with each microphone output.



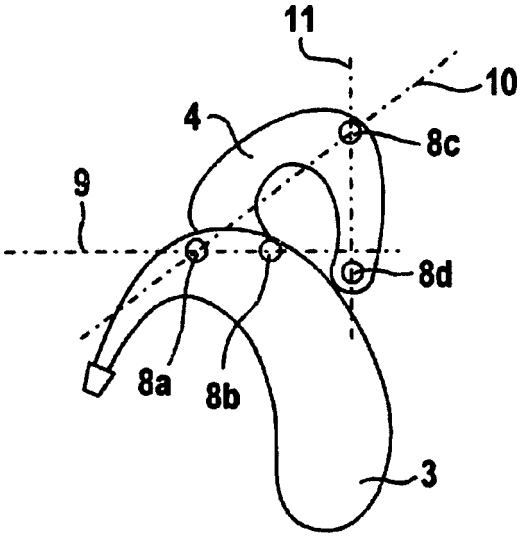
The 3-dB cutoff frequency of each of the high-pass filters is matched to the cutoff of the microphone in the opposite signal path. By using existing coupling capacitors and programmable means to adjust the values of series resistors that are already in the circuit, the two microphones are said to be matched without adding more components.—DAP

6,424,721

43.66.Ts HEARING AID WITH A DIRECTIONAL MICROPHONE SYSTEM AS WELL AS METHOD FOR THE OPERATION THEREOF

Werner Hohn, assignor to Siemens Audiologische Technik GmbH
23 July 2002 (Class 381/313); filed in Germany 9 March 1998

A hearing aid system containing a combination of nondirectional and directional microphones is useful for processing desired sounds from the sides and rear of the hearing aid wearer. To achieve the desired performance, the microphone outputs are cross-connected with different weightings and different delays. Interconnections can be in pairs of microphones **8a**, **8b** for making a first-order gradient directional system or in three or more microphones **8c**, **8d** to achieve a higher-order directional system. In order to



change the directions of maximum sensitivity and nulls and to facilitate packaging for hearing aid designs with many microphones, some of the microphones can be packaged in a hinged, leverlike auxiliary module that is attached to the main hearing aid case. A change in the maximum sensitivity direction may result, for example, if the energy from a particular direction or the peak-to-average level difference exceeds one or more predetermined stored thresholds. Alternately, a change in maximum sensitivity direction may occur if the difference in spectral content of the energy or modulation frequency received from several directions deviates from a stored reference value.—DAP

6,424,722

43.66.Ts PORTABLE SYSTEM FOR PROGRAMMING HEARING AIDS

Lawrence T. Hagen and David A. Preves, assignors to Micro Ear Technology, Incorporated
23 July 2002 (Class 381/314); filed 18 July 1997

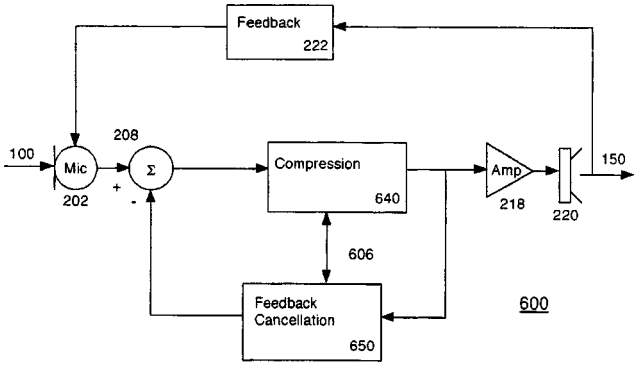
A host computer containing a PCMCIA interface downloads programs to a portable multi-program hearing aid programming system. The host computer may be a lap top, notebook, or hand-held type. Programming software is downloaded to the PCMCIA card when the card is inserted into the host computer, or stored in nonvolatile memory contained on the card. Different programs may be utilized for different ambient listening environments. Once programmed, the portable multiprogram unit can be decoupled from the PCMCIA interface and utilized by a dispenser or hearing aid wearer to modify the programs in the hearing aids. Parameters for selected programs are transmitted from the portable programmer to hearing aids via a wireless connection.—DAP

6,434,246

43.66.Ts APPARATUS AND METHODS FOR COMBINING AUDIO COMPRESSION AND FEEDBACK CANCELLATION IN A HEARING AID

James Mitchell Kates and John Laurence Melanson, assignors to GN ReSound A/S
13 August 2002 (Class 381/312); filed 2 October 1998

In one embodiment of the invention, the acoustic feedback signal is modeled and subtracted from the audio signal received from the microphone. Audio compression and feedback cancellation algorithms are designed to work together. The feedback cancellation algorithm provides information to the compression algorithm and the compression algorithm is



adjusted accordingly. For example, the maximum gain provided at low levels is reduced if the feedback signal detected is a sinusoid. The compression section also can provide information to the feedback cancellation section such as input signal power as a function of frequency. This information may be used to adjust the adaptation time constant of the feedback cancellation element.—DAP

6,434,247

43.66.Ts FEEDBACK CANCELLATION APPARATUS AND METHODS UTILIZING ADAPTIVE REFERENCE FILTER MECHANISMS

James Mitchell Kates and John Laurence Melanson, assignors to GN ReSound A/S
13 August 2002 (Class 381/312); filed 30 July 1999

A feedback cancellation system for hearing aids includes a first filter that models the quickly varying parts of the acoustic feedback path and is updated continuously. A second filter, which constrains adaptation and moni-

tors more slowly varying changes, is updated once there is an accurate estimate of the acoustic feedback path. As part of the update algorithm for the first filter, the system determines whether the audio signal is suitable or not for estimating the acoustic feedback path. In one embodiment, this decision is based on the audio signal being broadband, indicating no feedback oscillation is present.—DAP

6,396,954

43.72.Dv APPARATUS AND METHOD FOR RECOGNITION AND APPARATUS AND METHOD FOR LEARNING

Tetsujiro Kondo, assignor to Sony Corporation
28 May 2002 (Class 382/224); filed in Japan 26 December 1996

This patent for speech recognition in noisy environments uses images of mouth movements in addition to the voice data to improve the performance of the system. It would be especially useful in car navigation systems in which unwanted sounds, such as a CD player, an engine, or an air-conditioner, could affect the accuracy of a speech recognition system.—HHN

6,385,570

43.72.Gy APPARATUS AND METHOD FOR DETECTING TRANSITIONAL PART OF SPEECH AND METHOD OF SYNTHESIZING TRANSITIONAL PARTS OF SPEECH

Moo-young Kim, assignor to Samsung Electronics Company, Limited
7 May 2002 (Class 704/200); filed in the Republic of Korea 17 November 1999

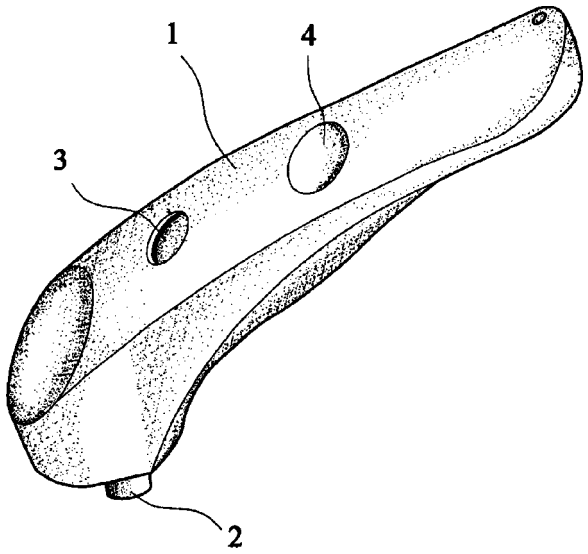
This patent argues, with some justification, that current speech coding methods concentrate on band amplitudes within steady-state portions, to the neglect of transitional portions of the signal. The present system measures the amplitude of peaks of the linear prediction residual signal, by samples rather than by frames. At the residual peaks, extra subframes are generated which include phase information such as is typically computed in a harmonic speech coder. This is said to result in much better rejection of background noises.—DLR

6,385,554

43.72.Ja DIGITAL TIRE GAUGE WITH VISUAL AND VOICE RESPONSE SYSTEM

Leo Wu, Chang Hua Hsien, Taiwan, Province of China
7 May 2002 (Class 702/140); filed 2 November 1999

These days, almost anything can be made to talk if there is room for a chip or two, a loudspeaker, and a small battery. In this case, it is a tire pressure gauge. The invention enhances the obvious design by adding a



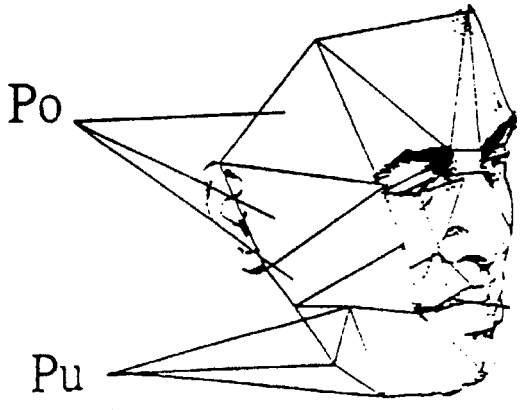
timer which counts how long it has been since the gauge was last used. At a preset interval, the gauge announces that it might be a good idea to check your tire pressure again. Let's just hope you happen to be within earshot when that happens.—DLR

6,385,580

43.72.Ja METHOD OF SPEECH SYNTHESIS

Bertil Lyberg and Mats Wiren, assignors to Telia AB
7 May 2002 (Class 704/258); filed in Sweden 25 March 1997

This is a method for simultaneously synthesizing a speech signal and a video talking head. One database contains the coordinates of key points on the head shape model. A second database contains speech segments from at least two speakers, chosen first according to the speech sounds, but then



modified according to the motions as required for the video display. This results in a new database containing speech sound units with duration information and specifying the motions of the key points on the head model.—DLR

6,382,206

43.72.Kb SPEECH TRANSMISSION ADAPTOR FOR USE WITH A RESPIRATOR MASK

Joyce B. Palazzotto and Harold R. Carpenter, assignors to 3M Innovative Properties Company
7 May 2002 (Class 128/201.19); filed 29 September 1997

This speech transducer is designed for use with a respirator mask such as used in hazardous environments. The design does not differ from earlier designs in that the speech pickup unit fits within the mask, allowing more natural speech with less influence of mask resonances. It differs from earlier designs in that it attaches to an air intake opening and may be fitted and later removed without requiring modification of an existing mask.—DLR

6,397,179

43.72.Ne SEARCH OPTIMIZATION SYSTEM AND METHOD FOR CONTINUOUS SPEECH RECOGNITION

Jean-Francois Crespo *et al.*, assignors to Nortel Networks Limited
28 May 2002 (Class 704/242); filed 4 November 1998

This patent covers a method for reducing the word grammar processing time in a continuous speech recognition system by the use of a semantically null word. The searching time is minimized by providing an exact *n*-best list of semantically meaningful words.—HHN

6,421,644

43.72.Ne INFORMATION APPARATUS FOR DISPATCHING OUTPUT PHRASE TO REMOTE TERMINAL IN RESPONSE TO INPUT SOUND

Hiromi Okitsu, assignor to Yamaha Corporation
16 July 2002 (Class 704/270.1); filed in Japan 8 March 1998

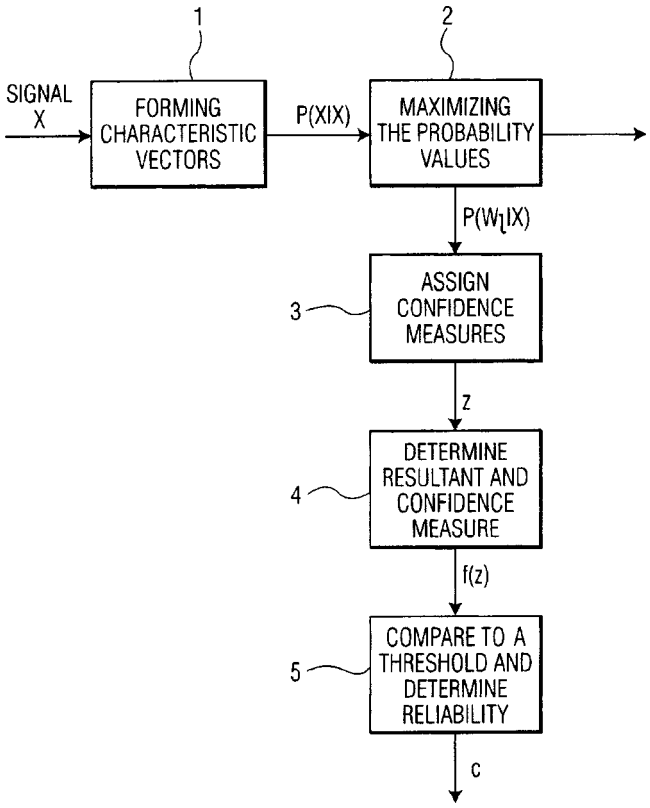
Obsfuscation can improve almost any idea, including the one buried in this conceptual patent. Somehow, multiple audio channels are analyzed (the FFT is ubiquitously mentioned) and ''commands'' are sent to a remote terminal. In contrast to most Yamaha patents, this one is conspicuously short on realistic details and concepts.—MK

6,421,640

43.72.Ne SPEECH RECOGNITION METHOD USING CONFIDENCE MEASURE EVALUATION

Jannes G. A. Dolfing and Andreas Wendemuth, assignors to Koninklijke Philips Electronics N.V.
16 July 2002 (Class 704/236); filed in Germany 16 September 1998

Speech recognition accuracy is improved, particularly for single-utterance command-and-control and dictation applications, by evaluating a global confidence (reliability) measure determined from a combination of several primary confidence measures. The parameter weights for linearly combining the primary confidence measures are determined by minimizing a



cross-entropy-error measure. The resulting global confidence measure is compared with a predetermined threshold to decide whether a recognition result represents the actual speech utterance. Further reduction in the error rate is obtained by adapting the confidence measure with a user-specific utterance.—DAP

6,430,532

43.72.Ne DETERMINING AN ADEQUATE REPRESENTATIVE SOUND USING TWO QUALITY CRITERIA, FROM SOUND MODELS CHOSEN FROM A STRUCTURE INCLUDING A SET OF SOUND MODELS

Martin Holzapfel, assignor to Siemens Aktiengesellschaft
6 August 2002 (Class 704/258); filed in Germany 8 March 1999

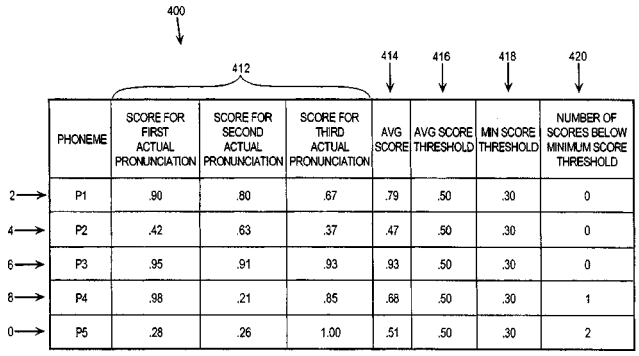
Hidden Markov models are now used habitually for speech recognition (see Rabiner and Huang for examples). Although the patent tries to broaden the claims to any class of sounds, all the examples are speech. This omission is notable because the feature selection is critical to any recognition scheme and the inventor says nothing about it.—MK

6,434,521

43.72.Ne AUTOMATICALLY DETERMINING WORDS FOR UPDATING IN A PRONUNCIATION DICTIONARY IN A SPEECH RECOGNITION SYSTEM

Etienne Barnard, assignor to Speech Works International, Incorporated
13 August 2002 (Class 704/244); filed 24 June 1999

Automatic determination of the accuracy of a pronunciation dictionary in a speech recognition system is achieved by comparing at a phomene level a pronunciation representation for a particular word from the dictionary to one or more actual utterances of the word. The accuracy scores for the



	PHONEME	SCORE FOR FIRST ACTUAL PRONUNCIATION	SCORE FOR SECOND ACTUAL PRONUNCIATION	SCORE FOR THIRD ACTUAL PRONUNCIATION	AVG SCORE	AVG SCORE THRESHOLD	MIN SCORE THRESHOLD	NUMBER OF SCORES BELOW MINIMUM SCORE THRESHOLD
2→	P1	.90	.80	.67	.79	.50	.30	0
4→	P2	.42	.63	.37	.47	.50	.30	0
6→	P3	.95	.91	.93	.93	.50	.30	0
8→	P4	.98	.21	.85	.68	.50	.30	1
0→	P5	.28	.26	1.00	.51	.50	.30	2

phonemes are compared to stored criteria to determine if the stored pronunciation satisfies the accuracy criteria. The dictionary is updated manually or automatically to reflect the actual pronunciations if the accuracy scores for each of the stored phoneme pronunciations is less than their accuracy thresholds.—DAP

6,420,641

43.75.Mn MULTIMEDIA KEYBOARD WITH INSTRUMENT PLAYING DEVICE

Yen-Liang Kuan, assignor to Behavior Tech Computer Corporation
16 July 2002 (Class 84/658); filed 6 July 2001

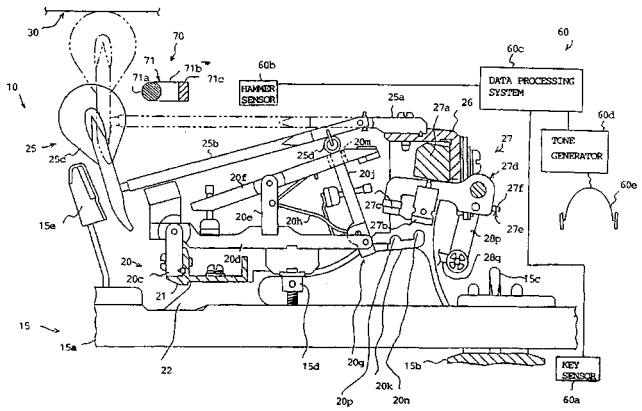
It would seem a reasonable idea to combine the computer keyboard with a musical instrument interface. This patent proposes just that. Using a series of resistive sensors, the keyboard is extended to music. But the touch and feel of these interfaces are completely unfamiliar to any performer, so this interface is at best a toy.—MK

6,423,889

43.75.Mn REGULATING BUTTON MECHANISM FOR EASILY REGULATING ESCAPE TIMING, SILENT SYSTEM COOPERATIVE THEREWITH AND KEYBOARD MUSICAL INSTRUMENT EQUIPPED THEREWITH

Satoshi Inoue, assignor to Yamaha Corporation
23 July 2002 (Class 84/236); filed in Japan 19 May 2000

A silent piano would seem to defeat the purpose, wouldn't it? Not if



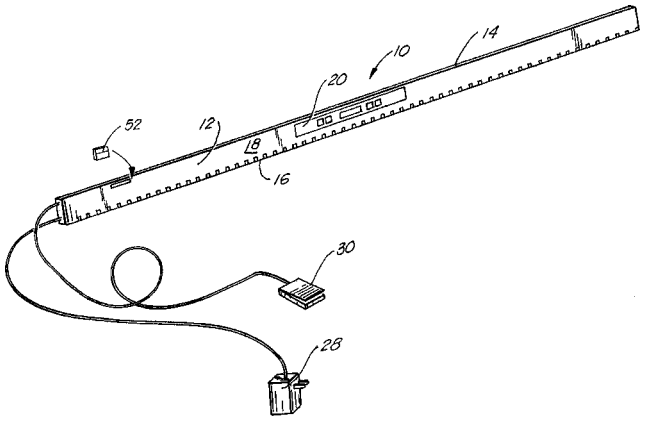
the piano keyboard is being used as an 88-key sensor. Accordingly, modifications are made to the action so that no sound will be generated.—MK

6,407,324

43.75.St PIANO INSTRUCTIONAL APPARATUS

Robert P. Hulcher, Mobile, Alabama
18 June 2002 (Class 84/478); filed 1 December 2000

The inventor proposes a detachable array of LEDs, one per key, for piano instruction. In addition, the LED color can be used to indicate use of



the right or left hand. Of course, this encourages key watching, which is not a desirable skill in either typists or pianists.—MK

6,385,581

43.75.Wx SYSTEM AND METHOD OF PROVIDING EMOTIVE BACKGROUND SOUND TO TEXT

Stanley W. Stephenson, Spencerport, New York
7 May 2002 (Class 704/270); filed 10 December 1999

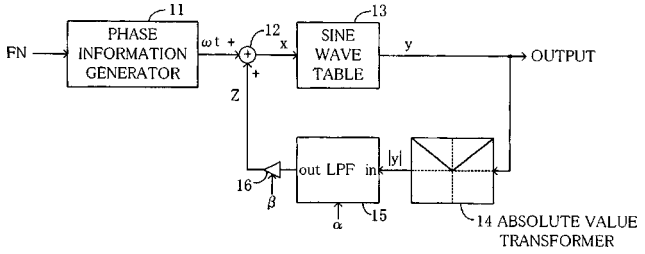
This patent describes a database of musical sounds or acoustic themes, each keyed to a specific word. Speech is recognized to produce a stream of words or words may simply be entered as text from a computer link. Each input word is checked for occurrence in the database. If found, the associated sound is played. The result is a sound track which may be played separately or mixed with an input speech track.—DLR

6,410,838

43.75.Wx MUSICAL SOUND SIGNAL SYNTHESIZER AND METHOD FOR SYNTHESIZING MUSICAL SOUND SIGNALS USING NONLINEAR TRANSFORMER

Chifumi Takeuchi, assignor to Yamaha Corporation
25 June 2002 (Class 84/607); filed in Japan 16 December 1999

FM generated sounds have a fairly distinctive timbre. The issue is how to use an FM tone generator with a simple operator—in this case table lookup with absolute value. Unfortunately, this idea is predated by the



wavetable lookup schemes of LeBrun and Arfib dating back to the mid-1970s. The additional wrinkle shown here is the feedback loop to the phase accumulator.—MK

6,411,225

43.75.Wx SAMPLE RATE CONVERTER

Adrianus Wilhelmus Maria Van Den Enden *et al.*, assignors to Koninklijke Philips Electronics N.V.
25 June 2002 (Class 341/61); filed in the European Patent Office 22 April 1999

It is well known that sampling rate conversion is easiest when the input and output rates are rational—then a polyphase filter can be used to implement an efficient computation. Irrational sampling rate conversion is much more complex since the input sample to output sample distance is irrational. The standard solution is to introduce an interpolator just as the authors propose. The use of Lagrange and bandlimited interpolation methods for audio signals was introduced by Smith and Gossett at ICASSP in 1984. So, what is new here?—MK

6,426,456

43.75.Wx METHOD AND APPARATUS FOR GENERATING PERCUSSIVE SOUNDS IN EMBEDDED DEVICES

Jean Khawand and Radu Frangopol, assignors to Motorola, Incorporated
30 July 2002 (Class 84/624); filed 26 October 2001

FM synthesis, as originally described (and patented) by Chowning in 1967, can be made to do many un-natural acts, including the simulation of percussion. Essentially, this patent is an FM patch together with filtered white noise. What is ignored is the classic FM problem: what are the values of the parameters and how do they relate to real instruments?—MK

6,423,007

43.80.Qf ULTRASONIC SYSTEMS AND METHODS FOR CONTRAST AGENT CONCENTRATION MEASUREMENT

Frederic Louis Lizzi and Cheri Xiaoyu Deng, assignors to Riverside Research Institute
23 July 2002 (Class 600/458); filed 11 December 2000

Contrast agent concentration or radii are determined by first estimating the contrast echo power spectrum and then correlating one or more parameters of the spectrum to a known size distribution function for the contrast agent. If the mean square radius of the particles is known, the concentration of the contrast agent can be determined. If the concentration of the agent is constant, relative variations in mean square radius can be determined.—RCW

6,425,874

43.80.Qf METHOD AND APPARATUS FOR CHARACTERIZING GASTROINTESTINAL SOUNDS

Richard H. Sandler and Hussein A. Mansy, assignors to Rush—Presbyterian—St. Luke's Medical Center
30 July 2002 (Class 600/586); filed 23 March 2000

This system for characterizing gastrointestinal sounds includes a microphone array to be positioned on a body. The microphone signals are digitized and their spectra and duration are determined by a processor. The state of the gastrointestinal tract is characterized by the spectra and duration of the sound or event.—DRR

6,428,479

43.80.Qf ULTRASONOGRAPHY OF THE PROSTATE

Anne Kirsti Aksnes *et al.*, assignors to Nycomed Imaging AS
6 August 2002 (Class 600/458); filed in the United Kingdom 17 December 1997

Prostate abnormalities such as cancer may be detected by ultrasonic determination of the in-flow kinetics of contrast agent-containing blood in the prostate and/or by observation of disease-related asymmetries in the spokelike vascular pattern of the prostate.—DRR

6,432,050

43.80.Qf IMPLANTABLE ACOUSTIC BIO-SENSING SYSTEM AND METHOD

Yariv Porat *et al.*, assignors to Remon Medical Technologies Limited
13 August 2002 (Class 600/300); filed 3 May 1999

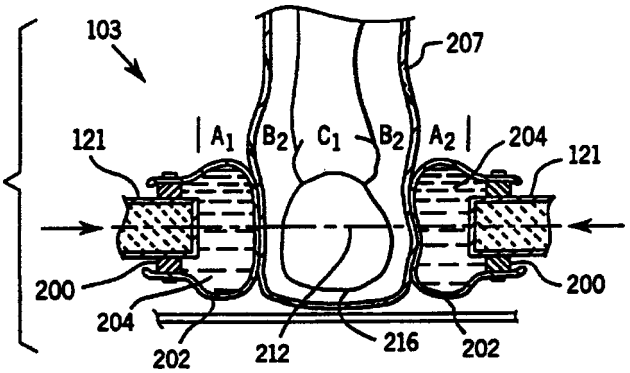
An implantable biosensor system is provided for monitoring and optionally alleviating a physiological condition in a patient. The system includes (a) at least one sensor for sensing at least one parameter of a physiological condition and for generating electrical sensor signals representative of the physiological condition and (b) an acoustic activatable transducer directly or indirectly coupled with the sensor. The transducer converts a received acoustic interrogation signal from outside the patient's body into electrical power for energizing the processor. The activatable transducer also converts the electrical sensor signals into acoustic signals receivable from outside the patient's body such that information pertaining to the physiological condition can be relayed outside the patient's body upon generation of an acoustic interrogation signal.—DRR

6,432,057

43.80.Qf STABILIZING ACOUSTIC COUPLER FOR LIMB DENSITOMETRY

Richard B. Mazess and Richard F. Morris, assignors to Lunar Corporation
13 August 2002 (Class 600/449); filed 1 June 2000

An ultrasonic densitometer has opposing ultrasonic transducers with inflatable bladders **202** containing a coupling fluid. The bladders are affixed to a housing with a simple mounting. The densitometer sizes the bladder diameter so as to be moved into contact with the patient **207** by inflation



only. The densitometer further controls the geometry of the inflated bladders so as to provide improved immobilization of the patient. The coupling fluid allows the bladder to couple the ultrasonic energy to the patient member.—DRR

6,419,648

43.80.Sh SYSTEMS AND METHODS FOR REDUCING SECONDARY HOT SPOTS IN A PHASED ARRAY FOCUSED ULTRASOUND SYSTEM

Shuki Vitek and Naama Brenner, assignors to Insightec-TxSonics Limited
16 July 2002 (Class 601/3); filed 21 April 2000

The patent describes methods for performing a therapeutic procedure using focused ultrasound with the intention of avoiding the generation of unintended ''hot spots.'' The system includes a piezoelectric transducer incorporating a number of transducer elements, e.g., a concave concentric ring array or a linear array of transducer elements. A drive circuit coupled to the transducer provides driving signals to the transducer elements at one of several discrete radio frequencies. A controller coupled to the drive circuitry periodically changes the drive signal frequency while controlling a phase component of the drive signals to maintain the focus of the transducer at a primary focal zone during a single sonification.—DRR

6,423,002

43.80.Sh INTRA-OPERATIVE DIAGNOSTIC ULTRASOUND MULTIPLE-ARRAY TRANSDUCER PROBE AND OPTIONAL SURGICAL TOOL

John A. Hossack, assignor to Acuson Corporation
23 July 2002 (Class 600/439); filed 24 June 1999

This patent covers a system for diagnostic ultrasound imaging with at least two but preferably three transducer arrays. The first array provides a primary image. The second is an array of transducer elements positioned near one end of the first array for providing an image in a different plane. An optional third array is positioned near the other end of the first array to provide an additional image in a different plane. One application of the invention is to collect tissue for medical procedures, wherein at least one surgical tool is attached to the probe near the imaging arrays and moves together with at least one imaging array. As the primary imaging array is drawn across tissue for collection, the operator observes a cross-section of the tissue prior to cutting the tissue. Another application of the invention allows an operator to selectively display pseudo or true 3-D images using the same probe. Another embodiment of the invention is said to permit an operator to determine the 3-D position of a foreign object in body tissue.—DRR

6,425,906

43.80.Sh ULTRASONIC CUTTING TOOL

Michael John Radley Young and Stephen Michael Radley Young, both of Ashburton, the United Kingdom
30 July 2002 (Class 606/169); filed in the United Kingdom 19 January 1998

A surgical tool for cutting and/or coagulating tissue includes a piezoelectric driver to generate ultrasonic energy in the form of torsional mode vibrations. A waveguide is operatively connected to the proximal end of the driver and extends a distance of $n\lambda$ where n designates an integer and λ represents the wavelength of ultrasonic vibration in the material of the horn or waveguide. The distal end of the waveguide is provided with a cutting and/or coagulating tool.—DRR

6,428,477

43.80.Sh DELIVERY OF THERAPEUTIC ULTRASOUND BY TWO DIMENSIONAL ULTRASOUND ARRAY

Martin K. Mason, assignor to Koninklijke Philips Electronics, N.V.
6 August 2002 (Class 600/437); filed 10 March 2000

A fully steerable two-dimensional ultrasound array delivers therapy by steering and selective focusing of beams. In some systems, the ultrasound array also includes an imaging functionality to simultaneously perform diagnostic imaging and therapy delivery. In one example, the two-dimensional ultrasound array includes a controller that controls beam forming and focusing to scan the focal point of the beam in a pattern within an identified structure of an image. Tissue is thus scanned using a sharply focused beam that is suitable for delivering a therapy such as hyperthermia or delivery of a pharmaceutical via microspheres. Imaging and therapy proceed either simultaneously or separately by operator selection. Simultaneous operation of imaging and therapy delivery are achieved by pulsing a focused, scanned beam to deliver intensity levels suitable for heating tissue or for bursting microspheres. Reflected signals from the pulses are detected and used to create an image in the manner of conventional ultrasound imaging—DRR

6,428,491

43.80.Sh DELIVERY OF ULTRASOUND TO PERCUTANEOUS AND INTRABODY DEVICES

Dan Weiss, Hadera, Israel
6 August 2002 (Class 601/2); filed in Israel 27 August 1999

A system is described for reducing morbidity associated with the insertion of a medical device into the body of a patient. Ultrasonic energy is applied to the device with sufficient intensity to inhibit accretion of biological matter onto the device while the device is in the body. Typically, the energy is applied to the device in the absence of clinically observable accretion so as to prevent such accretion.—DRR

6,432,067

43.80.Sh METHOD AND APPARATUS FOR MEDICAL PROCEDURES USING HIGH-INTENSITY FOCUSED ULTRASOUND

Roy W. Martin *et al.*, assignors to University of Washington; Sonic Concepts, Incorporated
13 August 2002 (Class 601/2); filed 3 September 1999

A number of methods are described for enabling substantially bloodless surgery and for stemming hemorrhaging. High-intensity focused ultrasound (HIFU) is applied to form cauterized tissue regions prior to surgical incision, for example, developing a cauterized tissue shell around a tumor to be removed. The procedure is referred to as ''presurgical volume cauterization.'' In one embodiment, the method is claimed to be particularly effective for use in surgical lesion removal or resection of tissue having a highly vascularized constitution, such as the liver or spleen, and thus a propensity for hemorrhaging. In other embodiments, methods and apparatus for hemostasis using HIFU are to be applied in both surgical, presurgical, and medical emergency situations. In one apparatus embodiment, a telescoping, acoustic coupler is provided so that the depth of focus of the HIFU energy is controllable. The apparatus can be designed to be portable for use in emergency medical situations.—DRR

6,432,069

43.80.Sh COUPLING MEDIUM FOR HIGH-POWER ULTRASOUND

Joseph Godo and Emmanuel Blanc, assignors to Technomed Medical Systems, S.A.
13 August 2002 (Class 601/2); filed in France 25 March 1999

The subject coupling medium for high-power ultrasound includes a liquid aqueous solution of a hydrophilic polymer, notably polyvinylpyrrolidone. It is designed to be a part of an ultrasound therapy apparatus that includes a therapy transducer array adapted to transmit high-power ultrasound in an enclosure, a portion of which is filled with the coupling medium.—DRR

6,432,070

43.80.Sh METHOD AND APPARATUS FOR ULTRASONIC TREATMENT OF REFLEX SYMPATHETIC DYSTROPHY

Roger J. Talish and Alan A. Winder, assignors to Exogen, Incorporated
13 August 2002 (Class 601/2); filed 9 May 2000

The invention relates to a system for therapeutically treating reflex sympathetic dystrophy using ultrasound. The apparatus includes at least one ergonomically constructed ultrasonic transducer configured to cooperate with a placement module or strip for placement in proximity to pain receptors of the sympathetic nervous system. The apparatus also utilizes a portable, ergonomically constructed main operating unit constructed to fit within a pouch worn by the patient. In operation, at least one ultrasonic transducer is excited for a predetermined period of time. To ensure proper positioning, and to insure compliance with a treatment protocol, a safety interlock is provided to prevent inadvertent excitation of the ultrasonic transducer. In an alternate embodiment, the apparatus includes a treatment basin having a plurality of ultrasonic transducer assemblies placed on the perimeter thereof. The patient places an injured part of the body therein and the transducer assemblies are excited to impinge ultrasonic waves to the injured part of the body—DRR

6,432,118

43.80.Sh MULTIFUNCTIONAL CURVED BLADE FOR USE WITH AN ULTRASONIC SURGICAL INSTRUMENT

Jeffrey D. Messerly, assignor to Ethicon Endo-Surgery, Incorporated
13 August 2002 (Class 606/169); filed 28 August 2000

The present invention relates, in general, to ultrasonic surgical clamping instruments and, more particularly, to a multifunctional curved shearing blade for an ultrasonic surgical clamping instrument. An ultrasonic surgical instrument is disclosed that combines end-effector geometry to best affect the multiple functions of a shears-type configuration. The shape of the blade is characterized by a radius cut to form a curved and potentially tapered geometry. This cut creates a curved surface including a concave surface and a convex surface. The convex surface transitions into a short, straight, flat surface. The length of this straight portion affects, in part, the acoustic balancing of the transverse motion induced by the curved shape. Relative to straight blade tips, the tip curvature of the present design is said to provide improved visibility of the transection site and improved access to targeted tissues.—DRR

6,419,633

43.80.Vj 2D ULTRASONIC TRANSDUCER ARRAY FOR TWO DIMENSIONAL AND THREE DIMENSIONAL IMAGING

Andrew L. Robinson *et al.*, assignors to Koninklijke Philips Electronics N.V.
16 July 2002 (Class 600/443); filed 15 September 2000

The elements in this array can be operated to form a beam and scan a three-dimensional region. The elements can also be operated to form a linear array and scan a beam in a plane. A probe containing the array can be quickly switched between the two- and three-dimensional imaging modes to produce both two- and three-dimensional images in real time.—RCW

6,423,003

43.80.Vj ULTRASONIC IMAGING SYSTEM AND METHOD WITH SNR ADAPTIVE PROCESSING

Kutay Ustuner and Anming He, assignors to Acuson Corporation
23 July 2002 (Class 600/443); filed 29 October 1999

A processor in this system varies the received signal path parameters as a function of the signal-to-noise ratio in the echo. Background noise is either acquired by receiving with the transmitters off, estimated by using known differences in bandwidth and correlation lengths of the signal and noise, or computed by using a system noise model with current system parameters. Echos are then processed using a nonlinear function that depends on a comparison of the echo signal and the background noise. The processor may include this kind of adaptive filtering in azimuth, elevation, and time dimensions, in an implementation of high-pass, band-pass, and whitening functions, and in image compounding.—RCW

6,423,004

43.80.Vj REAL-TIME ULTRASOUND SPATIAL COMPOUNDING USING MULTIPLE ANGLES OF VIEW

Fang Dong *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
23 July 2002 (Class 600/443); filed 30 May 2000

Multiple angles of view are obtained by operator movement of a probe and successive image frames are compounded by using a sum-of-absolute-difference registration algorithm.—RCW

6,425,867

43.80.Vj NOISE-FREE REAL TIME ULTRASONIC IMAGING OF A TREATMENT SITE UNDERGOING HIGH INTENSITY FOCUSED ULTRASOUND THERAPY

Shahram Vaezy *et al.*, assignors to University of Washington
30 July 2002 (Class 600/439); filed 17 September 1999

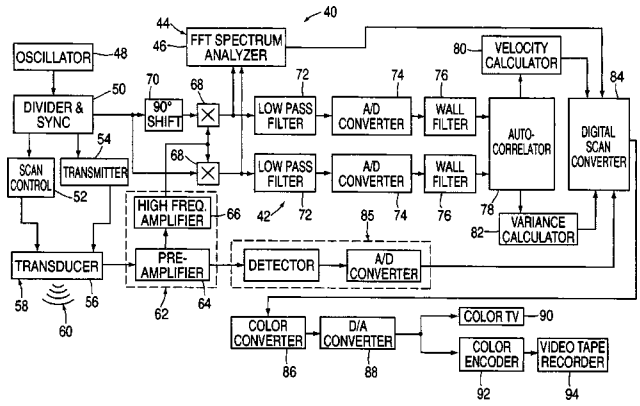
The probe in this system for simultaneous ultrasonic imaging and therapy limits undesirable interference noise in a display by synchronizing high-intensity focused ultrasonic waves with an imaging transducer to display the noise in an area that does not overlap the treatment site. For example, a low treatment beam power level that does not cause damage is used to identify the focus of the treatment beam by a change in echogenicity and then the power level is increased to a therapeutic level. The location of the focus is stored and displayed to facilitate spacing of the treatment sites in a region. As the treatment progresses, changes in the treatment site can be seen in the real-time, noise-free image.—RCW

6,425,868

43.80.Vj ULTRASONIC IMAGING SYSTEM

Tadashi Tamura, assignor to Aloka Company, Limited
30 July 2002 (Class 600/454); filed 26 July 1999

This system produces color flow and b-scan images with the line density in the color flow image comparable to that in the b-scan image. The scanned sequence of each transmitted and received color flow beam is laterally spaced through the imaging field to calculate the flow velocities. The color flow beams are transmitted and received only once for each position



along the direction of the imaging field to enable a large number of color flow lines to be acquired at a high frame rate. The large number of signals produced by the scanning technique are processed with high speed to produce and synchronize color flow and b-scan images for real-time display.—RCW

6,425,869

43.80.Vj WIDEBAND PHASED-ARRAY TRANSDUCER FOR UNIFORM HARMONIC IMAGING, CONTRAST AGENT DETECTION, AND DESTRUCTION

Patrick G. Rafter *et al.*, assignors to Koninklijke Philips Electronics, N.V.
30 July 2002 (Class 600/458); filed 18 July 2000

This transducer, in combination with transmit, receive, and display electronics, insonifies tissue at a fundamental frequency below 1.5 MHz. The contrast agent is detected by the harmonic response produced during the destruction of the agent by the ultrasonic pressure wave. The ultrasonic harmonic response of tissue is obtained by insonifying at a low frequency for increased uniformity in depth and uses a transmit apodization to improve harmonic generation in the near field.—RCW

6,432,055

43.80.Vj MEDICAL ULTRASONIC IMAGING SYSTEM WITH THREE-STATE ULTRASONIC PULSE AND IMPROVED PULSE GENERATOR

Stuart L. Carp *et al.*, assignors to Acuson Corporation
13 August 2002 (Class 600/437); filed 30 June 2000

This pulse generator produces a three-state, pulse width modulated, bipolar waveform by summing a waveform segment with an inverted, time-shifted version of the segment. Frequency characteristics of the waveform are controlled by selecting the interval of the time shift. The waveform is produced by a switched voltage source that has a low, constant impedance for each voltage state.—RCW

Acoustic ray chaos and billiard system in Hamiltonian formalism (L)

Tetsuji Kawabe,[a] Keisuke Aono, and Masakazu Shin-ya
Physics Department, Department of Acoustic Design, Kyushu Institute of Design, Shiobaru, Fukuoka 815-8540, Japan

(Received 11 December 2001; revised 9 August 2002; accepted 18 November 2002)

The acoustic ray model with a strong connection to the billiard problem is presented within the framework of the Hamiltonian form. Introducing the background function into the sound-speed profile to confine all rays in a closed space, we obtain the ray trajectories consistent with a billiard picture. The ray chaos is observed when the perturbation due to inhomogeneity of the medium is taken into account. Based on the Poincaré surface of section and the Lyapunov exponents, we confirm that the chaos is characterized by almost the same structure as one observed in many Hamiltonian systems with two degrees of freedom. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1536626]

PACS numbers: 43.25.Rq, 43.25.Ts, 43.30.Cq, 43.20.Dk [MFH]

The acoustic ray has a strong analogy with a particle trajectory in the billiard problem. The billiard problem is to investigate the dynamics of a particle moving freely with a constant speed in two dimensions surrounded by the hard wall and being perfectly reflected at impacts with the boundary.[1] The reflection on the boundary obeys the law that the angle of reflection equals the angle of incidence. Similarly, the acoustic ray under the high-frequency limit moves straight in the interior of a closed space and obeys the reflection law on the boundary with a constant speed in a homogeneous medium.[2]

The behavior of the trajectories in the billiard system sensitively depends on the shape of the boundary.[1] The trajectories behave regularly for a shape such as a circle, a rectangle and an oval, while they exhibit chaotic behavior for the shape of a stadium. This behavior of the trajectories in the billiard problem reflects the dynamics of the acoustic ray in a closed space as follows: In the domain with a rectangular or other simple shape, the acoustic ray propagates regularly in a manner that it moves straight in the interior and changes its direction at the wall according to the reflection law, while the ray behaves chaotic in the interior of the irregular shapes like the stadium type. Thus this analogy between the billiard and the ray system in a homogeneous medium seems to impose a strong restriction on the relation between the shape of closed space and the emergence of the acoustic ray chaos.

On the other hand, for a more realistic case that a medium has inhomogeneity, the ray motion is perturbed and its trajectory deviates from a straight line. For example, the ray trajectory is curved when there is a temperature fluctuation in the medium. In the conventional framework of the billiard problem studied so far, such an effect of the inhomogeneity has not been taken into account on the ray motion. From the point of view of the acoustic ray chaos, it is a very important issue to study the billiard problem for the perturbed ray propagating in the domain with regular shapes.

[a]Electronic mail: kawabe@kyushu-id.ac.jp

In this Letter we try to formulate the acoustic ray model that involves the billiard problem connected with the inhomogeneous medium. The ray equations are formulated by the Hamiltonian form. The billiard problem is realized by modeling the sound-speed profile so as to trap all rays in the bounded space. When the sound speed is perturbed by the inhomogeneity of medium, the ray trajectory exhibits chaotic behavior.

The propagation of wave is described by the acoustic ray equations,[3]

$$\frac{d\mathbf{r}}{dt}=\frac{\partial\omega}{\partial\mathbf{k}},\quad \frac{d\mathbf{k}}{dt}=-\frac{\partial\omega}{\partial\mathbf{r}},\quad \omega(\mathbf{r},\mathbf{k})=c(\mathbf{r})|\mathbf{k}|, \tag{1}$$

where ω is the angular frequency of the wave, $c(\mathbf{r})$ is the sound speed, and $\mathbf{r}=(z,r)\equiv(z,\sqrt{x^2+y^2})$ is the position vector, $\mathbf{k}=(k_z,k_r)$ is the wave number vector, and t is time. Since the angular frequency is a constant of motion, ω is regarded as the energy of the system E. Thus this $\omega(\mathbf{r},\mathbf{k})$ is equivalent to the autonomous Hamiltonian $H(z,r,k_z,k_r)$ with two degrees of freedom.[4] By defining the momentum $\mathbf{p}=(p_z,p_r)=E^{-1}\mathbf{k}$ with $|\mathbf{p}|=p$, the above Hamiltonian system becomes as follows:

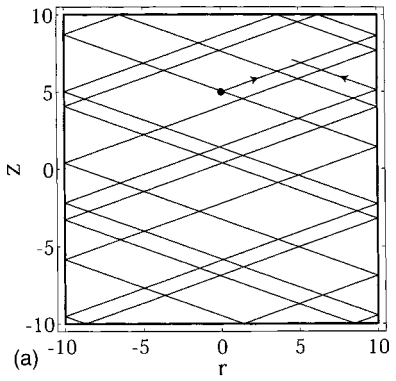
$$\frac{dz}{dt}=\frac{\partial\tilde{H}}{\partial p_z},\quad \frac{dp_z}{dt}=-\frac{\partial\tilde{H}}{\partial z},\quad \frac{dr}{dt}=\frac{\partial\tilde{H}}{\partial p_r},\quad \frac{dp_r}{dt}=-\frac{\partial\tilde{H}}{\partial r}. \tag{2}$$

Here $\tilde{H}$ is the rescaled Hamiltonian defined by

$$\tilde{H}=cp=c(z,r)\sqrt{p_z^2+p_r^2}, \tag{3}$$

where the relation $\tilde{H}=1$ holds. Thus, the final form of the acoustic ray equations is

$$\begin{aligned}&\frac{dz}{dt}=c^2p_z,\quad \frac{dp_z}{dt}=-\frac{1}{c}\frac{\partial c}{\partial z},\\ &\frac{dr}{dt}=c^2p_r,\quad \frac{dp_r}{dt}=-\frac{1}{c}\frac{\partial c}{\partial r},\end{aligned} \tag{4}$$



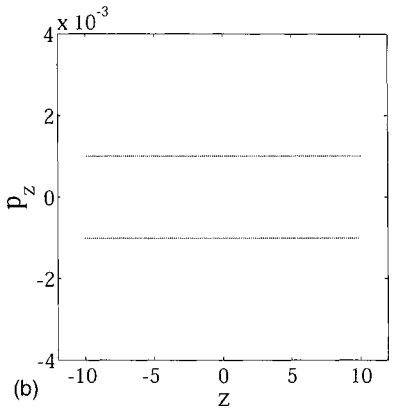
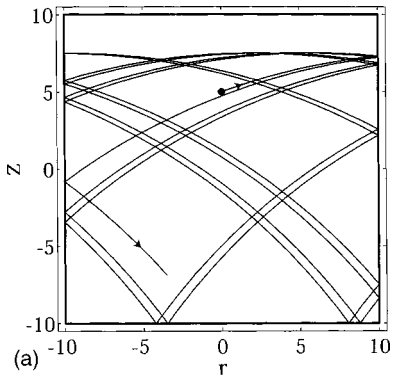


FIG. 1. (a) The ray trajectory and (b) the PSS for the case (a): a constant sound speed of (8), with $c_0=340$ m/s, $f=0$, and $h=0$. The initial conditions and parameters are as follows: $z(0)=5$ m, $r(0)=0$ m, $\theta(0)=\tan^{-1}(p_z/p_r)=20°$, $L_z=L_r=10$ m, $a=1$ m/s, and $n=4000$. The PSS is defined by a projection of z and p_z values on the (z,p_z) plane, at whose values the conditions $r=0$ and $dr/dt>0$ are satisfied. Note that the time evolution of the trajectory in (b) is much longer than in (a) to obtain the PSS points enough to draw the KAM curve.



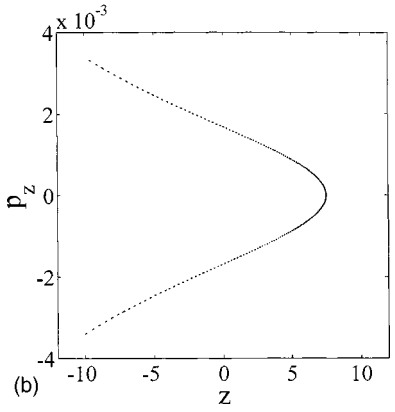


FIG. 2. (a) The ray trajectory and (b) the PSS for the case (b): a linearly increasing sound speed of (8), with $c_1=10$ m/s and $h=0$. The other parameters, the initial conditions, and the condition to draw the PSS are the same as those used in Fig. 1.

where three of the above equations are independent because of the constraint $cp=1$.

By differentiating the first and the third equation in (4) with respect to t, we can obtain the novel formula of the equations of motion for the ray trajectory as follows:

$$\frac{d^2z}{dt^2}=\frac{\partial W}{\partial z}\left(\frac{dz}{dt}\right)^2+\frac{\partial W}{\partial z}\frac{dr}{dt}\frac{dz}{dt}-\frac{\partial U}{\partial z}, \tag{5}$$

$$\frac{d^2r}{dt^2}=\frac{\partial W}{\partial r}\left(\frac{dr}{dt}\right)^2+\frac{\partial W}{\partial r}\frac{dz}{dt}\frac{dr}{dt}-\frac{\partial U}{\partial r}, \tag{6}$$

where $U=c^2/2$ and $W=\ln(c/c_0)^2$, and c_0 is a reference sound speed. Note that the formula of (5) and (6) is more transparent form than that of (4) to grasp the nonlinear structure in the present formalism. Owing to this nonlinearity, some ray trajectories will exhibit chaotic behavior for given functions U and W. Thus the sound speed c is a fundamental quantity and has a very important effect on the ray behavior.

In order to realize the two-dimensional billiard problem in the (z, r) space, we need to introduce a specific function into the sound-speed profile so as to trap all rays in a confined space. As such a function suitable for this purpose, we define the background function by

$$F_{\mathrm{BG}}(z,r)=a\left(\left(\frac{z}{L_z}\right)^n+\left(\frac{r}{L_r}\right)^n\right), \tag{7}$$

where a is a constant with dimension of velocity, and L_z and L_r are parameters to fix the scale of the region. With very large positive n, the background function $F_{\mathrm{BG}}(z,r)$ virtually becomes zero inside the domain of $|z|<L_z$ and $|r|<L_r$ while it becomes infinity for $|z|\geqslant L_z$ and $|r|\geqslant L_r$. Thus, the background function plays much the same role as a two-dimensional square well potential with rigid walls in which all rays are confined. By including this background function in the sound-speed profile, we can write the sound-speed profile as

$$c(z,r)=c_0+f(z)+\epsilon h(z,r)+F_{\mathrm{BG}}(z,r), \tag{8}$$

where $f(z)$ is the z dependent sound speed and $h(z,r)$ is the perturbation due to the inhomogeneity of medium, and ϵ is the strength of the perturbation.

In order to see how the sound-speed profile produces the ray motion consistent with the billiard picture, we consider the following two cases of (8): (a) a constant sound speed c_0, i.e., $f=0$, $h=0$, and (b) a linearly increasing sound speed c_0+c_1z, i.e., $f(z)=c_1z$, $h=0$. Figure 1 shows an example of the ray trajectory in the (z, r) space and its Poincaré surface of section (PSS) on the (z,p_z) plane for the case (a). We see that the trajectory of Fig. 1(a) is the same as one ex-

pected for a free particle motion in the billiard system and the PSS of Fig. 1(b) shows the Kolmogorov–Arnold–Moser (KAM) curves[5] corresponding to a quasiperiodic motion. Figure 2 shows the ray trajectory and the PSS for the case (b). We see from Fig. 2(a) that the trajectory curves downward as z increases, whose behavior is a reasonable one expected from the physical viewpoint. The formation of the invariant KAM curve on the PSS of Fig. 2(b) seems to indicate that the ray motion is regular. Thus, the present model describes well a ray motion in the closed space traveling not only in straight lines but also in curved lines, where the ray repeats many reflections with the wall according to the reflection law and draws regularly its trajectory.

In order to study the more realistic situation that the acoustic ray propagates in the inhomogeneous medium, we need to specify the perturbation h in (8). For this purpose, we consider the following two cases: (c) $h(z,r)=c_0\ \alpha z\cos\beta r$, and (d) $h(z,r)=c_0\exp\alpha z\cos\beta r$, where α is the inverse of a height scale, and $\beta=\pi/2B$, B is the perturbation period. The sinusoidal form is assumed to describe the inhomogeneity in the r direction because this is considered one of the simplest forms that make both the analytical and numerical study easy. The form (d) is essentially the same as one widely used in the study of the ray chaos in the underwater acoustics.[4,6] The sign of z in the exponent is different by a simple reason that the depth z of the underwater is defined positive downward opposite to the direction of z in our model. Using the form $f(z)=c_1 z$, we can elucidate how the perturbed profile affects on the ray motion. Figures 3(a) and (b) show some examples of the PSS for the sound-speed profile of (8) with the perturbations (c) and (d), respectively, when ϵ is large. Both cases show qualitatively similar results, i.e., the coexistence with the invariant KAM curves, islands, and stochastic sea in the phase space. We can understand that the sinusoidal variation in the r direction is mainly responsible for the onset of ray chaos and the characteristics of the chaos. On the other hand, as the distinctive shape of the invariant KAM curves is similar to the one of Fig. 2(b), the main structure of the ray motion seems to be controlled by the form of $f(z)$ rather than the z dependence in $h(z,r)$.

All calculations here were done by the fourth-order Runge–Kutta routine of double precision with a time step equal to 10^{-5} s. This value was chosen small enough to maintain the constraint $cp=1$ to a relative error of 10^{-3} after 10^6 iterations. We also calculated the trajectories for many different values of both initial conditions and the parameters in the sound-speed profile and checked that the qualitative results remain almost the same.

Let us briefly comment on our results. It is worthwhile considering how and why the regular trajectory appears for the sound-speed profile of (8) when $\epsilon=0$, as shown in Figs. 1 and 2. As the sound-speed profile contains the background function F_{BG}, there will be a possibility that some trajectories exhibit chaos due to the nonlinear structure of (5) and (6), even when $\epsilon=0$. In order to investigate whether such a chaotic trajectory can indeed appear in this case, we have calculated the Lyapunov exponent[7] (LE) because this quantity is a reliable indicator of chaos. We have confirmed that the LE virtually becomes zero, which indicates that this system is an integral one. This means that background function (7) with very large n can play the role of the square well potential with infinite walls. This picture is perfectly consistent with the standard billiard system. For the case of $\epsilon\neq 0$, we have estimated the spectrum $(\sigma_1,\sigma_2,\sigma_3,\sigma_4)$ of the LE and obtained (15.421, 0.0513, −0.0524, −15.420) for Fig. 3(a) and (24.84, 0.507, −0.509, −24.84) for Fig. 3(b). The positive maximal LE, σ_1, assures us that the system is chaotic. Since the present model is a two-dimensional Hamiltonian system, the spectrum of the LE should hold the symmetry relation,[5] $\Sigma_i\sigma_i=0$, which seems to be satisfied in our calculation.

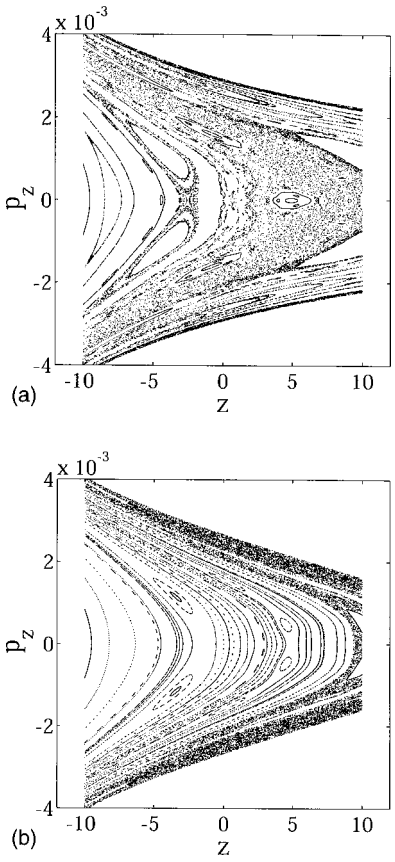


FIG. 3. The PSS for the cases (c) and (d): the perturbed sound speed with $\alpha^{-1}=6$ m, $B=10$ m. The strength of the perturbation is (c) $\varepsilon=0.01$ and (d) $\varepsilon=0.1$. The other parameters are the same as those used in Fig. 2. The PSS is constructed by the ray trajectories with several different initial values.

Next, we would like to point out the feature of the acoustic ray chaos observed in our model. The chaotic trajectories come from the perturbation h. They show the typical feature of chaos, depending on the strength ϵ, such as the formation and destruction of the KAM curves, islands, and chaotic sea, as seen in Fig. 3, which all are observed in many Hamiltonian systems with two degrees of freedom.[5] Since the onset of chaos depends on ϵ and the initial conditions, we will need to systematically study these dependences for the sound-speed profile derived from suitable data.

Finally, we would like to comment on the possible connection between the present model equation and the parabolic ray equation.[4,6] Both equations are based on the same Hamiltonian of (1), but the final forms differ from each other due to different assumptions and constraints to them. The

parabolic ray equation is equivalent to a nonautonomous equation with one degree of freedom (z, p_z) derived from the Hamiltonian after imposing the condition $dr/dt > 0$, where owing to this condition the variable r turns into the time variable.[4] This equation shows a ray chaos when a sound speed depends on time, r, and the ray motion is confined in a finite z region. In our model, with adjusting the background function it is possible to reproduce the main features in their numerical results.[4,6] Since we treat the z and r variables as the space coordinates, we can apply our model to the problem of the ray chaos in a two-dimensional closed region, not only in the underwater acoustics but also in the room acoustics, which must be the advantage of the present formulation.

ACKNOWLEDGMENT

We are very grateful to S. Yoshikawa for useful conversations and comments.

[1] See, e.g., M. V. Berry, ''Regularity and chaos in classical mechanics, illustrated by three deformations of a circular billiard,'' Eur. J. Phys. **2**, 91–102 (1981).
[2] See, e.g., H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1980).
[3] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959); *The Classical Theory of Fields* (Pergamon, Oxford, 1987).
[4] K. B. Smith, M. G. Brown, and F. D. Tappert, ''Ray chaos in underwater acoustics,'' J. Acoust. Soc. Am. **91**, 1939–1949 (1992); K. B. Smith, M. G. Brown, and F. D. Tappert, ''Acoustic ray chaos induced by mesoscale ocean structure,'' *ibid.* **91**, 1950–1959 (1992).
[5] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics* (Springer-Verlag, Berlin, 1992).
[6] D. R. Palmer, M. G. Brown, F. D. Tappert, and H. F. Bezdek, ''Classical chaos in nonseparable wave propagation problems,'' Geophys. Res. Lett. **15**, 569–572 (1988); X. Li, Y. Zhang, and G. Du, ''Influence of perturbations on chaotic behavior of the parabolic ray system,'' J. Acoust. Soc. Am. **105**, 2142–2148 (1999); Z. Jiang, T. A. Pitts, and J. F. Greenleaf, ''Analytic investigation of chaos in a class of parabolic ray systems,'' *ibid.* **101**, 1971–1980 (1997).
[7] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, ''Determining Lyapunov exponents from a time series,'' Physica D **16**, 285–317 (1985).

A mixed finite element method for acoustic wave propagation in moving fluids based on an Eulerian–Lagrangian description

Fabien Treyssède,[a] Gwénaël Gabard, and Mabrouk Ben Tahar
Université de Technologie de Compiègne, Laboratoire Roberval UMR 6066, Secteur Acoustique, BP 20529, 60205 Compiègne Cedex, France

(Received 24 May 2002; revised 5 October 2002; accepted 8 November 2002)

A nonstandard wave equation, established by Galbrun in 1931, is used to study sound propagation in nonuniform flows. Galbrun's equation describes exactly the same physical phenomenon as the linearized Euler's equations (LEE) but is derived from an Eulerian–Lagrangian description and written only in term of the Lagrangian perturbation of the displacement. This equation has interesting properties and may be a good alternative to the LEE: only acoustic displacement is involved (even in nonhomentropic cases), it provides exact expressions of acoustic intensity and energy, and boundary conditions are easily expressed because acoustic displacement whose normal component is continuous appears explicitly. In this paper, Galbrun's equation is solved using a finite element method in the axisymmetric case. With standard finite elements, the direct displacement-based variational formulation gives some corrupted results. Instead, a mixed finite element satisfying the inf-sup condition is proposed to avoid this problem. A first set of results is compared with semianalytical solutions for a straight duct containing a sheared flow (obtained from Pridmore–Brown's equation). A second set of results concerns a more complex duct geometry with a potential flow and is compared to results obtained from a multiple-scale method (which is an adaptation for the incompressible case of Rienstra's recent work). [DOI: 10.1121/1.1534837]

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

I. INTRODUCTION

Propagation of acoustic disturbances in nonuniform flows is a subject of great interest in many practical problems, particularly in transport engineering with automotive exhaust systems, aeronautical turbofan engine inlet ducts, etc. The understanding of this phenomenon is a central feature for the prediction of noise and for designing components that efficiently attenuate sound. In practice, the shape of these components is often complex and flows are not uniform. Thus, the basic equations that describe such a problem must be able to cope with those complexities. Two kinds of formulations are mainly used: the linearized Euler's equations (LEE) and the full-potential formulation. However as discussed further in the following, there exists another wave equation, which is a reformulation of the LEE.

The full-potential formulation is obtained from the LEE by assuming both flow and disturbances irrotationality. Thus, it constitutes a specific case of the general LEE. The corresponding propagation equation is scalar and written only in term of the acoustic velocity potential. This makes its resolution far easier and explains why it is much more widespread in the literature. On the basis of the full-potential equation, some authors studied the effect of flow variation and multidirectionality upon sound propagation in ducts with variable cross sections, using a finite element method (FEM)[1–3] or a boundary element method.[4] Besides, the analysis of pure propagation phenomena has naturally been extended to radiation by many authors by the use of various techniques (FEM combined with a boundary element method[5–9] or a wave-envelope element technique,[10–13] dual reciprocity boundary element method,[14] etc.).

Recently, Rienstra[15] developed an analytical model based on a multiple-scales method to study pure propagation in slowly varying cross-section ducts. Rienstra and Eversman[16] made comparisons between the multiple-scales and FEM solutions that validate both models (though it highlights the limits of a multiple-scales method when reflections or conversions into other modes occur). In this paper, Rienstra's analytical model is used to give a reference solution that validates the FEM implementation of Galbrun's equation.

Because of its relative simplicity, the full-potential equation is a powerful formulation and may be sufficient to study sound propagation in flows. However, its main drawback is that it cannot take into account rotational mean flows. For instance, flow rotationality cannot be neglected when the effect of boundary layer refraction is important (see, for instance, Refs. 17–19) or when a mean flow swirl is present (this typically happens behind a rotor stage—see Ref. 20).

Actually, if the mean flow is rotational, the decomposition of the perturbations in terms of independent acoustic and rotational modes is no longer valid. The LEE must be directly solved. These equations represent a system of five equations and five unknowns, which can be reduced to four if the flow is supposed to be homentropic (pressure and density are then directly related). Given the complexity of the system, some attempts were made to simplify its solution.

[a]Electronic mail: fabien.treyssede@utc.fr

For instance, Nayfeh *et al.*[21] and Uenishi and Myers[22] developed a numerical method for solving acoustic propagation in ducts with variable cross sections and mean sheared flow.

In the late 1970s, Abrahamson[23] and Astley and Eversman[24,25] implemented the direct LEE for axisymmetric and two-dimensional geometries, respectively, using a FEM. Good results were presented for simple cases but unfortunately, their work has not been continued. Astley and Eversman[26] also outlined the existence of spurious modes when using a FEM to formulate the eigenvalue problem for a straight-lined duct with shear flow.

More recently, Golubev and Atassi[27] studied a straight duct containing a mean flow with swirl and showed the coupling that occurs between acoustic and rotational modes. Cooper and Peake[20] extended Golubev's study to slowly varying lined ducts by applying a multiple-scales method. Results showed the influence of the mean flow swirl, i.e., co-rotating modes are always much more damped than those in a nonswirling flow and counter-rotating modes may be amplified.

Furthermore, it must be noticed that the LEE have recently been implemented by several authors (see, for instance, Refs. 28–30) using numerical methods based on finite difference schemes. These references proved the operator ability to adequately propagate sound wave but, so far, numerical applications are still limited to simple geometries because of numerical difficulties and high computer resource requirements.

Through what has been cited previously, it can be seen that the use of the LEE may be essential and inescapable in several practical cases. However, this system is far more complex than the scalar full-potential equation, and it is thus rarely solved for general cases. This paper attempts to develop a general method based on a FEM to solve sound propagation problems with rotational mean flows. Instead of choosing the LEE, a nonstandard wave equation originally established by Galbrun[31] in the early 1930s is considered. As explained in Sec. II, this equation is derived from an Eulerian–Lagrangian description and is an exact reformulation of the LEE. Thus, Galbrun's equation *a priori* describes exactly the same physical phenomenon. Its main feature is that it constitutes a second-order linear partial differential equation written only in terms of the displacement perturbation.

Although only few works deal with this equation, it may be an interesting alternative to the LEE. Only acoustic displacement is involved (even in nonhomentropic cases), which yields a gain of one to two unknowns compared to the LEE; it also provides exact expressions of acoustic intensity and energy; besides, boundary conditions are easily expressed because acoustic displacement (whose normal component is generally continuous at any interface between two media) appears explicitly, which avoid the somewhat difficult use of Myers' condition.[32]

In 1985, Poirée[33] detailed the Eulerian–Lagrangian description in order to derive Galbrun's equation and its extension to nonlinear problems. He also derived some general continuity conditions valid at any straight interface. Godin[34] independently obtained Galbrun's equation in a quite different way. In particular, he derived exact expressions of acoustic energy and intensity, and boundary conditions for free, rigid, and absorbing surfaces in terms of the displacement. Ben Tahar and Goy[35] developed a variational formulation to study vibroacoustic problems with mean flow.

Recently, Peyret and Elias[36] proposed a direct displacement-based formulation of Galbrun's equation solved using a FEM. They also derived the same energy conservation law as Godin but with a different approach. Bonnet *et al.*[37] pointed out the fact that the direct displacement-based formulation associated with Galbrun's equations does not necessarily converge with standard finite elements and proposed a method to regularize the variational formulation in the case of a uniform flow. These last two references are discussed in detail later (see Sec. III).

In this paper, a mixed variational formulation based on the pressure-displacement variables is presented in order to avoid some spurious solutions. Though the overall method is quite general, finite element discretization and numerical results are presented for the axisymmetric case. A first set of results consists in comparing FEM with semi-analytical solutions for a straight duct containing a sheared flow (obtained from the Pridmore–Brown equation). A second set of results concerns a more complex duct geometry with a potential flow and is compared to results obtained from a multiple-scale method (which is an adaptation of Rienstra's work for the incompressible case—see Ref. 15).

II. THEORY

Compared to the classical Eulerian description, the Eulerian–Lagrangian description is not usual. This section briefly recalls the latter before giving Galbrun's equation and the associated expression of acoustic intensity. This section does not give details about calculations but reviews can be found in Refs. 31, 33, and 34.

A. The Eulerian–Lagrangian description

In the Eulerian description, perturbations are Eulerian and expressed with Eulerian variables. This description is implicitly used when Euler's equations are directly linearized. The Eulerian–Lagrangian description consists in considering Lagrangian perturbations of quantities expressed in terms of Eulerian variables. In order to give a physical comprehension of what Lagrangian and Eulerian perturbations mean, the perturbed (or total) and nonperturbed (or mean flow) states have to be described first.

Define the geometrical position $\mathbf{x}$ of a given particle in the mean flow configuration and its position $\mathbf{y}$ in the perturbed configuration. Then, if $\mathbf{w}^L$ denotes the displacement perturbation of this particle, $\mathbf{x}$ and $\mathbf{y}$ are related by

$$\mathbf{y}=\mathbf{x}+\varepsilon\mathbf{w}^L. \tag{2.1}$$

In the remainder of this paper, Ψ represents any physical quantity (tensor of arbitrary order) and the subscript 0 is used to distinguish mean flow quantities from their total (or perturbed) counterpart. Then, two kinds of perturbations can be defined:

$$\varepsilon\Psi^E=\Psi(\mathbf{y},t)-\Psi_0(\mathbf{y},t),$$
$$\varepsilon\Psi^L=\Psi(\mathbf{y},t)-\Psi_0(\mathbf{x},t). \tag{2.2}$$

Superscripts E and L denote, respectively, Eulerian and Lagrangian perturbations. From these definitions, it can be seen that Eulerian perturbations are associated with the same geometrical point but not the same particle, whereas Lagrangian perturbations are associated with the same particle. Using Eq. (2.1) into Eq. (2.2), a Taylor expansion up to the first order gives the fundamental relationship between Eulerian and Lagrangian perturbations:

$$\Psi^L=\Psi^E+\mathbf{w}^L\cdot\boldsymbol{\nabla}\Psi_0. \tag{2.3}$$

As stated earlier, Ψ represents any physical quantity and Eq. (2.3) holds for pressure fluctuations, density, velocity, etc.

B. Galbrun's equation

Equation (2.3) is now applied to pressure, density, and velocity perturbations. This yields:

$$p^L=p^E+\mathbf{w}^L\cdot\boldsymbol{\nabla}p_0,\quad \rho^L=\rho^E+\mathbf{w}^L\cdot\boldsymbol{\nabla}\rho_0,$$
$$\frac{d\mathbf{w}^L}{dt}=\mathbf{v}^E+\mathbf{w}^L\cdot\boldsymbol{\nabla}\mathbf{v_0}, \tag{2.4}$$

where $d/dt=\partial/\partial t+\mathbf{v_0}\cdot\boldsymbol{\nabla}$ is the material derivative. Equation (2.4) allows one to express every Eulerian perturbation in terms of Lagrangian perturbations. Replacing every fluctuation by its Lagrangian counterpart into the LEE leads (after tedious calculations) to Galbrun's equation:

$$\rho_0\frac{d^2\mathbf{w}^L}{dt^2}-\nabla(\rho_0c_0^2\boldsymbol{\nabla}\cdot\mathbf{w}^L)+(\boldsymbol{\nabla}\cdot\mathbf{w}^L)\boldsymbol{\nabla}p_0-{}^T\boldsymbol{\nabla}\mathbf{w}^L\cdot\boldsymbol{\nabla}p_0=0 \tag{2.5}$$

and the Lagrangian density perturbation is explicitly given by

$$\rho^L=-\rho_0\boldsymbol{\nabla}\cdot\mathbf{w}^L. \tag{2.6}$$

Equation (2.6) constitutes the mass continuity equation obtained from a Lagrangian–Eulerian description. It simply states that density fluctuations are balanced by dilatation fluctuations.

In order to derive Galbrun's equation, perfect fluid and isentropic assumptions have implicitly been made by starting from the LEE. For Lagrangian perturbations, the isentropic assumption leads to the well-known pressure–density relationship:

$$p^L=\rho^Lc_0^2, \tag{2.7}$$

which yields the following explicit equation for the Lagrangian pressure:

$$p^L=-\rho_0c_0^2\boldsymbol{\nabla}\cdot\mathbf{w}^L. \tag{2.8}$$

Some important remarks should now be addressed. Unlike Eulerian perturbations, Eq. (2.7) remains valid even for inhomogeneous media and/or for nonhomentropic flows. In these situations, solving the LEE would require an additional equation, that is the linearized energy equation (see, for instance, Ref. 38). In fact, for Lagrangian perturbations, Eq. (2.7) still holds because the equation of state applies for a given particle so that the thermodynamical system is closed.

Thus, Galbrun's equation obviously provides a first advantage (compared to the LEE) because this equation is expressed with the Lagrangian displacement as a unique unknown, which yields a gain of one to two variables.

Another asset is the existence (but not uniqueness) of a Lagrangian density associated with Galbrun's equation (see Ref. 36). This yields an exact energy conservation law and exact expressions for the acoustic energy and intensity. These expressions can be found in Refs. 34 and 36. In particular, the acoustic intensity is given by

$$\mathbf{i}=\rho_0\left(\frac{\partial\mathbf{w}^L}{\partial t}\cdot\frac{d\mathbf{w}^L}{dt}\right)\mathbf{v_0}+(p^L-\mathbf{w}^L\cdot\boldsymbol{\nabla}p_0)\frac{\partial\mathbf{w}^L}{\partial t}. \tag{2.9}$$

It can be noticed that the Lagrangian displacement inevitably appears in this expression, which likely explains why no exact formulation for the intensity has been found based on a pure Eulerian description.

III. NUMERICAL METHOD

In this section, Galbrun's equation is solved using a FEM. From now on, fluctuations are assumed to have an $e^{-i\omega t}$ time dependence. Besides, the mean flow is steady and the mean pressure is supposed to be constant for simplicity [dropping the last two terms of Eq. (2.5)]. This assumption is not valid for aeroacoustic problems but our purpose is to solve Galbrun's equation with a FEM (as outlined by Peyret and Elias,[36] terms with p_0 do not present a major interest from a numerical point of view). Under these assumptions, Eq. (2.5) becomes

$$-\rho_0\omega^2\mathbf{w}^L-2i\omega\rho_0\mathbf{v_0}\cdot\boldsymbol{\nabla}\mathbf{w}^L+\rho_0\mathbf{v_0}\cdot\boldsymbol{\nabla}(\mathbf{v_0}\cdot\boldsymbol{\nabla}\mathbf{w}^L)$$
$$-\rho_0c_0^2\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{w}^L)=0. \tag{3.1}$$

Section III A gives a brief review of the numerical difficulties of the direct displacement-based formulation associated with Eq. (3.1). In Sec. III B, a displacement-pressure-based mixed formulation is proposed to overcome these difficulties. Section III C deals with boundary conditions and Sec. III D gives some important details about the finite element discretization of the mixed formulation.

A. Displacement-based formulation

Equation (3.1) is multiplied by a trial field $\mathbf{w}^*$ and integrated over the domain Ω. Then, integrating by part the last two terms (which imply second-order derivatives) and half the second term yields the following direct displacement-based variational formulation:

$$\int_\Omega \rho_0 c_0^2 (\nabla\cdot\mathbf{w}^*)(\nabla\cdot\mathbf{w}^L)d\Omega - \omega^2 \int_\Omega \rho_0 \mathbf{w}^*\cdot\mathbf{w}^L d\Omega$$
$$- i\omega \int_\Omega \rho_0 \mathbf{w}^*\cdot(\mathbf{v_0}\cdot\nabla\mathbf{w}^L)d\Omega$$
$$+ i\omega \int_\Omega \rho_0 (\mathbf{v_0}\cdot\nabla\mathbf{w}^*)\cdot\mathbf{w}^L d\Omega$$
$$- \int_\Omega \rho_0 (\mathbf{v_0}\cdot\nabla\mathbf{w}^*)\cdot(\mathbf{v_0}\cdot\nabla\mathbf{w}^L)d\Omega$$
$$+ \int_S \mathbf{w}^*\cdot\left\{\rho_0(\mathbf{v_0}\cdot\mathbf{n})\frac{d\mathbf{w}^L}{dt} + p^L\cdot\mathbf{n}\right\}dS = 0, \quad \forall \mathbf{w}^*. \tag{3.2}$$

It can be noticed that Eq. (3.2) is presented in such a way that the first line corresponds to the no-flow case and the second, third, and fourth ones to the presence of mean flow (the last line is a boundary integral).

This formulation is used in Ref. 36. However, when standard finite elements are implemented to discretize the formulation, solutions are generally corrupted, even in the no-flow case. This phenomenon is purely numerical. An example of spurious solution is given in Sec. IV A.

Bonnet-Ben Dhia *et al.*[37] have recently proposed a regularized formulation of Galbrun's equation in the uniform flow case [with this method, some specific terms are added in the formulation (3.2)]. A good convergence is obtained but limitations of the method arise for the generalization to arbitrary mean flows.

The no-flow case was first studied in the 1970s when considering vibrations of coupled fluid-structure systems (see, for instance, Ref. 39) and was proved to exhibit spurious circulation modes with nonzero frequencies. Basically, this phenomenon is due to a bad accuracy of the divergence and curl (calculated from derivatives of displacements), which in turn affects the displacement prediction itself. In order to cope with this numerical phenomenon, several methods have been proposed, such as the penalty method,[39] edge finite element,[40] and mixed finite element methods.[41,42]

In the presence of mean flow, the penalty method and edge finite element method cannot be directly applied because the displacement field is generally no more irrotational. Thus, the method chosen in this paper is naturally based on a mixed finite element formulation. To conclude this section, one emphasizes that the overall problem in the no-flow case is typically analogous to incompressible elasticity or fluid[43] and electromagnetics,[44] and is often referred to as ''locking'' in the literature.

B. Mixed formulation

A mixed variational formulation based on pressure-displacement variables is now derived. Although pressure is now an explicit unknown of Galbrun's equation (as in the LEE), the efficiency of such a formulation to prevent corruption by spurious solutions has already been demonstrated by Wang and Bathe[42] in the no-flow case when considering fluid-structure interactions. It was also extensively and rigorously analyzed by Brezzi and Fortin[45] in a general way (see also Bathe[43] for a more engineering-oriented approach applied to incompressible elasticity).

In the presence of arbitrary mean flow, an analogous mixed formulation of Ref. 42 can be obtained by replacing the last term of Eq. (3.1) with the pressure gradient (use of Eq. (2.8) is made). It yields the following system:

$$\begin{aligned} &-\rho_0\omega^2\mathbf{w}^L - 2i\omega\rho_0\mathbf{v_0}\cdot\nabla\mathbf{w}^L + \rho_0\mathbf{v_0}\cdot\nabla(\mathbf{v_0}\cdot\nabla\mathbf{w}^L) + \nabla p^L = 0, \\ &p^L + \rho_0 c_0^2 \nabla\cdot\mathbf{w}^L = 0. \end{aligned} \tag{3.3}$$

Multiplying both equations by trial fields $\mathbf{w}^*$ and p^*, respectively, integrating over the domain Ω and then by parts gives the following mixed variational formulation:

$$-\int_\Omega \frac{1}{\rho_0 c_0^2} p^* p^L d\Omega + \int_\Omega \nabla p^*\cdot\mathbf{w}^L d\Omega + \int_\Omega \mathbf{w}^*\cdot\nabla p^L d\Omega$$
$$- \omega^2 \int_\Omega \rho_0 \mathbf{w}^*\cdot\mathbf{w}^L d\Omega - i\omega \int_\Omega \rho_0 \mathbf{w}^*\cdot(\mathbf{v_0}\cdot\nabla\mathbf{w}^L)d\Omega$$
$$+ i\omega \int_\Omega \rho_0 (\mathbf{v_0}\cdot\nabla\mathbf{w}^*)\cdot\mathbf{w}^L d\Omega$$
$$- \int_\Omega \rho_0 (\mathbf{v_0}\cdot\nabla\mathbf{w}^*)\cdot(\mathbf{v_0}\cdot\nabla\mathbf{w}^L)d\Omega$$
$$+ \int_S \mathbf{w}^*\cdot\left\{\rho_0(\mathbf{v_0}\cdot\mathbf{n})\frac{d\mathbf{w}^L}{dt}\right\}dS - \int_S p^*(\mathbf{w}^L\cdot\mathbf{n})dS$$
$$= 0 \quad \forall\{\mathbf{w}^*, p^*\}, \tag{3.4}$$

where S is the surface enclosing the acoustic domain Ω and $\mathbf{n}$ is the outward normal.

As in Eq. (3.2), the first line of the mixed formulation (3.4) represents the no-flow operators. These operators are almost identical to those used by Wang and Bathe[42] in their mixed formulation. The only slight difference is that one has chosen to integrate by parts the second term of the second equation of system (3.3) (instead of the last term of the first equation) in order to let the normal displacement appear explicitly at the boundary.

Normal displacement at walls can thus be easily imposed (the second surface integral of Eq. (3.4) simply disappears for perfectly rigid walls). Besides, for fluid–structure interactions (not considered in this paper), normal displacement continuity could also be easily imposed by replacing the fluid normal displacement with the structure normal displacement.

In order to show the efficiency of a mixed formulation, a comparison with the displacement-based formulation is given in Sec. IV A.

C. Boundary conditions

Boundary conditions associated with Galbrun's equation must be carefully applied. Figure 1 represents a typical problem of propagation inside a duct carrying flow. Two types of

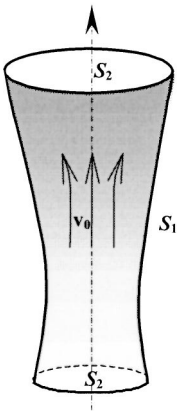


FIG. 1. Typical geometry of a duct carrying flow (S_1 denotes the wall boundaries and S_2 the inlet and outlet boundaries).

boundary conditions must be distinguished: those imposed at walls (boundary S_1) and those imposed inside the fluid (boundary S_2).

At walls, the first surface integral of Eq. (3.4) always vanishes because the mean flow normal velocity is zero. For rigid walls, normal displacement also equals zero and the second integral also vanishes. For an absorbing wall, the adequate boundary condition is obtained from the normal displacement fluid–wall continuity. Given the wall impedance Z, this condition is simply given by the following relationship:

$$p^L|_{S_1} = -i\omega Z \mathbf{w}^L \cdot \mathbf{n}, \tag{3.5}$$

which is simply applied by replacing the normal displacement in the second boundary integral. The fact that displacement is an explicit variable in the proposed formulation makes the impedance condition (3.5) simpler to implement than in the LEE case, which would have required the use of Myer's condition.[32]

When an arbitrary mean flow is present, the displacement field may not be irrotational. Thus, it is obvious that a fixed pressure inside the fluid (on S_2) is not sufficient to determine a unique solution. For instance, applying pressure (which represents the displacement divergence from Eq. (2.8)) at the duct inlet and outlet means that the rotational part of the displacement is left free. Consequently, one must impose the total displacement field everywhere a boundary condition is required inside the fluid (typically at the duct inlet–outlet). This condition is explicitly given by

$$\mathbf{w}^L|_{S_2} = \overline{\mathbf{w}} \tag{3.6}$$

and is directly enforced at nodes as a constraint in the FEM model. This makes the first surface integrals in Eq. (3.4) vanish because $\mathbf{w}^* \equiv \delta\mathbf{w} = 0$ on S_2 (i.e., forced boundary condition). Note that in many instances, it is difficult to specify the particle displacement on both surfaces labeled S_2 because only the incident component is known. A boundary condition based on a multimodal decomposition technique overcomes this difficulty.

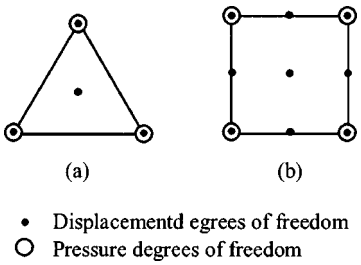


FIG. 2. Examples of 2D elements satisfying the inf-sup condition. (a) Element "4-3c" (used in this paper), which may be referred to as the "$P_1^+ - P_1$" or "MINI" element in the literature. (b) Example of higher order element, with 9 nodes for the displacement and 4 nodes for the pressure (element "9-4c").

As a side remark, in the no-flow case, a surface with fixed pressure inside the fluid is sufficient to uniquely determine the solution because displacement is implicitly irrotational.

D. Finite element discretization

It has been proven that interpolations for displacement and pressure must be adequately chosen. The choice of a mixed formulation is not sufficient to avoid locking, and one also has to choose appropriate interpolation functions. Based on Refs. 43 and 45, this section gives some details about what kind of finite elements has to be used.

If a bad choice of interpolating functions is made, the element may lock and/or give some spurious pressure modes. A criterion that ensures convergence and stability of the finite element is given by the so-called inf-sup condition. Though not necessary, this condition is a strong guaranty of reliability. Details about the inf-sup condition are not given in this paper but can be found in Refs. 43 and 45. For a more practical use, a numerical test of this condition has also been developed by Chapelle and Bathe.[46] One example of a triangular two-dimensional (2D) element satisfying the inf-sup condition is given in Fig. 2. This element may be referred to as the "$P_1^+ - P_1$", "4-3c" or "MINI" element[43,45] and is the one used in this paper. An example of higher order element is also given in Fig. 2 (element "9-4c"). Note that some other 2D (or three-dimensional) elements can be found in Refs. 43 and 46.

In the no-flow case, elements that satisfy the inf-sup condition have successfully been implemented by Wang and Bathe.[42] The originality of Galbrun's equation is that the pressure-displacement relationship given by Eq. (2.8) is not altered by the presence of flow and is strictly identical to the no-flow case. This enables one to directly apply the inf-sup condition to a Galbrun-based formulation. Thus, under the assumption that the additional operators introduced by the presence of flow (terms in $\mathbf{v_0}\cdot\nabla$) does not alter convergence properties of elements satisfying the inf-sup condition, it is expected that the proposed mixed FEM for solving Galbrun's equation is robust.

When flow is present, it should be noticed that the mass continuity equation obtained from the LEE implies material derivatives and no equation similar to Eq. (2.8) can be obtained. The inf-sup condition cannot then be directly applied

when the LEE are considered. Thus, given the current knowledge about mixed formulations, Galbrun's equation seems to be more interesting from a numerical point of view.

The general variational formulation (3.4) is now restricted to axisymmetric geometries. Introducing the cylindrical coordinates, the following θ dependence is set:

$$\begin{aligned} \mathbf{w}^L(r,\theta,z) &= \mathbf{w}^L(r,z)e^{-im\theta}, \\ p^L(r,\theta,z) &= p^L(r,z)e^{-im\theta}, \end{aligned} \tag{3.7}$$

where the angular mode number m is a parameter of the solution. The weighting functions are taken as

$$\begin{aligned} \mathbf{w}^*(r,\theta,z) &= \mathbf{w}^*(r,z)e^{+im\theta}, \\ p^*(r,\theta,z) &= p^*(r,z)e^{+im\theta}. \end{aligned} \tag{3.8}$$

The (2D) acoustic domain Ω is chosen to be meshed with 4-3c elements. On the reference element, displacement and pressure variables are thus interpolated as follows:

$$\begin{aligned} \mathbf{w}^L(u,v) &= (1-u-v)\mathbf{w_1} + u\mathbf{w}_2 + v\mathbf{w}_3 + (1-u-v)uv\mathbf{a}, \\ p^L(u,v) &= (1-u-v)p_1 + up_2 + vp_3, \end{aligned} \tag{3.9}$$

where the subscripts i ($i=1,2,3$) denote the node number. It can be seen for the displacement, that the standard linear interpolation is enriched by introducing a generalized variable **a**. The term $(1-u-v)uv$ represents a bubble function: this polynomial is null on the three side of the triangle and thus maintains the compatibility (C^0 continuity).

Then, elements have four degrees of freedom per node, plus three internal (three for each component of $\mathbf{w}^L$). However, these internal degrees of freedom can be condensed out before the element is assembled, which is attractive from a computational point of view.

After assembling and applying boundary conditions, the global discretized variational formulation yields the following algebraic system:

$$\mathbf{K}_r\mathbf{w}_r = \mathbf{f}_r. \tag{3.10}$$

$\mathbf{K}_r$ is a ω-dependent complex band matrix, unsymmetrical when flow is present. A sparse storage is chosen. For a fixed ω, the unknown nodal displacement vector $\mathbf{w}_r$ is finally obtained by using a LU decomposition.

IV. RESULTS

In Sec. IV, the FEM numerical method is validated with two semianalytical models. The first model corresponds to the well-known Pridmore–Brown equation, the second to Rienstra's multiple scale approximation (in the incompressible mean flow case in order to fit with the assumption of the FEM model—see Sec. III). Both models represent the propagation of a given (m,n) mode in an infinite duct (m and n denote, respectively, the angular and radial mode number).

They may be considered as complementary. The Pridmore–Brown equation deals with a simple straight duct but a possibly sheared mean flow. Boundary layer effects upon propagation can thus be considered. In Rienstra's multiple-scales method, the mean flow must be potential but the duct is slowly varying, which permits one to study more complex geometries.

Unlike the Pridmore–Brown equation, Rienstra's model constitutes an approximation. It cannot be exact because of modal reflection and scattering that may occur in a varying duct. These limitations of a multiple-scales method have been highlighted by Rienstra and Eversman[16] and are also demonstrated in this paper.

In the following, iso-pressure contours are given in modulus for all plots. Units are chosen to be in pascals in order not to minimize errors. The averaged intensity vector may also be given on the $\theta=0°$ plane. Mean flow velocities are defined in Mach number (M). Propagation and flow directions are also sketched in order to explicitly show if wave propagation is upstream or downstream. Besides, typical values of $\rho_0=1.2\ \mathrm{kg\,m^{-3}}$ and $c_0=340\ \mathrm{m\,s^{-1}}$ are used.

Test cases sweep a nondimensional frequency range up to about $kR=20$ and the duct geometry is always meshed with a $\lambda/10$ finite element length. This is the estimated criterion for a satisfactory convergence but a $\lambda/8$ criterion may be sufficient to give good results. These criteria are yet modulated by a $\sqrt{1-M^2}$ factor due to the Doppler effect when flow is present.

A. Validation for straight ducts (Pridmore–Brown equation)

For FEM computations in this section, the methodology is as follows. The Lagrangian displacement is calculated from the Pridmore–Brown model and is then imposed at the duct inlet of the FEM model (in the remaining, the terms ''inlet'' and ''outlet'' are used from an acoustical point of view—see the direction of propagation upon each figure—and not from a mean flow point of view). A modal nonreflective boundary condition is preferred at the outlet, which is less constraining (in particular, the phase and amplitude are left free). Hence, the mode being enforced at the inlet/outlet boundaries, the FEM solution inside the duct is computed and compared to the semianalytical one.

The nonreflective boundary condition is inspired from Peyret and Elias.[36] It simulates the one-way propagation and applies only for a given (m,n) mode. In classical acoustics (no-flow case), one usually imposes on the cross section an anechoic termination based on the wave impedance (i.e., the ratio of the pressure and the normal acoustic velocity). Here, an additional condition is required with the displacement variable in order to determine a unique solution. The velocity and the normal displacement which appear in the boundary integral of Eq. (3.4) are thus replaced with the following:

$$\rho_0(\mathbf{v_0}\cdot\mathbf{n})\frac{d\mathbf{w}^L}{dt} = \mathbf{Z}_{\mathrm{nr}}\mathbf{w}^L, \quad \mathbf{w}^L\cdot\mathbf{n} = -\frac{1}{i\omega Z_{\mathrm{nr}}}p^L. \tag{4.1}$$

Z_{nr} is the standard modal nonreflecting impedance and $\mathbf{Z}_{\mathrm{nr}}$ can be viewed as a modal nonreflective matrix impedance. In the Pridmore–Brown model, material derivatives are simply given by $d/dt=-i(\omega-v_0k_z)$ because an $e^{i(k_zz-m\theta-\omega t)}$ dependence is chosen for the modal acoustic variables. This

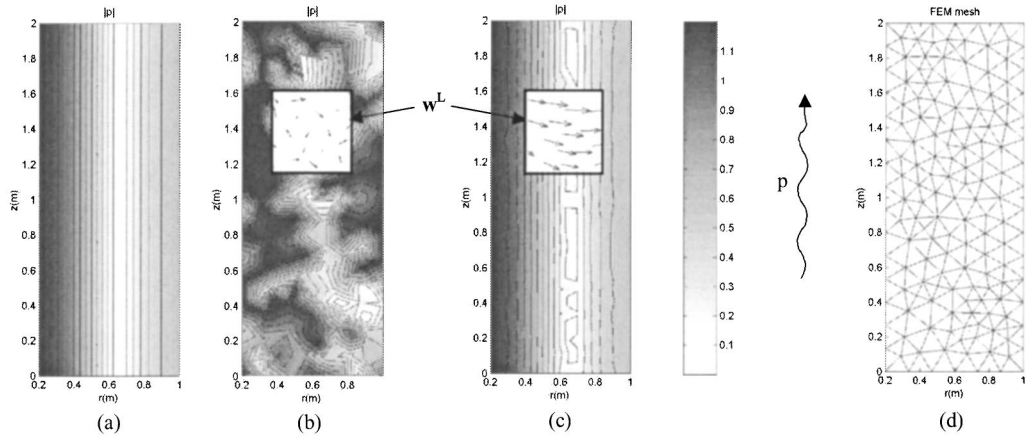


FIG. 3. Pressure modulus in pascals of the (0,1) mode at $f=250$ Hz (no flow and rigid walls): (a) semianalytical solution, (b) displacement-based FEM solution, (c) mixed FEM solution, (d) FEM mesh.

yields the following explicit expressions for the modal impedances defined by Eq. (4.1):

$$\mathbf{Z}_{\mathrm{nr}}=-i\rho_0 v_0(\omega-v_0k_z)\mathbf{I}, \quad Z_{\mathrm{nr}}=\rho_0\frac{(\omega-v_0k_z)^2}{\omega k_z}, \quad (4.2)$$

where v_0 is the axial mean flow velocity and k_z is the outlet modal axial wave number, which is part of the semianalytical solution. **I** denotes the identity matrix.

In this section, an annular straight duct is considered. The inner and outer radius are, respectively, 0.2 and 1 m. Geometry with a typical mesh is shown by Fig. 3.

The first test case, given in Fig. 3, shows the efficiency of a mixed FEM compared to a displacement-based one. A comparison between the semianalytical, displacement-based FEM and mixed FEM solutions is given for the pressure modulus. The example concerns a (0,1) mode at $f=250$ Hz propagating in a hard wall duct without flow. Solution obtained with a displacement-based formulation is totally corrupted by rotational fields. These spurious solutions are small-scale artifacts which are trapped within the model grid. Unfortunately, this numerical problem does not disappear at all when the mesh is refined. For a mixed FEM, results are

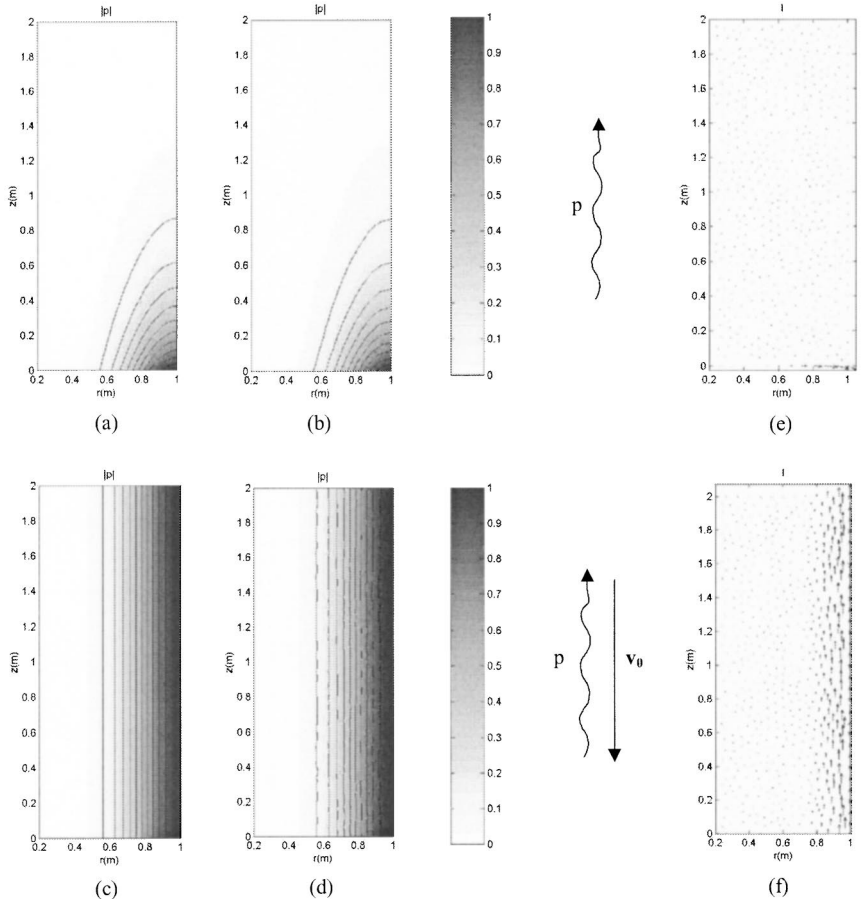


FIG. 4. Pressure modulus in pascals of the (8,0) mode at $f=500$ Hz (rigid walls). (a) Pridmore-Brown and (b) mixed FEM solutions with $M=0.0$. (c) Pridmore–Brown and (d) mixed FEM solutions with $M=-0.4$. (e) and (f) intensity vector (computed from the FEM model) for $M=0.0$ and $M=-0.4$, respectively.

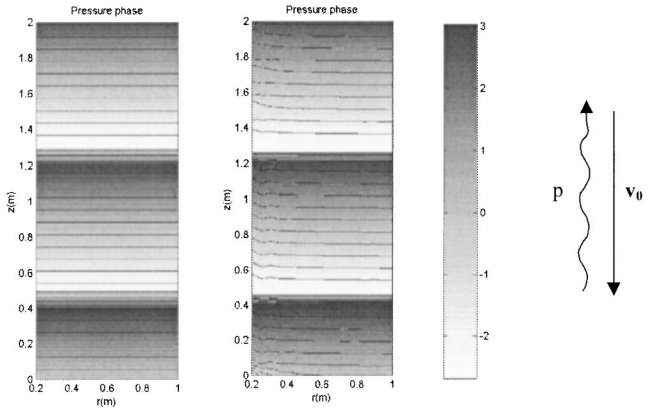


FIG. 5. Pressure phase in radians of the (8,0) mode at $f=500$ Hz with $M=-0.4$ (rigid walls). (a) Pridmore–Brown and (b) mixed FEM solutions.

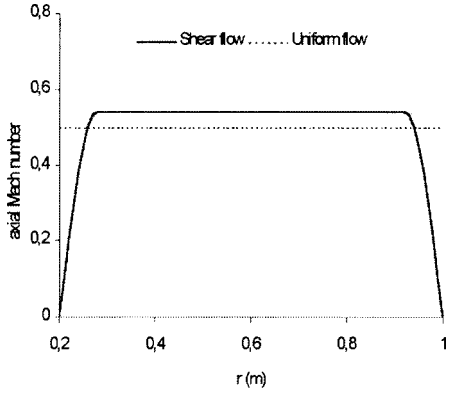


FIG. 7. Axial velocity profiles for the shear and uniform mean flows. Both flows have the same mass flow rates (the mean Mach number is 0.5).

clearly in good agreement with the semianalytical solution.

The second test case (Fig. 4) concerns an (8,0) mode propagating at $f=500$ Hz in a perfectly rigid wall duct. This mode is calculated for $M=0$ and $M=-0.4$ (upstream propagation). Agreement between Pridmore–Brown and FEM solutions is perfect for both Mach numbers. This test case aims at pointing out one of the effects of convection upon cutoff frequencies, say a decrease by a $\sqrt{1-M^2}$ factor. Figure 4 clearly shows that without flow, the (8,0) mode is cut-off whereas for $M=-0.4$, it becomes cut-on. In fact, for the duct dimensions, the (8,0) mode has an exact cut-off frequency of 522.0 Hz without flow, decreased to 478.5 Hz at $M=-0.4$. This frequency is lower than 500 Hz, which explains why the mode fully propagates along the duct. Examining the intensity plot for both Machs number gives directly the nature of modes: unlike at $M=-0.4$, the acoustic intensity vector is null at $M=0.0$, which proves that the mode is cut-off (energy does not propagate). For the $M=-0.4$ case, a comparison in terms of the pressure phase is also given by Fig. 5, where a good agreement between solutions can be observed.

The third test case (Fig. 6) exhibits a (10,1) mode at $f=1000$ Hz. Walls are lined and the impedance value is $Z=2040(1-i)$ for both inner and outer walls. The cross-section averaged Mach number is $M=-0.5$ (upstream propagation). In this example, a comparison between uniform and sheared flow is given in order to demonstrate the capability of the FEM approach to take into account refraction phenomena. The shear flow is arbitrarily chosen to have a boundary layer thickness of 10% ($\delta=0.08$ m), with the same mass flow rate as the uniform profile (see Fig. 7 for the mean flow velocity profiles). This thickness is not realistic but voluntarily exaggerated in order to illustrate refraction. Note that flows with a boundary layer are obviously rotational and cannot be considered with the full-potential propagation equation.

As seen in Fig. 6, Pridmore–Brown and FEM solutions show satisfactory agreement. In particular, modal shape and attenuation are conserved in the FEM model. Results show a strong difference in amplitude between the uniform and sheared cases. In fact, for an upstream propagation, the mean flow velocity gradient due to the presence of the boundary layer tends to refract waves toward the center, thus yielding a weaker attenuation than in the uniform case. (This is the opposite for a downstream propagation, for which waves are refracted toward walls—see, for instance, Ref. 17.) Attenuation coefficient values can directly be obtained from the semianalytical model. These coefficients are 5.6 and 0.9 dB m^{-1}, respectively, for the uniform and sheared cases.

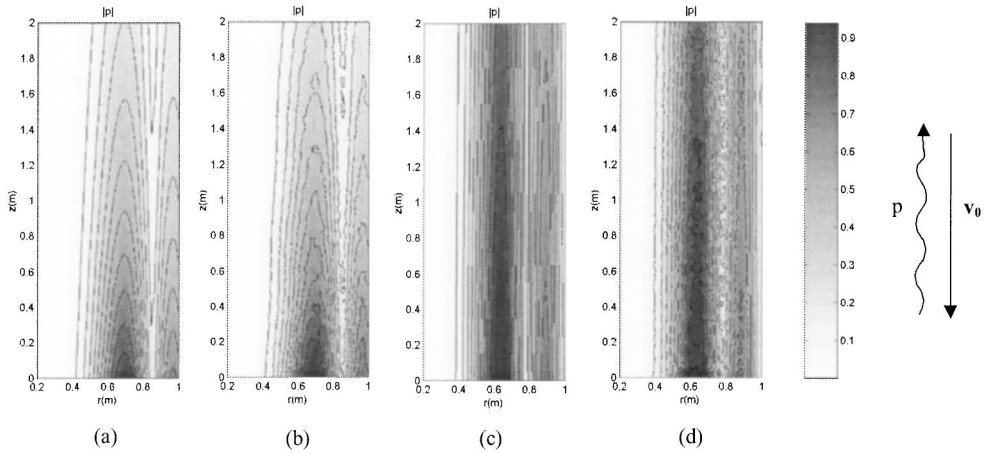


FIG. 6. Pressure modulus in pascals of the (10,1) mode at $f=1000$ Hz and $M=-0.5$ with lined walls ($Z=2040$-$2040i$). (a) Pridmore–Brown and (b) mixed FEM solutions for a uniform flow. (c) Pridmore–Brown and (d) mixed FEM solutions with a boundary layer thickness of 10%.

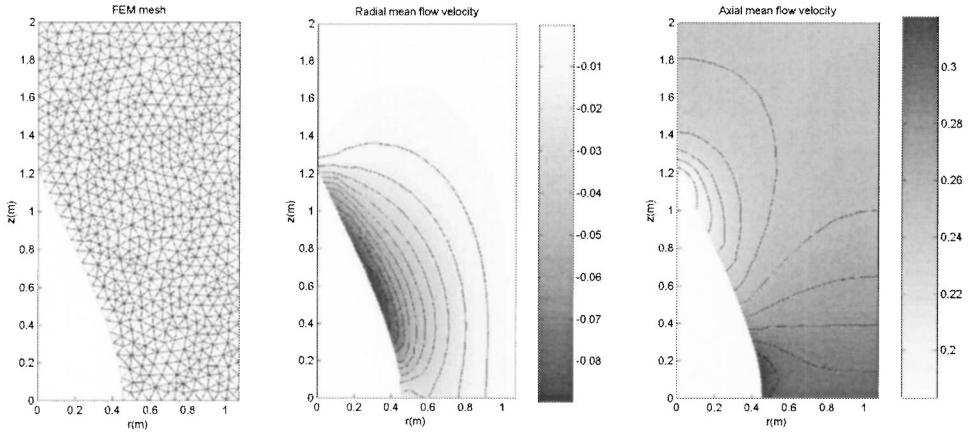


FIG. 8. Rienstra's duct geometry with a straight outer wall: (a) FEM mesh, (b) radial and (c) axial potential mean flow velocities (in Mach number) computed from a FEM model of Laplace's equation.

Thus, for the duct being considered, an error of about 9 dB is made at the outlet when a uniform profile is supposed.

B. Validation for varying ducts (Rienstra's model)

The test geometry taken is now varying (but the flow is restricted to be potential). This geometry is the same as in Rienstra's papers[15,16] and is representative of a turbofan aircraft engine. It includes a circular-to-annular transition (a central body is thus present).

There are differences between the multiple scale method and the FEM. The FEM formulation admits the propagation of many modes (reflection and scattering are integral parts of the solution). On the contrary, the multiple scale approximation lies in supposing that a single given mode is propagating in a single direction inside a duct. Hence, this kind of approximation neglects reflection and scattering into other modes, as clearly demonstrated by Rienstra and Eversman's study.[16] Reflection and scattering limitations of Rienstra's model are also both highlighted in this section by choosing adequate test cases.

The procedure used for FEM computation is not the same as previously exposed in Sec. IV A. The previous mono-modal nonreflecting boundary condition cannot be used when reflections or scattering into other modes become significant. Instead, a modal decomposition technique such as in Ref. 16 is used. This technique consists in recasting the acoustic variables at the duct inlet and outlet via an eigenmode expansion. For FEM calculations, the method is as follows. On the inlet plane, the complex amplitude of the appropriate mode is specified via a forced boundary condition. On the outer plane, reflected mode amplitudes are set to zero, imposing a nonreflecting boundary condition. Inlet reflected and outlet transmitted mode amplitudes are unknown and part of the solution. In this paper, ten radial modes have been used for the expansion (given the cut-off frequencies of higher order modes, this is sufficient for the test cases presented in the following). Note that the eigenmode expansion must be done with hard walls for orthogonality (soft wall modes are not orthogonal). Thus, for test cases with lined walls, the FEM geometry has been extended up to 0.5 m both at the inlet and outlet by adding pieces of straight hard wall ducts (not shown in the figures).

In the FEM model, the potential mean flow is first computed via a FEM solving Laplace's equation and then used for solving Galbrun's equation. Results presented in this section give comparisons between the solutions obtained by Rienstra's multiple-scales method (mono-modal) and the FEM model proposed in the present paper (FEM solutions of Rienstra and Eversman are not shown).

The first test case is depicted in Fig. 8, where geometry, mesh, and mean flow are presented. At the inlet and outlet, the axial local Mach number values are $M=0.3$ and 0.25, respectively. It must be observed that the outer wall is straight, which constitutes a slight difference from Rienstra's geometry (considered in the second test case). This essentially aims at minimizing reflections due to the outer wall.

In this first set of results (Fig. 9), a downstream (1,1) mode propagating in a rigid wall duct is considered. Computations are achieved for three frequencies. At 300 and 360 Hz, a very good agreement is obtained between FEM and multiple-scales solutions, which validates the FEM code. At these frequencies, the only cut-on modes with $m=1$ are (1,0) and (1,1). Other cut-on modes with $m\neq 1$ are not considered by the FEM model because m is a fixed parameter in the code. Consequently, the good convergence between both models demonstrates that scattering into the (1,0) mode as well as reflection are almost negligible.

However, when the frequency $f=420$ Hz is reached, a strong difference is observed. This disagreement is likely explained by partial scattering into the (1,2) mode. In fact, analyzing local cut-off frequencies shows that this mode is cut-on at the outlet (its local cut-off frequency is 416.8 Hz). Hence, it can be deduced from the difference observed that for $f=420$ Hz, the multiple-scales approximation fails.

The last test case (see Fig. 10) concerns a (7,0) mode at $f=500$ Hz propagating into the exact Rienstra's geometry. The outer wall is lined, the impedance value is $Z=4080(1+i)$. The central body is left perfectly rigid. Mesh and flow are not shown for conciseness ($M=0.5$ and 0.49 at the inlet and outlet, respectively). The goal of this example is to outline the reflection phenomenon that limits the use of a multiple-scales method. The frequency of $f=500$ Hz is chosen in order for the (7,0) mode to be the only cut-on mode. Calculations effectively give a maximum local cut-off frequency along the duct of 411.2 Hz for this mode and a mini-

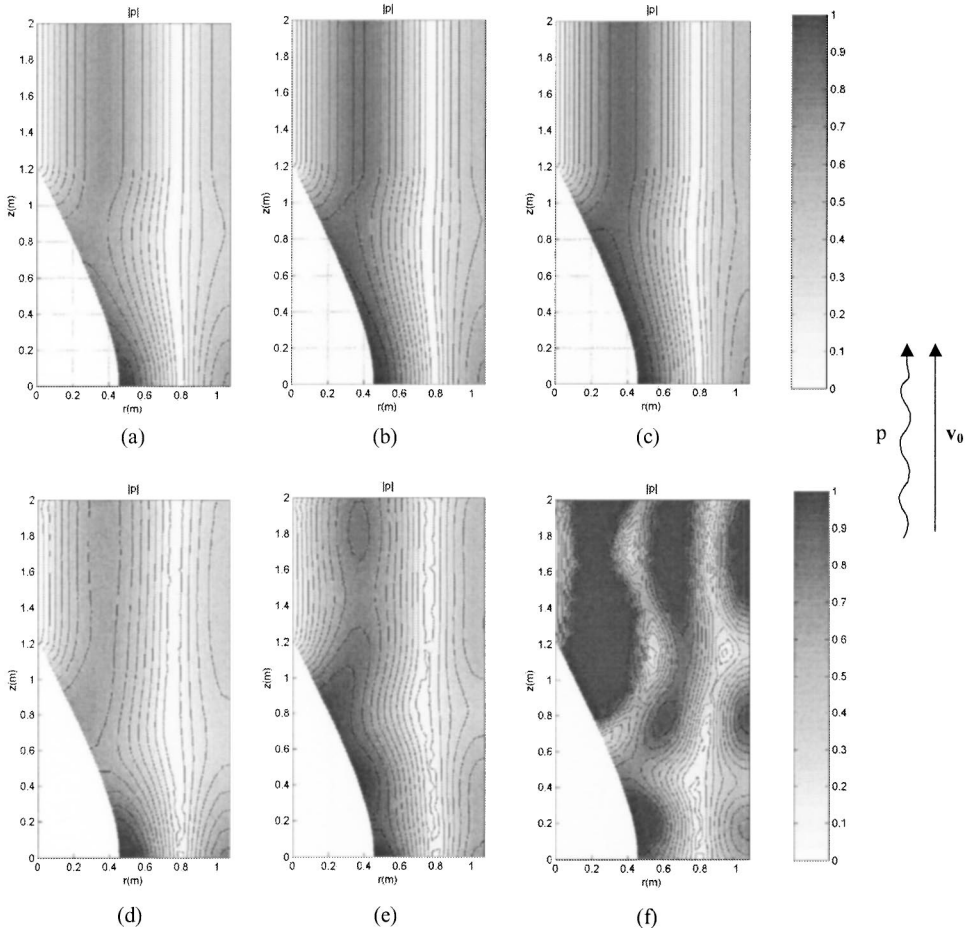


FIG. 9. Pressure modulus in pascals of the (1,1) mode (rigid walls). (a), (b), (c) Multiple-scale semianalytical solutions at $f=300$, 360, and 420 Hz, respectively. (d), (e), (f) Mixed FEM solutions at $f=300$, 360, and 420 Hz.

mum cut-off frequency of 558.6 Hz for the (7,1) mode. This indicates that the (7,0) mode is always cut-on and the (7,1) always cut-off. This permits one to avoid any significant scattering into other modes and thus to focus on autoreflection of the (7,0) mode only. Computations are made for both downstream and upstream propagation.

In the downstream case, a good agreement is obtained. This shows that only few reflections are produced inside the duct for this direction of propagation. In the upstream propagation case, some differences occur (on this plot, wave is propagating from the top to the bottom). Some wiggles appear and iso-pressure contours are not totally smooth. At the acoustic outlet (bottom), it can be seen on plots that the attenuation obtained by the FEM is a little greater than the semianalytical one, which tends to prove that reflections of the (7,0) mode on itself are not negligible. This may be attributed to the central body as well as the abrupt change of outer radius located at the acoustic inlet, both viewed as a narrowing for an upstream propagation.

Finally, it may be interesting to look at the acoustic intensity vector that has been plotted for both the upstream and downstream cases. Because the lining of the wall absorbs

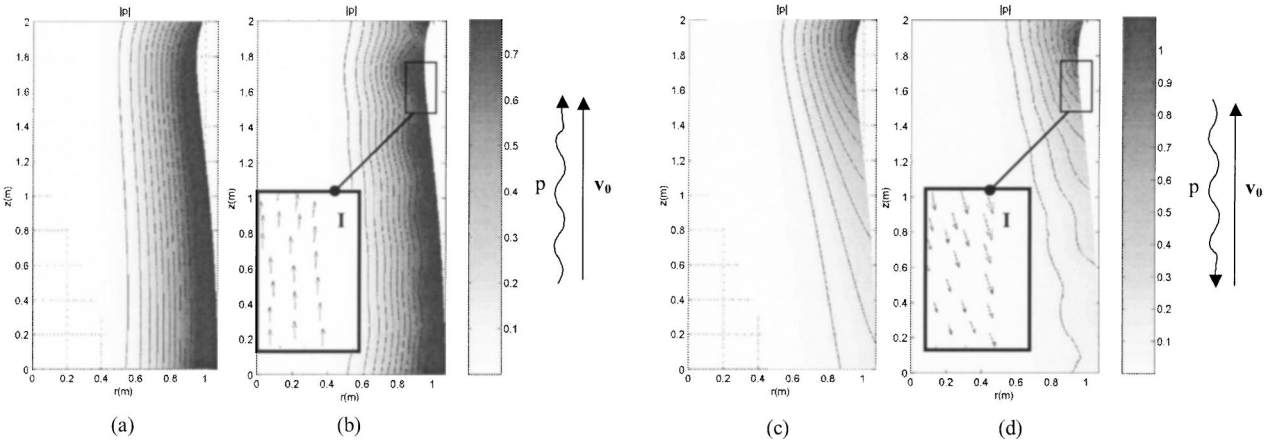


FIG. 10. Pressure modulus in pascals of the (7,0) mode at 500 Hz with lined walls ($Z=4080+4080i$) for the exact Rienstra's duct geometry. (a) and (b) Comparison between multiple-scales and mixed FEM solutions for a downstream propagation. (c) and (d) Comparison for an upstream propagation.

some energy, intensity is not exactly parallel to the wall but penetrates into it. This is more visible in the upstream case, which is coherent with the fact that the downstream wave is less attenuated.

V. CONCLUSION

In this paper, a mixed FEM based on Galbrun's equation has been proposed to solve sound wave propagation in arbitrary flows. Compared to the LEE, this equation may have several advantages. From a theoretical point of view, it yields an exact expression for the acoustic intensity and simpler boundary conditions (especially in the case of absorbing walls). From a numerical point of view, Galbrun's equation allows the direct application of the inf-sup condition, already encountered in the study of mixed FEM for incompressible media.

Results obtained with the proposed mixed FEM have been compared with two complementary semianalytical models and have been found to be in very good agreement. The comparison with the Pridmore–Brown equation has proven the efficiency of the numerical method to take into account convection and refraction from a boundary layer, which cannot be considered with a full-potential formulation. Comparisons with a multiple-scales method have fully validated the FEM for complex geometry and have also confirmed limitations of a multiple-scales approach when some significant reflection or diffraction occur.

Those results show that a mixed FEM method based on Galbrun's equation could be an interesting alternative to a finite-difference method based on the LEE, for solving aeroacoustic problems.

[1] R. K. Sigman, R. K. Majjigi, and B. T. Zinn, "Determination of turbofan inlet acoustics using finite elements," AIAA J. **16**, 1139–1145 (1978).
[2] R. Fuller and D. A. Bies, "The effects of flow on the performance of a reactive acoustic attenuator," J. Sound Vib. **62**, 73–92 (1979).
[3] A. Cabelli, "The influence of flow on the acoustic characteristics of a duct bend for higher order modes—a numerical study," J. Sound Vib. **82**, 131–149 (1982).
[4] Z. L. Ji, Q. Ma, and Z. H. Zhang, "A boundary element scheme for evaluation of four-pole parameters of ducts and mufflers with low Mach number non-uniform flow," J. Sound Vib. **185**, 107–117 (1995).
[5] S. J. Horowitz, R. K. Sigman, and B. T. Zinn, "An iterative finite element-integral technique for predicting sound radiation from turbofan inlets in steady flight," AIAA J. **24**, 1256–1262 (1986).
[6] K. J. Baumeister and S. J. Horowitz, "Finite element-integral acoustic simulation of JT15D turbofan engine," J. Vib., Acoustics, Stress, Reliability Design **106**, 405–413 (1984).
[7] R. J. Astley and J. G. Bain, "A three-dimensional boundary element scheme for acoustic radiation in low Mach number flows," J. Sound Vib. **109**, 445–465 (1986).
[8] M. Ben Tahar, "Formulation variationnelle par équations intégrales pour le rayonnement acoustique en présence d'écoulement non-uniforme" ("Variational formulation for acoustic radiation in non uniform flows using a boundary element method"), Thèse de docteur d'état, Université de Technologie de Compiègne, 1991.
[9] P. Zhang, T. Wu, and L. Lee, "A coupled FEM/BEM formulation for acoustic radiation in a subsonic non-uniform flow," J. Sound Vib. **192**, 333–347 (1996).
[10] R. J. Astley, "A finite element wave envelope formulation for acoustic radiation in moving flows," J. Sound Vib. **103**, 471–485 (1985).
[11] D. Roy and W. Eversman, "Improved finite element modeling of the turbofan engine inlet radiation problem," J. Vibr. Acoust. **117**, 109–115 (1995).
[12] W. Eversman and D. Okunbor, "Aft fan duct acoustic radiation," J. Sound Vib. **213**, 235–257 (1998).
[13] W. Eversman, "Mapped infinite wave envelope elements for acoustic radiation in a uniformly moving medium," J. Sound Vib. **224**, 665–687 (1999).
[14] L. Lee, T. W. Wu, and P. Zhang, "A dual-reciprocity method for acoustic radiation in a subsonic non-uniform flow," Eng. Anal. Boundary Elem. **13**, 365–370 (1994).
[15] S. W. Rienstra, "Sound transmission in slowly varying circular and annular lined ducts with flow," J. Fluid Mech. **380**, 279–296 (1999).
[16] S. W. Rienstra and W. Eversman, "A numerical comparison between the multiple-scales and finite-element solution for sound propagation in lined flow ducts," J. Fluid Mech. **437**, 367–384 (2001).
[17] P. Mungur and H. E. Plumbee, "Propagation and attenuation of sound in as soft-walled annular duct containing a sheared flow," NASA SP-207, 305–327 (1969).
[18] H. Nayfeh, J. E. Kaiser, and D. P. Telionis, "Acoustics of aircraft engine-duct systems," AIAA J. **13**, 130–153 (1975).
[19] N. K. Agarwall and M. K. Bull, "Acoustic wave propagation in a pipe with fully developed turbulent flow," J. Sound Vib. **132**, 275–298 (1989).
[20] J. Cooper and N. Peake, "Propagation of unsteady disturbances in a slowly varying duct with mean swirling flow," J. Fluid Mech. **445**, 207–234 (2001).
[21] H. Nayfeh, B. S. Shaker, and J. E. Kaiser, "Transmission of sound through nonuniform circular ducts with compressible mean flows," AIAA J. **18**, 515–525 (1980).
[22] K. Uenishi and M. K. Myers, "Two-dimensional acoustic field in a non-uniform duct carrying compressible flow," AIAA J. **22**, 1242–1248 (1984).
[23] L. Abrahamson, "A finite element algorithm for sound propagation in axisymmetric ducts containing compressible mean flow," AIAA Fourth Aeroacoustics Conference, Atlanta, 1977.
[24] R. J. Astley and W. Eversman, "A finite element for transmission in non-uniform ducts without flow: Comparison with the method of weighted residuals," J. Sound Vib. **57**, 367–388 (1978).
[25] R. J. Astley and W. Eversman, "Acoustic transmission in non-uniform ducts with mean flow. II. The finite element method," J. Sound Vib. **74**, 103–121 (1981).
[26] R. J. Astley and W. Eversman, "A finite element formulation of the eigenvalue problem in lined ducts with flow," J. Sound Vib. **65**, 61–74 (1979).
[27] V. Golubev and H. M. Atassi, "Acoustic-vorticity waves in swirling flows," J. Sound Vib. **209**, 203–222 (1998).
[28] L. Greverie and C. Bailly, "Construction d'un opérateur de propagation à partir des équations d'Euler linéarisées" ("Formulation of an acoustic wave operator based on linearized Eulere quations"), C. R. Acad. Sci., Ser. IIb: Mec., Phys., Chim., Astron. **326**, 741–746 (1998).
[29] E. Longatte and P. Lafon, "Computation of acoustic propagation in two-dimensional sheared ducted flow," AIAA J. **38**, 389–394 (2000).
[30] C. Bailly and D. Juve, "Numerical solution of acoustic propagation problems using linearized Euler equations," AIAA J. **38**, 22–29 (2000).
[31] H. Galbrun, *Propagation d'Une OndeSono re Dans l'Atmosphère et Théorie des Zones de Silence (Propagation of an Acoustic Wave in the Atmosphere and Theory of Zones of Silence)* (Gauthier-Villars, Paris, 1931).
[32] M. K. Myers, "On the acoustic boundary condition in the presence of flow," J. Sound Vib. **71**, 429–434 (1980).
[33] B. Poiree, "Les équations de l'acoustique linéaire et non linéaire dans un écoulement de fluide parfait" ("Equations of linear and non linear acoustics in a perfect fluid flow"), Acustica **57**, 5–25 (1985).
[34] O. A. Godin, "Reciprocity and energy theorems for waves in a compressible inhomogeneous moving fluid," Wave Motion **25**, 143–167 (1997).
[35] M. Ben Tahar and E. Goy, "Resolution of a vibroacoustic problem in the presence of a nonuniform mean flow," Fourth AIAA Joint Aeroacoustics Conference, Paper No. 98-2215, 1998.
[36] C. Peyret and G. Elias, "Finite-element method to study harmonic aeroacoustics problems," J. Acoust. Soc. Am. **110**, 661–668 (2001).
[37] S. Bonnet-Ben Dhia, G. Legendre, and E. Luneville, "Analyse mathématique de l'équation de Galbrun en écoulement uniforme" ("Mathematical analysis of Galbrun's equation with uniform flow"), C. R. Acad. Sci., Ser. IIb: Mec., Phys., Chim., Astron. **329**, 601–606 (2001).
[38] D. Blokhintzev, "The propagation of sound in an inhomogeneous and moving medium I," J. Acoust. Soc. Am. **18**, 322–328 (1946).

[39] M. A. Hamdi and Y. Ousset, ''A displacement method for the analysis of vibrations of coupled fluid-structure systems,'' Int. J. Numer. Methods Eng. **13**, 139–150 (1978).
[40] A. Bermudez and L. Hervella-Nieto, ''Finite element computation of three-dimensional elastoacoustic vibrations,'' J. Sound Vib. **219**, 279–306 (1999).
[41] K. J. Bathe, C. Nitikitpaiboon, and X. Wang, ''A mixed displacement-based finite element formulation for acoustic fluid-structure interaction,'' Comput. Struct. **56**, 225–237 (1995).
[42] X. Wang and K. J. Bathe, ''Displacement/pressure based mixed finite element formulations for acoustic fluid-structure interaction problems,'' Int. J. Numer. Methods Eng. **40**, 2001–2017 (1997).
[43] K. J. Bathe, *Finite Element Procedures* (Prentice Hall, Englewood Cliffs, NJ, 1996).
[44] J.-M. Jin, *The Finite Element Method in Electromagnetics* (Wiley, New York, 1993).
[45] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods* (Springer, New York, 1991).
[46] D. Chapelle and K. J. Bathe, ''The inf-sup test,'' Comput. Struct. **47**, 537–545 (1993).

Measurement of surface wave transmission coefficient across surface-breaking cracks and notches in concrete

Won-Joon Song[a)]
Department of Engineering Science & Mechanics, The Pennsylvania State University, University Park, Pennsylvania 16802

John S. Popovics[b)]
Department of Civil & Environmental Engineering, The University of Illinois, Urbana, Illinois 61801

John C. Aldrin[a)]
Computational Tools, 6797 Roanoake Court, Gurnee, Illinois 60031

Surendra P. Shah
The Center for Advanced Cement-based Materials, Northwestern University, Evanston, Illinois 60208

(Received 3 May 2002; revised 20 November 2002; accepted 22 November 2002)

In this paper, a technique for measuring a surface wave transmission coefficient across surface-breaking cracks and notches in a heterogeneous but globally isotropic material (concrete) is presented. Once the transmission coefficient across a surface discontinuity is known, its depth may be estimated. There are many difficulties in measuring the transmission coefficient experimentally owing to effects of wave path dependence, unknown characteristics of the receiver and the wave source, and the variation of impact event or receiver coupling. To eliminate the undesired effects, a self-calibrating measurement scheme is applied to obtain the surface wave transmission coefficient across notches and surface-breaking cracks in concrete. The obtained signal transmission coefficient is not affected by the experimental setup or the heterogeneous nature of the material. The testing scheme is described and experimental results obtained from concrete specimens with notches and surface-breaking cracks are presented. Repeatable and reliable measurements of surface wave transmission coefficient are obtained, which demonstrate a strong relation to normalized discontinuity depth. A numerical study using the boundary element method is presented, which verifies the experimental findings.
[DOI: 10.1121/1.1537709]

PACS numbers: 43.20.Gp, 43.35.Pt, 43.35.Zc [JGH]

I. INTRODUCTION

Techniques that non-destructively detect defects, measure the mechanical properties, or monitor the state of deterioration in concrete structures are of considerable interest to civil engineers.[1,2] Concrete structures suffer damage from environmental deterioration or repeated service loads, where the damage most often takes the form of cracking. A distinct single surface-breaking crack is a common and significant defect that can eventually lead to failure of a concrete structure. Thus, it is important to monitor in a nondestructive fashion the distance that the crack extends into the concrete. Several studies on nondestructive techniques to determine surface-breaking crack depth in concrete have been reported. Most efforts make use of time-of-flight methods: the time required for waves to travel around the tip of a surface-breaking discontinuity is measured, from which the depth is estimated. Several studies report success in determining the depth of simulated cracks (notches with well-defined tips) in concrete when the velocity of wave propagation in concrete is known and a particular wave pulse–crack interaction is realized.[3–5] The time-of-flight method, however, is not effective when realistic concrete cracks are tested, that is, when the crack tip is ill defined and the crack is tightly closed.[6]

Wave transmission or attenuation measurements do show high sensitivity to realistic cracking in concrete under laboratory conditions.[7] However, the practical measurement of the wave transmission coefficient across cracks in concrete structures has been restricted up to now because of disrupting effects of wave path dependence in heterogeneous materials, incoherent signal noise and source, receiver, and coupling variability. Practical one-sided surface wave transmission coefficient measurements in relatively homogeneous materials such as metals have been achieved through the use of a self-compensating testing scheme: after the disrupting experimental variability is eliminated, the obtained surface wave transmission coefficient is used to obtain the depth of surface-breaking cracks.[8] The self-compensating approach has several advantages: the wave velocity of the material need not be known in advance and the technique is independent of the type of wave transmitter and receiver and coupling conditions. Numerical studies[9] and preliminary self-compensating surface wave measurements on concrete[6,10,11] demonstrate great sensitivity of self-compensating surface wave transmission to the presence of cracking along the wave path, although the precise relation between the crack

[a)]Portions of this work were carried out at The Center for Quality Engineering and Failure Prevention at Northwestern University.
[b)]Electronic mail: johnpop@uiuc.edu

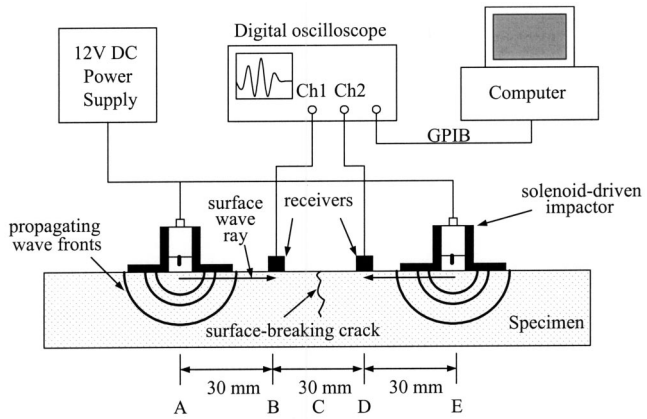


FIG. 1. Experimental setup for self-calibrating surface wave transmission measurements.

depth in concrete and the surface wave transmission coefficient from the crack has yet to be established. The testing scheme for concrete is a modified version of that reported by Achenbach *et al.*;[8] the modification was needed to allow testing in inhomogeneous materials. Signal transmission obtained with the modified self-compensating scheme reduces sharply as the crack depth relative to the wavelength of the surface wave increases, but is largely unaffected by concrete composition, the nature of a crack (crack opening condition), and the testing characteristics such as the type of wave transmitter and receiver.[11] In order to apply the results effectively, however, the accuracy of the measurement must be established and a unique relationship between the surface wave transmission coefficient across the crack and the depth of the crack in concrete obtained. That work is the focus of this paper, where refined experimental results obtained using the modified self-compensating scheme and numerical data are presented to verify the accuracy and illustrate the relationship.

II. SELF-CALIBRATING SURFACE WAVE TRANSMISSION MEASUREMENT TECHNIQUE

A. Principle

Two wave sensors and two wave sources (point impactors) are placed along a line on the surface of a concrete specimen at the given spacing, as shown in Fig. 1. A surface-breaking crack or notch lies at location C, midway between the two wave sensors. When a transient point load is applied at location A, surface waves generated by the point load propagate along the surface of the specimen and the resulting motion is captured by the near sensor (location B) and then by the far sensor (location D) across the surface-breaking crack. In the frequency domain, the signals that are sent from the impact source at location A and detected by the receivers at locations B and D can be represented as a simple product of terms in frequency domain:

$$V_{\mathrm{AB}}^{\mathrm{cr}}=S_{\mathrm{A}}^{\mathrm{cr}}d_{\mathrm{AB}}^{\mathrm{cr}}R_{\mathrm{B}}^{\mathrm{cr}}, \tag{1}$$

$$V_{\mathrm{AD}}^{\mathrm{cr}}=S_{\mathrm{A}}^{\mathrm{cr}}d_{\mathrm{AB}}^{\mathrm{cr}}d_{\mathrm{BC}}^{\mathrm{cr}}T_{\mathrm{cr}}\,d_{\mathrm{CD}}^{\mathrm{cr}}R_{\mathrm{D}}^{\mathrm{cr}}, \tag{2}$$

where $V_{\mathrm{AB}}^{\mathrm{cr}}$ and $V_{\mathrm{AD}}^{\mathrm{cr}}$ are modulation of Fourier transformed signals, $S_{\mathrm{A}}^{\mathrm{cr}}$ the generating response term, $R_{\mathrm{B}}^{\mathrm{cr}}$ and $R_{\mathrm{D}}^{\mathrm{cr}}$ the receiving response terms, and $d_{\mathrm{AB}}^{\mathrm{cr}}$, $d_{\mathrm{BC}}^{\mathrm{cr}}$, and $d_{\mathrm{CD}}^{\mathrm{cr}}$ the signal transmission function between locations A and B, B and C, and C and D.[8] As surface waves propagate along the surface along a cylindrical wave front, some of the energy will be dissipated, and as a result the wave amplitude will attenuate. Therefore, signal transmission function terms include geometrical attenuation. $T_{\mathrm{cr}}(f)$ is the transmission coefficient of surface waves incident on a surface-breaking crack as a function of the frequency. The S_i^{cr} and R_i^{cr} terms include unknown effects from the variation of impact events and receiver type and coupling. In order to eliminate the S_i^{cr}, R_i^{cr}, and d_{ij}^{cr} terms, and thus measure the surface wave transmission response across a surface-breaking crack between two receivers, a complimentary set of signals $V_{\mathrm{ED}}^{\mathrm{cr}}$ and $V_{\mathrm{EB}}^{\mathrm{cr}}$ must be collected. These signals are generated by a normal point load at location E and detected by the receivers at locations D and B, respectively, and expressed as a product of terms:

$$V_{\mathrm{ED}}^{\mathrm{cr}}=S_{\mathrm{E}}^{\mathrm{cr}}d_{\mathrm{ED}}^{\mathrm{cr}}R_{\mathrm{D}}^{\mathrm{cr}}, \tag{3}$$

$$V_{\mathrm{EB}}^{\mathrm{cr}}=S_{\mathrm{E}}^{\mathrm{cr}}d_{\mathrm{ED}}^{\mathrm{cr}}d_{\mathrm{DC}}^{\mathrm{cr}}T_{\mathrm{cr}}\,d_{\mathrm{CB}}^{\mathrm{cr}}R_{\mathrm{B}}^{\mathrm{cr}}. \tag{4}$$

We may set $d_{\mathrm{BC}}^{\mathrm{cr}}=d_{\mathrm{CB}}^{\mathrm{cr}}$ and $d_{\mathrm{CD}}^{\mathrm{cr}}=d_{\mathrm{DC}}^{\mathrm{cr}}$ since those terms describe the same wave path and experience the same geometrical attenuation. An expression for the signal transmission between locations B and D, $d_{\mathrm{BD}}^{\mathrm{cr}}$, can then be obtained with the following manipulation:

$$|d_{\mathrm{BD}}^{\mathrm{cr}}(f)|=|d_{\mathrm{BC}}^{\mathrm{cr}}T_{\mathrm{cr}}\,d_{\mathrm{CD}}^{\mathrm{cr}}|=\left|\sqrt{\frac{V_{\mathrm{AD}}^{\mathrm{cr}}V_{\mathrm{EB}}^{\mathrm{cr}}}{V_{\mathrm{AB}}^{\mathrm{cr}}V_{\mathrm{ED}}^{\mathrm{cr}}}}\right|. \tag{5}$$

The superscript, "cr," implies that the signal transmission is obtained along a path that contains a surface-breaking crack, although the signals $V_{\mathrm{AB}}^{\mathrm{cr}}$ and $V_{\mathrm{ED}}^{\mathrm{cr}}$ are not influenced directly by the surface-breaking crack. This formulation assumes that the mounted receivers have no effect on the passing waves and that only surface-guided waves are monitored. $d_{\mathrm{BD}}^{\mathrm{cr}}$ is a function of frequency and the expected range of the $d_{\mathrm{BD}}^{\mathrm{cr}}$ value is 0 (complete attenuation) to 1 (complete transmission). The propagating surface waves are reflected and scattered by the presence of the surface-breaking crack. Therefore, $d_{\mathrm{BD}}^{\mathrm{cr}}$ should, in principle, decrease as the depth of the surface-breaking crack increases. $d_{\mathrm{BD}}^{\mathrm{cr}}$ also depends on the relative spacing between the receivers and wave sources: $d_{\mathrm{BD}}^{\mathrm{cr}}$ should decrease as the sensor spacing increases for a fixed spacing between wave source and sensor. This behavior arises from geometrical attenuation of the waves, which is included in the $d_{\mathrm{BC}}^{\mathrm{cr}}$ and $d_{\mathrm{CD}}^{\mathrm{cr}}$ terms.

The transmission coefficient of the surface wave across the surface-breaking crack itself, $T_{\mathrm{cr}}(f)$, can be obtained by collecting the additional signal transmission from a crack-free wave path on the same specimen using the same impactor–receiver configuration. By repeating the same procedure along the crack-free path, four more signals are obtained. The signals from the crack-free path are represented as

$$V_{\mathrm{AB}}=S_{\mathrm{A}}d_{\mathrm{AB}}R_{\mathrm{B}}, \tag{6}$$

$$V_{\mathrm{AD}}=S_{\mathrm{A}}d_{\mathrm{AB}}d_{\mathrm{BC}}d_{\mathrm{CD}}R_{\mathrm{D}}, \tag{7}$$

$$V_{\mathrm{ED}}=S_{\mathrm{E}}d_{\mathrm{ED}}R_{\mathrm{D}}, \tag{8}$$

$$V_{EB} = S_E d_{ED} d_{DC} d_{CB} R_B . \quad (9)$$

The signal transmission between location B and D from the crack-free area, d_{BD}, is obtained using the same manipulation as in Eq. (5):

$$|d_{BD}(f)| = |d_{BC} d_{CD}| = \left| \sqrt{\frac{V_{AD} V_{EB}}{V_{AB} V_{ED}}} \right| . \quad (10)$$

d_{BD} includes only the material signal transmission functions between locations B and C and locations C and D. Within a particular concrete specimen, we assume $d_{BC}^{cr} = d_{BC}$ and $d_{CD}^{cr} = d_{CD}$ since the values are independent of the crack itself and the material composition and curing history are identical, even though the wave paths are different. The transmission coefficient of the surface wave incident on the surface-breaking crack is then obtained by dividing the signal transmission across the surface-breaking crack by the signal transmission from a crack-free path on the same specimen,

$$|T_{cr}(f)| = \left| \frac{d_{BD}^{cr}}{d_{BD}} \right| = \left| \frac{d_{BC} T_{cr} d_{CD}}{d_{BC} d_{CD}} \right| . \quad (11)$$

A relationship between $T_{cr}(f)$ and crack depth is now obtained by carrying out this procedure on a collection of concrete specimens with varying crack depths.

B. Test specimens and crack/notch generation

For the experiments, two batches of normal strength concrete specimens were cast: one batch contains notches with successively increasing depth and the other contains a 10 mm deep starter notch to initiate a surface-breaking crack of varying depth. All the specimens have 10 cm thickness and 41 cm×41 cm dimension in plan. All specimens are comprised of Type I portland cement (C), water (W), well-graded washed sand (FA), and gravel coarse aggregate (CA). The proportions of the concrete by mass are 1: 0.5:2:2 (C:W: FA:CA, respectively). After casting, the concrete specimens were cured at 23 °C (73 °F) and 95% relative humidity for a period of 28 days. Companion 101.6 mm×203.2 mm cylinders were cast for each batch in order to measure the 28-day compressive strength, f'_C, of the concrete as specified by ASTM.[12] The average value of f'_C was 33.4 MPa. The surface wave velocity (C_R) was measured on all of the specimens using the one-sided wave velocity measurement technique described in Ref. 13. The wavelength of the surface wave, λ_R, can then be computed for a given frequency. The longitudinal wave velocity (C_L) for each specimen was obtained with the through-transmission pulse velocity method, specified by ASTM.[14] The transverse wave velocity (C_T) and Poisson's ratio were computed using the well-known relation between wave velocities and the elastic constants of a material.[15] The measured wave velocities, and the computed material properties based on the wave velocities, for all specimens are given in Tables I and II.

For the notched specimens C1–C7, the notches were placed along the centerline of the specimen directly after casting by inserting a 0.5 mm thick aluminum sheet into the fresh concrete. The aluminum sheet was removed after the concrete set, leaving behind the formed notch. The notch tip is well defined and the depth can be established accurately. The starter notches for the controlled crack specimens were made in the same way.

TABLE I. Description of concrete specimens with varying notch depth.

Specimen label	C_L (m/s)	C_R (m/s)	C_T (m/s)	Poisson's ratio ν	Density, ρ (kg/m^3)
C1	4563	2302	2487	0.289	2264.7
C2	4613	2248	2420	0.310	2292.0
C3	4533	2366	2567	0.264	2229.9
C4	4506	2174	2338	0.316	2209.1
C5	4442	2283	2472	0.276	2230.5
C6	4434	2143	2305	0.315	2284.0
C7	4397	2304	2501	0.261	2215.0

In order to generate controlled cracking in the concrete a closed-loop loading scheme was used, which is described in Refs. 11 and 16. This closed-loop loading scheme enables controlled crack propagation without unstable and sudden failure of the specimen. The complete load-deformation response, including the post-peak portion, may be obtained. The concrete slabs were subjected to three-point bending with simple support on the bottom surface. The line load is applied along the centerline of the specimen on the top surface using a 1 MN capacity MTS servohydraulic testing machine. Thus, a crack initiates from the starter notch and propagates into the concrete from the tension side of the plate. The signal transmission measurements were performed in the central region of the specimen across the surface-breaking crack. The load was applied using average crack mouth opening displacement (CMOD) control, which was monitored by two linear variable displacement transducers (LVDTs) mounted across the crack mouth at a distance of 65 mm from an edge. In order to generate surface-breaking cracks with different depths, each concrete specimen was continuously loaded until a specified point in the post-peak region of the load-CMOD response was reached. The load was maintained at this point and the signal transmission data was collected. The surface wave signal transmission measurements were performed under the loaded state (open crack case). Each concrete specimen has a single, controlled surface-breaking crack with a certain depth. The load-CMOD response for specimen C10 generated by the closed-loop loading scheme is shown in Fig. 2.

After the loading (cracking) and signal transmission measurements were completed, the specimen was cut along the centerline perpendicular to the crack plane. The surface-breaking crack at the location where the signal transmission measurements were performed was thereby exposed. The ac-

TABLE II. Description of concrete specimens with varying surface-breaking crack depth.

Specimen label	C_L (m/s)	C_R (m/s)	C_T (m/s)	Poisson's ratio ν	Density, ρ (kg/m^3)
C8	4410	2248	2432	0.282	2268.6
C9	4528	2304	2492	0.283	2283.9
C10	4402	2237	2419	0.284	2248.8
C11	4343	2207	2386	0.284	2236.4
C12	4517	2333	2528	0.272	2305.3

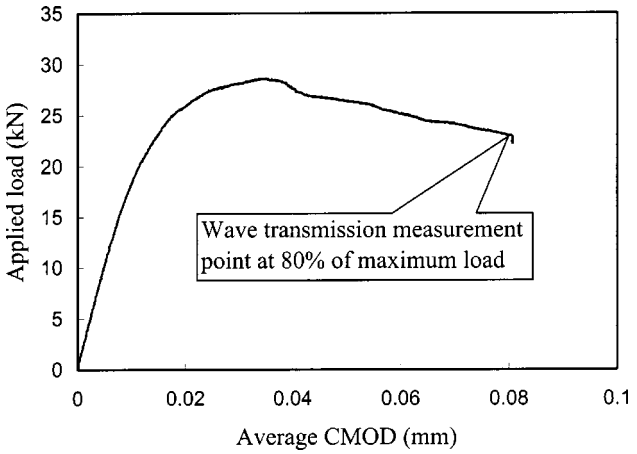


FIG. 2. The load-CMOD response for concrete specimen C10 obtained by the closed-loop loading scheme.

tual depth of the surface-breaking crack on the cut surface was measured by the laser-based Phase Measurement Interferometry (PMI) technique, which provides sensitive crack length determination by measuring in-plane horizontal displacements on the surface of a loaded specimen.[11,17,18] A typical phase fringe map obtained from a cracked concrete specimen is shown in Fig. 3. The surface-breaking crack is indicated by fringe discontinuities in the phase fringe map (heavy line). The image provides information about the length and location of an existing crack from the surface displacement contours. The image in Fig. 3 shows the full thickness of the specimen. Thus, the final surface-breaking crack depth, with respect to the full slab depth, can be calculated by measuring a percentage of the crack depth on the image.

C. Data collection and analysis

The experimental setup for self-calibrating surface wave transmission measurement consists of two solenoid-driven impactors, two receivers, a digital oscilloscope, and a personal computer that collects the data from the oscilloscope

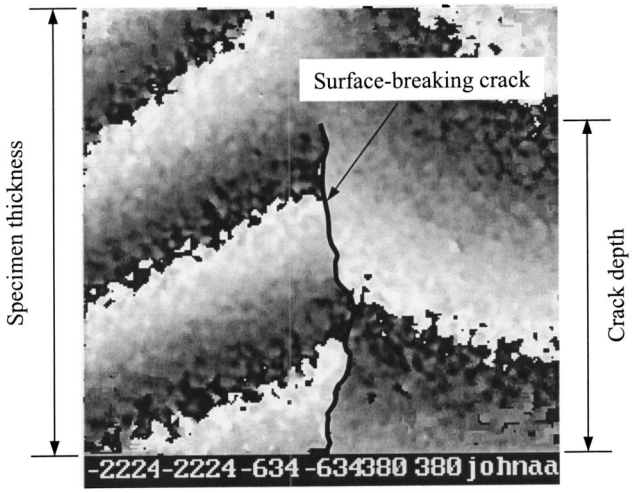


FIG. 3. Phase fringe map obtained with a PMI technique on the cut face of a specimen. The heavy line indicates the location and depth of the surface-breaking crack.

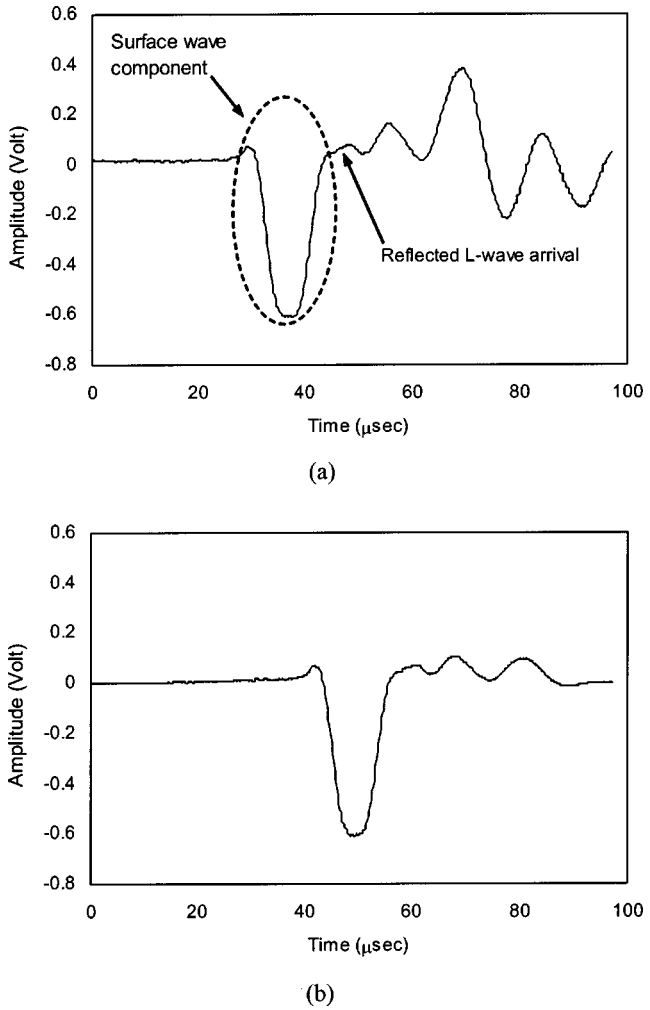


FIG. 4. The time–domain signal generated by an impact event and collected from a receiver on a crack-free concrete specimen. $C_L = 4510$ m/s, $C_R = 2350$ m/s, thickness $= 101.37$ mm: (a) the typical raw signal before processing and (b) the signal after shifting with respect to the center of the time window at 50 μs and smoothing with a Hanning window.

using a GPIB interface, as shown in Fig. 1. After an electric current pulse is applied, the solenoid drives a 9 mm steel ball mounted on a spring-loaded shaft, and the ball then strikes the surface of the concrete once. In this manner, consistent point source wave pulses are generated without damaging the surface of the concrete specimen. The receivers are miniature accelerometers with the following properties: contact area $= 25\ \text{mm}^2$, mass $= 0.7$ g, nominal voltage acceleration sensitivity $= 1.02\ \text{mV}/(\text{m/s}^2)$, and nominal $\pm 10\%$ flat frequency response over 1–25 kHz. Two solenoid-driven impactors and two accelerometers are mounted along a line on the same surface of the test specimen with a spacing between each of 30 mm. The surface-breaking crack (or notch) is located midway between the two receivers. When the specimen is subjected to a sudden impact on the surface by the impactor (locations A and E on the upper surface), waves propagate in all the directions, including along the surface of the specimen. The propagating waves are detected by receivers (locations B and D) and the signals are manipulated using a computer program. Figure 4(a) shows a typical unprocessed time–domain signal generated by the impact event and

detected by a surface-mounted receiver. The signal is sampled for a duration of 97.28 μs and digitized with sampling frequency of 5.26 MHz. The captured waveform includes wave components that propagate along the surface (direct longitudinal wave and Rayleigh surface wave), and through the thickness reflecting from the opposite free surface (reflected longitudinal wave). The Rayleigh surface wave arrival coincides with the sharp drop in signal value immediately after the positive peak. The arrival of the reflected longitudinal wave follows soon thereafter. The expected arrival times of the Rayleigh surface wave and the reflected longitudinal wave are marked in Fig. 4(a). The amplitude of the reflected longitudinal wave component is very small compared to the Rayleigh surface wave. Therefore, it is assumed that the captured signal is dominated by the surface wave. In order to reduce the amplitude of other wave components that may exist in the signal, the signal is digitally processed. The signal is shifted in time so that a common signal feature, the large negative peak, coincides with the center of the time window. The shifted signal is then smoothed with a Hanning window, which has a Gaussian distribution. This process transforms the shifted signal to have consistent amplitude at the center and zero amplitude at each end, as shown in Fig. 4(b). Therefore, the Rayleigh surface wave component retains the same amplitude while the amplitudes of other wave components are reduced. A series of zeros are then added to the end of the smoothed signal until the total signal duration is 8192 points. The windowed and zero-padded time signal is then transformed to the frequency domain using a FFT algorithm, where the frequency resolution is 642.48 Hz. In order to improve the signal-to-noise ratio, five time domain signals, resulting from five repeated impact events, are collected from each receiver. Each signal is individually processed by shifting, windowing, and FFT. Using these signals, five repeated self-compensating signal transmission values are obtained using Eq. (5). These five transmission functions are then arithmetically averaged at each frequency point to give one transmission function as a function of frequency, $d_{\mathrm{BD}}^{\mathrm{cr}}$. This averaging reduces incoherent noise content in the signal, so the signal-to-noise ratio is improved. In addition, the five transmission functions are also used to compute the frequency-dependent signal consistency SC(f), which is used to define the frequency range within which the signal transmission measurement is acceptable. SC(f) is defined as the quotient of the geometric average and the arithmetic average of the five transmission values as a function of frequency:

$$\mathrm{SC}(f)=\frac{\sqrt[5]{d_{\mathrm{BD}}d_{\mathrm{BD2}}d_{\mathrm{BD3}}d_{\mathrm{BD4}}d_{\mathrm{BD5}}}}{(d_{\mathrm{BD1}}+d_{\mathrm{BD2}}+d_{\mathrm{BD3}}+d_{\mathrm{BD4}}+d_{\mathrm{BD5}})/5}. \qquad (12)$$

The value of SC(f) may range from 0 (no consistency among signals) to 1 (perfect consistency) at each frequency. SC(f) has been shown to be very useful for defining usable frequency ranges in impact-generated signals that contain a broad range of frequencies with signal to signal variation and incoherent noise, regardless of the expected operating frequency range of the sensors. SC(f) below 0.99 indicates unsuitably high variability or noise content at a given frequency.[11] In this study, an averaged signal transmission datum was accepted only if SC(f) was greater than 0.99 at that given frequency. A usable frequency range of 0–85 kHz was typically obtained using the described testing setup and the signal consistency criterion, which is much broader than the nominal flat frequency range for the sensors (1–25 kHz). The entire procedure was then repeated to obtain d_{BD} from the same concrete sample but across an uncracked wave path. The surface wave transmission coefficient was then computed using Eq. (11).

TABLE III. The thickness and the final notch depth of the notched concrete specimens.

Specimen	Specimen thickness (mm)	Final notch depth (mm)
C1	101.95	10.19
C2	101.17	15.24
C3	106.75	20.88
C4	103.31	27.51
C5	100.16	31.83
C6	98.85	41.96
C7	102.45	51.71

III. EXPERIMENTAL RESULTS

A. Surface-breaking notch

The surface wave transmissions coefficient was measured for the notched concrete specimens using the self-calibrating scheme. The notch depth varied from 10 to 52 mm while the width was approximately 0.3 mm. The specimen thickness and the final notch depth for each concrete specimen are given in Table III.

The signal transmission measurements were performed across the notch ($d_{\mathrm{BD}}^{\mathrm{cr}}$) and along a notch-free path (d_{BD}) using Eqs. (5) and (10), respectively. Only signal transmission data with signal consistency greater than 0.99 were accepted. The surface wave transmission coefficient versus the notch depth normalized with respect to the wavelength of the surface wave, a/λ_R, for all notched concrete specimens are plotted in Fig. 5. The transmission coefficient data show a clear trend with regard to notch depth: as the normalized notch depth increases, the transmission coefficient decreases. The transmission coefficient values for a normalized notch

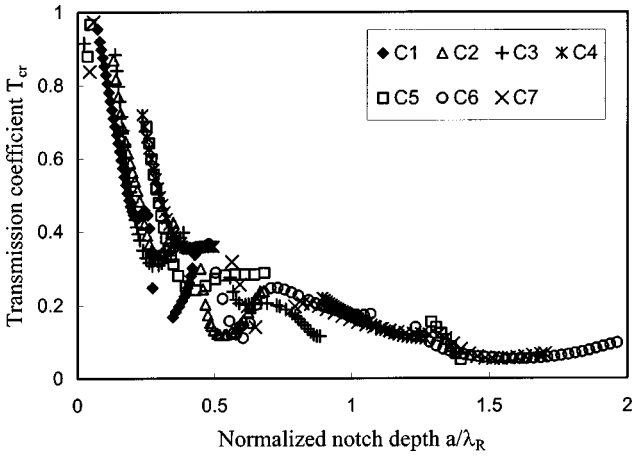


FIG. 5. The surface wave transmission coefficient as a function of normalized notch depth for concrete specimens with varying notch depth.

TABLE IV. The thickness and the final crack depth of cracked concrete specimens.

Specimen	Specimen thickness (mm)	Final load in post-peak/maximum load (%)	Final crack depth (mm)
C8	102.02	100%	33
C9	98.46	90%	43
C10	102.95	80%	49
C11	97.71	70%	52
C12	98.11	60%	57

depth less than 0.35 are consistent and very sensitive to the change of the normalized notch depth. However, the transmission coefficient data in the normalized notch depth from 0.35 to 0.6 show a relatively large variation. The transmission coefficient values in the normalized notch depth greater than 1.0 approach a constant level of about 0.06, so in this region the transmission coefficient is not sensitive to a change in the normalized notch depth.

B. Surface-breaking crack

The surface-breaking cracks were initiated from the 10 mm starter notch and induced into the concrete specimen using the closed-loop loading scheme. All surface-breaking cracks propagated into the concrete specimen normally with respect to the surface. Each concrete specimen was subjected to a different magnitude of final load in the post-peak region of the load-CMOD response: 100%, 90%, 80%, 70%, and 60% of the maximum load. The final crack depth of each specimen was determined with the PMI technique after the signal transmission measurements. Table IV gives the actual specimen thickness, final load in the post-peak region, and the final crack depth of all specimens.

$d_{\mathrm{BD}}^{\mathrm{cr}}$ was collected across the surface-breaking crack while the specimens were under the desired magnitude of the final load (open crack state). d_{BD} was similarly collected along a crack-free path on the same specimen. The surface wave transmission coefficient across the surface-breaking crack as a function of normalized crack depth, a/λ_R, was obtained as in the notched case. The results for the cracked specimens are shown in Fig. 6. The measured transmission coefficient data again show a clear relationship between the surface wave transmission coefficient and the normalized crack depth, although the transmission coefficient data across cracks show more scatter than that across notches. The transmission coefficient decreases as the normalized crack depth increases. The measured transmission coefficient values for a normalized crack depth from 0 to 0.4 show high sensitivity to the change of the normalized crack depth, although some inconsistency in T_{cr} values from concrete specimens is observed. T_{cr} for the normalized crack depth between 0.4 and 1.0 show no consistent behavior. The sensitivity of T_{cr} to the change of the normalized crack depth reduces significantly for a normalized crack depth greater than 1.0, as seen in the notch case, and T_{cr} approaches a constant value.

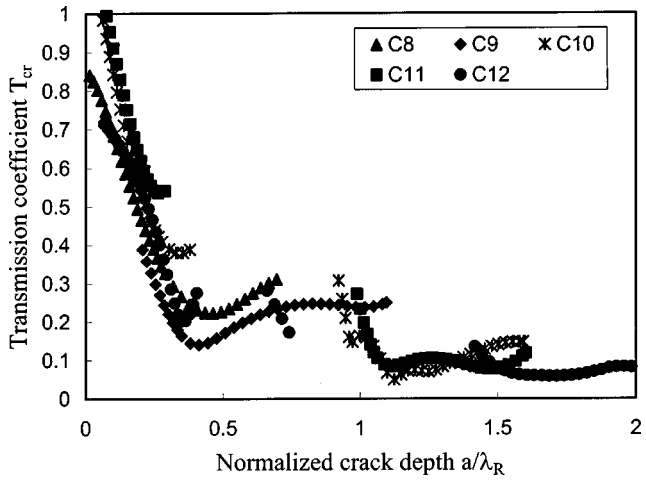


FIG. 6. The surface wave transmission coefficient as a function of normalized crack depth for concrete specimens with varying crack depth.

IV. NUMERICAL RESULTS

Work on analytical models for the scattering behavior of waves incident on cracks have been reported in the literature.[19–22] In these references, the scattered displacement field of waves incident on cracks are computed. In addition, Angel and Achenbach derived the reflection and transmission coefficients for a plane surface wave obliquely incident on a surface breaking crack.[21] This work demonstrated the potential for measured specular reflection to provide information for crack sizing. However, for this solution of the reflection and transmission coefficients, asymptotic expansions were applied where the location of the displacement field approached infinity.

For this study, a model is desired to represent the impact source on the material surface and the sensor response at locations near the crack in the experiment. Boundary integral equations in conjunction with the boundary element method provide an effective numerical technique to model the scattered wave field. Previously, the boundary integral equation method has been successfully applied to a wide range of problems in linear and nonlinear elasticity.[22–27] The method is based on the Betti–Rayleigh reciprocity theorem for time harmonic elastodynamic fields. This theorem can be used to express the scattered field in terms of an integral over the surface of the crack. Then, the boundary domain is discretized and numerical integration methods are applied to obtain a system of equations that can be solved.

In this numerical analysis, the surface wave transmission coefficient, $T_{\mathrm{cr}}(f)$, of a plane time-harmonic surface wave normally incident on both surface-breaking cracks and notches was calculated with the boundary element method in order to verify the experimentally measured surface wave transmission coefficient. The total displacement field at a point on the surface, u^{total}, generated by the interaction of the incident surface wave with the crack can be written as the sum of the incident field u^{in} and the scattered field u^{sc}:

$$u^{\mathrm{total}} = u^{\mathrm{in}} + u^{\mathrm{sc}}. \tag{13}$$

For a given incident field, the scattered field is determined using the boundary integral equations in conjunction with the

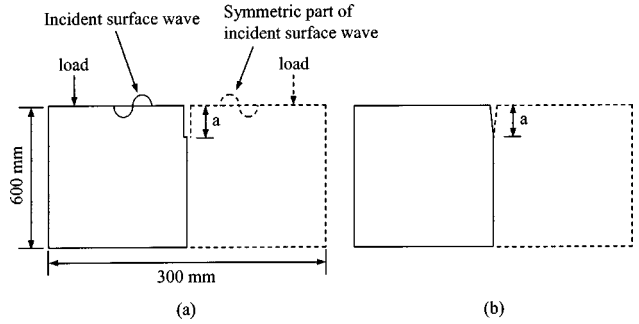


FIG. 7. An illustration of the boundary configurations of the numerical model and the calculation of the scattered field in the symmetric case: (a) surface-breaking notch case and (b) surface-breaking crack case.

boundary element method. Then, the transmission coefficient is defined by

$$T_{\mathrm{cr}}(f)=\frac{u^{\mathrm{total}}}{u^{\mathrm{in}}}. \tag{14}$$

Numerical calculations of the scattered displacement field by a surface-breaking crack have been reported.[20–25] The existing numerical codes for the boundary element method were adapted for this analysis.[27] In the experiments, a surface wave with a cylindrical wave front was generated and the surface wave transmission was measured along a line that is normal to the crack plane. Therefore, it is assumed that the signals collected by the receivers are not affected by the displacement field away from this line. A plane surface wave assumed in the numerical procedure is therefore appropriate for this case. In the analysis, the two-dimensional problem of a plane surface wave normally incident on a surface-breaking crack/notch of depth "*a*" is investigated. The shape of the surface-breaking crack/notch is assumed to be uniform.

Figure 7 shows the boundary configurations of the models for the surface-breaking notch and surface-breaking crack cases. In the calculation of the scattered displacement field, a symmetric/antisymmetric scheme was used.[19–21,27] The scattered displacement field was decomposed into symmetric and antisymmetric fields with respect to the plane of the crack. Figure 7 illustrates the calculation of the scattered field for the symmetric case. The boundary conditions along the line of symmetry satisfy the symmetric loading case. In the antisymmetric case the boundary conditions on the symmetric plane correspond to antisymmetric loading, thus the load indicated with the dashed line should point in the opposite direction. The transient elastodynamic problem was solved by transforming the time signals into the frequency domain using Fourier transforms (FFT), solving a series of frequency domain problems using the boundary element method, and acquiring the time domain solution using an inverse Fourier transform. With the transient solution at points B and D, the same signal processing approach (filtering, zero-padding, transform) as in the experiment was applied. Using the calculated incident and total normal displacement field, the surface wave transmission coefficient was calculated numerically for both notch and crack cases. The material elastic constants used in the numerical analysis were based on the measured wave velocities of the concrete specimens. Since each specimen had different wave velocities, average values were used. Table V shows the wave velocities and the material elastic constants used in the calculation. For the simulation, the width of the notch was defined to be 0.3 mm, which is the actual width of the notch in concrete specimens.

TABLE V. The wave velocities and the material constants used in the numerical analysis.

	C_L (m/s)	C_R (m/s)	C_T (m/s)	Density, ρ (kg/m^3)	Poisson's ratio, ν	Shear modulus (GPa)
Notch case	4500	2271	2455	2251.1	0.289	13.56
Crack case	4448	2270	2456	2268.6	0.281	13.68

Figure 8 shows the computed surface wave transmission coefficients as a function of crack/notch depth normalized with respect to the wavelength of the surface wave. The computed surface wave transmission coefficient data for the crack and notch cases are similar. In particular, the transmission coefficients for the normalized crack/notch depth from 0 to 0.3 and above 1.7 show close agreement, while some discrepancy between the transmission coefficient values of the two cases is seen for normalized crack/notch depths from 0.3 to 0.6. These small discrepancies can be contributed to differences in the tip geometry (single crack tip versus notch end with two square corners), resulting in slight differences in the reflection, transmission, and diffraction of the incident surface wave.

The experimental and numerical results are compared in Figs. 9 and 10 for the notch and crack cases, respectively. In these figures the experimental data are replotted as crosses and the numerical computations as solid lines. It can be seen that the surface wave transmission coefficient data obtained from the experiment show excellent agreement with that computed using BEM, although some discrepancies are seen for normalized notch depths of 0.4–0.8 (Fig. 9) and normalized crack depths of 0.4–1.0 (Fig. 10).

V. CONCLUSIONS

In this paper, a self-calibrating technique for the measurement of the surface wave transmission coefficient (T_{cr})

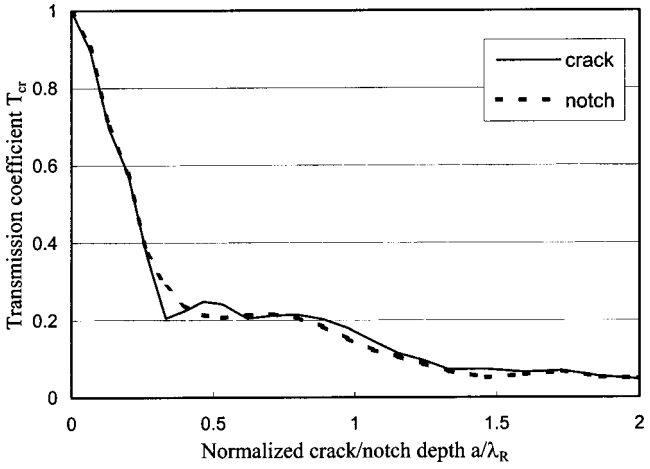


FIG. 8. The surface wave transmission coefficient as a function of the normalized crack/notch depth with a surface-breaking crack and a notch computed using the boundary element method.

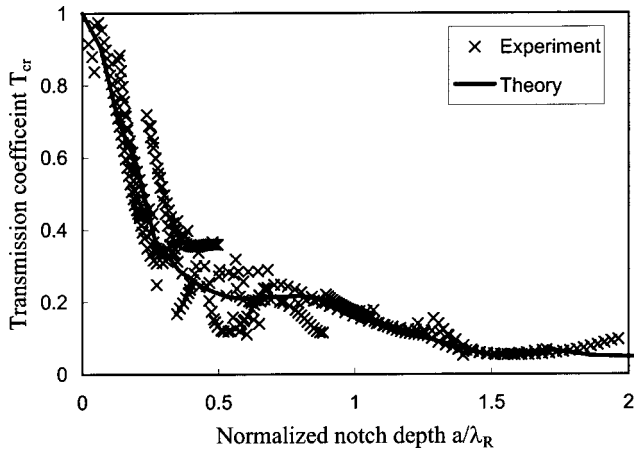


FIG. 9. The surface wave transmission coefficient across a notch obtained from the self-calibrating transmission measurement and the boundary element method.

across notches and surface-breaking cracks in concrete is presented. Boundary element method (BEM) computations are compared to experimental results. Experimental and numerical values of T_{cr} show the expected general decrease with increasing normalized notch and crack depth up to $a/\lambda_R = 1.5$: a sharp decrease is seen for $a/\lambda_R < 0.3$ and a gradual decrease for $0.3 < a/\lambda_R < 1.5$. T_{cr} approaches a constant value for $a/\lambda_R > 1.5$.

The agreement between the experimental and numerical data verifies that results obtained using the described test setup provide good estimates of T_{cr} for surface-breaking cracks and notches in concrete, especially for small ($a/\lambda_R < 0.4$) and large ($a/\lambda_R > 1.0$) values of normalized notch and crack depth. In terms of T_{cr} values, these regions of good agreement are defined by $T_{cr} > 0.4$ and $T_{cr} < 0.1$. The observed discrepancies at intermediate values of normalized notch and crack depth ($0.4 < a/\lambda_R < 1.0$) are likely a result of nonuniform (rough) discontinuity faces that exist in the experimental specimens and the variability inherent in experimental measurements. Variation in the cross-sectional shape of the discontinuity also affects T_{cr} for intermediate values of a/λ_R, as illustrated by differences in the BEM crack and notch predictions. (See Fig. 8.) Ideal discontinuities are assumed in the numerical analysis: a smooth face and a well-defined cross-section.

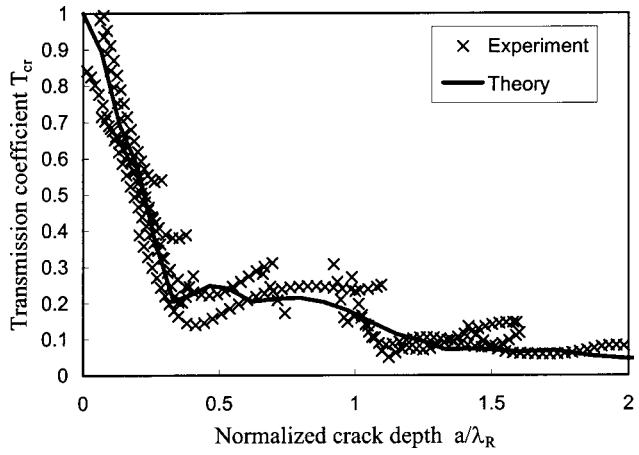


FIG. 10. The surface wave transmission coefficient across a surface-breaking crack obtained from the self-calibrating transmission measurement and the boundary element method.

Based on the obtained results, the self-calibrating surface wave transmission measurement technique shows excellent potential as a practical and reliable nondestructive method for the detection and sizing of *in-situ* surface-breaking cracks and notches in concrete structures. However T_{cr} obtained using the described self-compensating scheme is sensitive to changing crack depth and also a good estimate of the surface wave transmission across the notch or crack only when $T_{cr} > 0.4$, which corresponds to $a/\lambda_R < 0.3$. The following testing scheme is proposed based on this observed behavior. T_{cr} values across a range of frequency are obtained using the described self-calibrating measurement scheme on concrete containing a single surface-breaking crack. All T_{cr} values greater than 0.4 are retained and each value is mapped to the appropriate a/λ_R value using the numerically computed relation shown in Fig. 8, while retaining the frequency value associated with each datum. Note that the notch and crack responses (Fig. 8) are equivalent in this region. The surface wave velocity of the concrete is determined, and the wavelength (λ_R) is computed for each datum frequency. A crack depth estimate (a) from each datum is then computed. Multiple, redundant estimates of crack depth are thereby obtained from a single measurement. The effectiveness of this approach is being evaluated in ongoing research efforts.

ACKNOWLEDGMENTS

This paper was prepared in the course of research sponsored by the Federal Aviation Administration under research Grant No. 95-C-001 through the Center of Excellence for Airport Pavement Research.

[1] *CRC Handbook on Nondestructive Testing of Concrete*, edited by V. M. Malhotra and N. J. Carino (CRC Press, Boca Raton, 1991).
[2] J. H. Bungey and S. G. Millard, *Testing of Concrete in Structures*, 3rd edition (Blackie Academic & Professional, London, 1996).
[3] Y. Lin and W. Su, "Use of stress waves for determining the depth of surface-breaking cracks in concrete structures," ACI Mater. J. **93**, 494–505 (1996).
[4] M. Sansalone, J. Lin, and W. B. Streett, "Determining the depth of surface-opening cracks using impact-generated stress waves and time-of-flight technique," ACI Mater. J. **95**, 168–177 (1998).
[5] T. T. Wu, J. S. Fang, and P.-L. Liu, "Detection of the depth of a surface-breaking crack using transient elastic waves," J. Acoust. Soc. Am. **97**, 1678–1686 (1995).
[6] W. Song, J. S. Popovics, and J. D. Achenbach, "Crack depth determination in concrete slabs using wave propagation measurements," *Proceedings of the Federal Aviation Administration Technology Transfer Conference*, Atlantic City, NJ, April 1999.
[7] W. Suaris, and V. Fernando, "Ultrasonic pulse attenuation as a measure of damage growth during cyclic loading concrete," ACI Mater. J. **84**, 185–193 (1987).
[8] J. D. Achenbach, I. N. Komsky, Y. C. Lee, and Y. C. Angel, "Self-calibrating ultrasonic technique for crack depth measurement," J. Nondestruct. Eval. **11**, 103–108 (1992).
[9] G. Hévin, O. Abraham, H. A. Pedersen, and M. Campillo, "Characterization of surface cracks with Rayleigh waves: a numerical model," NDT & E Int. **31**, 289–298 (1998).
[10] J. S. Popovics, W. Song, and J. D. Achenbach, "A study of surface wave attenuation measurement for application to pavement characterization," *Structural Materials Technology III: An NDT Conference*, edited by R. D. Medlock and D. C. Laffey, Proc. SPIE **3400**, 300–308 (1998).

[11] J. S. Popovics, W. Song, M. Ghandehari, K. V. Subramaniam, J. D. Achenbach, and S. P. Shah, ''Application of wave transmission measurements for crack depth determination in concrete,'' ACI Mater. J. **97**, 127–135 (2000).
[12] ''Standard test method for compressive strength of cylindrical concrete specimens,'' (ASTM C39-94), *Annual Book of ASTM Standards*, 04.02, ASTM, West Coshohocken, PA.
[13] J. S. Popovics, W. Song, J. D. Achenbach, J. Lee, and R. F. Andre, ''One-sided stress wave velocity measurement in concrete,'' ASCE J. Eng. Mech. **124**, 1346–1353 (1998).
[14] ''Standard test method for pulse velocity through concrete'' (ASTM C597-83), *Annual Book of ASTM Standards*, 04.20, ASTM, West Coshohocken, PA.
[15] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, New York, 1973).
[16] R. Gettu, B. Mobasher, S. Carmona, and D. C. Jasen, ''Testing of the concrete under closed-loop control,'' J. Adv. Cement Based Materials **3**, 54–71 (1996).
[17] K. Creath, ''Phase measurement interferometry technique,'' in *Progress in Optics XXVI*, edited by E. Wolf (Elsevier, New York, 1988).
[18] M. Ghandehari, S. Krishnaswamy, and S. P. Shah, ''Technique for evaluating kinematics between rebars and concrete,'' ASCE, J. Eng. Mech. **125**, 234–241 (1999).
[19] J. D. Achenbach, L. M. Keer, and D. A. Mendelsohn, ''Elastodynamic analysis of an edge crack,'' J. Appl. Mech. **47**, 551–556 (1980).
[20] D. A. Mendelsohn, J. D. Achenbach, and L. M. Keer, ''Scattering of elastic waves by a surface-breaking crack,'' Wave Motion **2**, 277–292 (1980).
[21] Y. C. Angel and J. D. Achenbach, ''Reflection and transmission of obliquely incident Rayleigh waves by a surface-breaking crack,'' J. Acoust. Soc. Am. **75**, 313–319 (1984).
[22] D. E. Budreck and J. D. Achenbach, ''Scattering from three-dimensional planar cracks by the boundary integral equation method,'' J. Appl. Mech. **55**, 405–412 (1988).
[23] C. Zhang and J. D. Achenbach, ''Numerical analysis of surface wave scattering by the boundary element method,'' Wave Motion **10**, 365–374 (1988).
[24] Z. L. Li, J. D. Achenbach, I. Komsky, and Y. C. Lee, ''Reflection and transmission of obliquely incident surface waves by an edge of a quarter space: Theory and experiment,'' J. Appl. Mech. **59**, 349–355 (1992).
[25] C. Zhang and J. D. Achenbach, ''A new boundary integral equation formulation for elastodynamic and elastostatic crack analysis,'' J. Appl. Mech. **56**, 284–290 (1989).
[26] C. Zhang and J. D. Achenbach, ''Scattering by multiple crack configurations,'' J. Appl. Mech. **55**, 104–110 (1988).
[27] J. Dominguez, *Boundary Elements in Dynamics* (Computational Mechanics, Southhampton, 1993).

Numerical investigation and electro-acoustic modeling of measurement methods for the in-duct acoustical source parameters

Seung-Ho Jang[a] and Jeong-Guon Ih[b]
Center for Noise and Vibration Control, Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Science Town, Taejon 305-701, Korea

(Received 9 July 2001; revised 25 June 2002; accepted 3 July 2002)

It is known that the direct method yields different results from the indirect (or load) method in measuring the in-duct acoustic source parameters of fluid machines. The load method usually comes up with a negative source resistance, although a fairly accurate prediction of radiated noise can be obtained from any method. This study is focused on the effect of the time-varying nature of fluid machines on the output results of two typical measurement methods. For this purpose, a simplified fluid machine consisting of a reservoir, a valve, and an exhaust pipe is considered as representing a typical periodic, time-varying system and the measurement situations are simulated by using the method of characteristics. The equivalent circuits for such simulations are also analyzed by considering the system as having a linear time-varying source. It is found that the results from the load method are quite sensitive to the change of cylinder pressure or valve profile, in contrast to those from the direct method. In the load method, the source admittance turns out to be predominantly dependent on the valve admittance at the calculation frequency as well as the valve and load admittances at other frequencies. In the direct method, however, the source resistance is always positive and the source admittance depends mainly upon the zeroth order of valve admittance. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1504850]

PACS numbers: 43.20.Mv, 43.20.Ye, 43.50.Gf [DEC]

I. INTRODUCTION

For the acoustic analysis of intake or exhaust systems in fluid machines, the linear acoustic theory is usually used in the frequency domain. The acoustical source is characterized by two complex quantities, viz., the source strength and the source impedance, by assuming a linear and time-invariant source condition. Because the sound generating mechanisms of most fluid machine sources are too much complex, it is quite hard to obtain the source parameters by analytical means only. Consequently, source parameters are generally obtained by either experimental methods or numerical simulations. Experimental methods can be classified into direct and indirect methods according to whether or not an additional external source is utilized. Many previous works[1–3] record the accurate characterization of linear sources such as loudspeakers and blowers. However, it was reported that the indirect measurement technique yields negative source resistances in the application to the intake or exhaust system of compressors and IC-engines, although the predicted insertion loss and sound pressure level are in fairly good agreement with experimental results. Ross and Crocker[4] and Prasad and Crocker[5] tried to measure the source impedance of the exhaust of an eight-cylinder engine by using the direct method. They found that the source impedance is similar to the impedance of an anechoic source model. Bodén[2] and Albertson and Bodén[6] measured the source impedance of exhaust and intake systems of an engine by using the load method and they observed source impedance with a negative real part. Desmons *et al.*[7] also applied the load method in measuring the source characteristics of the second and fourth harmonics of a four-cylinder engine. In contrast to Bodén's result, they obtained source impedance with a positive real part in most of engine speed range. Gupta and Munjal[8] numerically simulated an engine exhaust by using the method of characteristics and they got negative source resistance values. Bodén[9] studied a model of the modified compressor having time-varying volume and reported that the nonlinear and time variant terms could not be neglected if an accurate description is to be obtained for all possible acoustic loads. Albertson[10] analyzed the steady state properties of a forced simple-harmonic damped oscillator and showed that negative resistances can easily be created by a simple nonlinearity. Ih and Peat[11] gave a general overview on this matter with an emphasis on the role of the time-varying nature of the source. Peat and Ih[12] also pointed out that the time-varying nature of the source is the major cause of a negative resistance, in which a very simplified and idealized source-load system was employed. It is theoretically known that one can end up with the same source characteristics by applying the direct method and the load method to a linear time-invariant source.[13] However, it is rare to see a comparison of application results of two methods to a time-varying fluid machine, mainly due to the difficulties in implementing the direct method.[2,7]

In this paper, it is clearly shown that the time-varying nature of fluid machines is the most probable origin of the

[a]Currently working at the PSR Group, Nuclear Power Laboratory, Korea Electric Power Research Institute, Science Town, Taejon 305-380, Korea.
[b]Author to whom correspondence should be addressed. Electronic mail: ihih@sorak.kaist.ac.kr

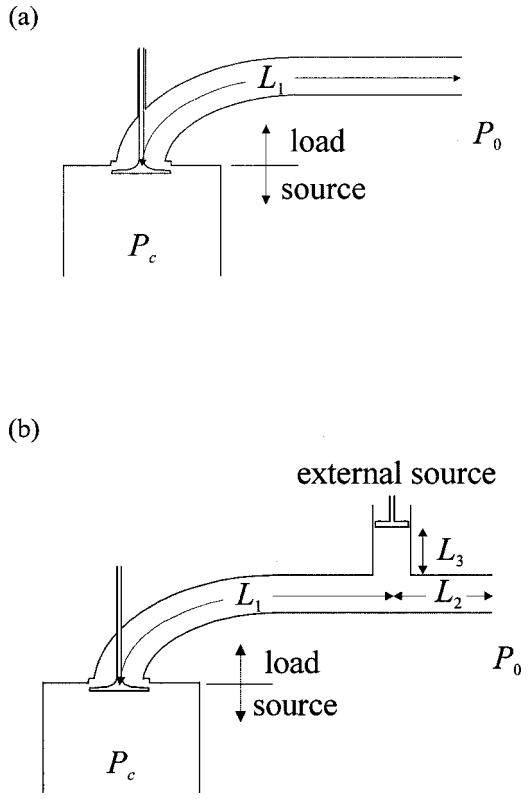


FIG. 1. A schematic layout of the reservoir–valve–pipe system for the simulation of measurement methods for source properties. (a) Load method, (b) direct method.

aforementioned looking-unphysical phenomenon. For this purpose, a simplified fluid machine consisting of a reservoir, a valve, and a pipe is considered as representing a typical periodic, time-variant system and the measurement techniques including direct and load methods are simulated using the time domain numerical analysis and the electro-acoustic analogy. In the numerical simulation, the method of characteristics is employed and the source characteristics are calculated from the resultant data. Then, equivalent circuits for the simulations are analyzed by considering the system as a linear time-variant source. Emphasis is given to the general trend in source characteristics, not to the precise prediction of radiated sound pressure level.

II. TIME DOMAIN NUMERICAL SIMULATION

A. Simulation method

Figure 1 shows a schematic layout of an acoustic system comprised of a reservoir, a valve, and a pipe for the simulation of experimental characterization techniques. The source represents a simplified one port system of a compressor or an internal combustion engine, in which an exhaust pipe is connected to a cylinder. It is assumed that the mass change in the cylinder interior due to the gas scavenging through the valve is negligible. Then, the cylinder can be regarded as a reservoir having a constant internal pressure P_c. The initial pressure and temperature in the pipe were set to the corresponding atmospheric values, i.e., 1 bar and 300 K, and the term P_0 denotes the atmospheric pressure. Changes of valve open area due to its lift are shown in Fig. 2, which are similar to

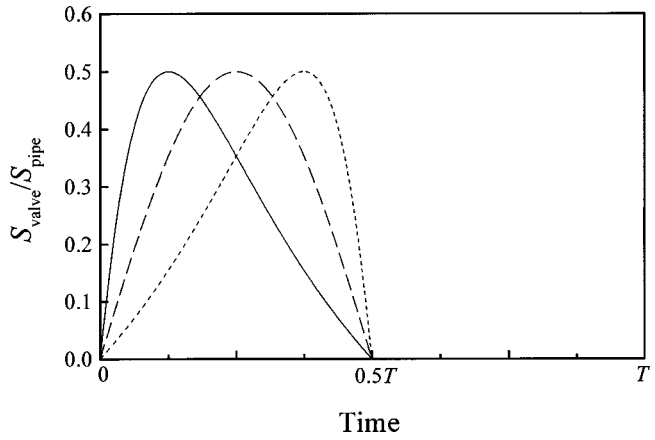


FIG. 2. The valve area diagram in which T denotes a period and S the sectional area: ——, curve 1; — — —, curve 2; ···, curve 3. The valve is completely shut off during the latter half of a cycle.

the valve motion in an actual fluid machine, but are more simplified. Curve 2 is the positive locus of a sine curve such as

$$\left.\frac{S_{\text{valve}}(t)}{S_{\text{pipe}}}\right|_{\text{curve 2}} = \begin{cases} 0.5\sin(2\pi t/T), & \text{if } \sin(2\pi t/T) \geqslant 0, \\ 0, & \text{if } \sin(2\pi t/T) < 0. \end{cases} \tag{1}$$

Here, S_{valve} and S_{pipe} are the areas of the valve throat and pipe section, respectively. The period T of the valve opening was taken as 0.025 s. Curves 1 and 3 are given as

$$\left.\frac{S_{\text{valve}}(t)}{S_{\text{pipe}}}\right|_{\text{curve 1}} = \begin{cases} \left.\dfrac{S_{\text{valve}}(t_1)}{S_{\text{pipe}}}\right|_{\text{curve 2}}, & \text{if } nT \leqslant t < (n+0.5)T, \\ 0, & \text{if } (n+0.5)T \leqslant t < (n+1)T, \end{cases} \tag{2}$$

and

$$\left.\frac{S_{\text{valve}}(t)}{S_{\text{pipe}}}\right|_{\text{curve 3}} = \begin{cases} \left.\dfrac{S_{\text{valve}}(t_2)}{S_{\text{pipe}}}\right|_{\text{curve 2}}, & \text{if } nT \leqslant t < (n+0.5)T, \\ 0, & \text{if } (n+0.5)T \leqslant t < (n+1)T, \end{cases} \tag{3}$$

where n is 0 or positive integer and

$$t_1 = \frac{3Tt}{4t+T}, \tag{4a}$$

$$t_2 = \frac{Tt}{3T-4t}. \tag{4b}$$

The source-load interface was positioned at the point just after the valve aperture in the downstream pipe. The load set for applying the indirect method consisted of two open pipes with different lengths.

It is assumed that the flow is one-dimensional and the medium is an ideal gas with constant specific heats. For computational conveniences, all flow fields are assumed to have

a constant entropy level, i.e., homentropic. For a uniform duct under the foregoing conditions, the continuity equation is given by

$$\frac{1}{\rho}\frac{\partial \rho}{\partial t}+\frac{u}{\rho}\frac{\partial \rho}{\partial x}+\frac{\partial u}{\partial x}=0, \tag{5}$$

and the momentum equation can be expressed as

$$\frac{1}{\rho}\frac{\partial p}{\partial x}+\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=0. \tag{6}$$

The speed of sound a for an ideal gas can be defined as

$$a^2=\gamma\frac{p}{\rho}, \tag{7}$$

where γ means the specific heat ratio of the medium. Pressure and density in the isentropic process would satisfy the following relations:

$$\frac{p}{p_{\mathrm{ref}}}=\left(\frac{a}{a_{\mathrm{ref}}}\right)^{2\gamma/(\gamma-1)}, \tag{8a}$$

$$\frac{\rho}{\rho_{\mathrm{ref}}}=\left(\frac{a}{a_{\mathrm{ref}}}\right)^{2/(\gamma-1)}. \tag{8b}$$

Here, p_{ref} and ρ_{ref} denote the pressure and density, respectively, at a reference condition. From the direction and compatibility equations,[14] Riemann variables λ, β can be defined as

$$\lambda=\frac{a}{a_{\mathrm{ref}}}+\frac{\gamma-1}{2}\frac{u}{a_{\mathrm{ref}}}, \tag{9a}$$

$$\beta=\frac{a}{a_{\mathrm{ref}}}-\frac{\gamma-1}{2}\frac{u}{a_{\mathrm{ref}}}. \tag{9b}$$

The mesh method was employed in the calculation of characteristics: The geometric range for calculation was divided by a uniform grid and λ, β were calculated at each grid point at a time corresponding to each time step. In this study, all calculations used the same grid size of 0.02 m that assures a satisfactory accuracy. Further reduction in grid size was not effective in enhancing the output result. The time interval for calculation was variably given by the Courant–Friedrichs–Levy stability criterion.[14] Because a periodic change of acoustic pressure and particle velocity occurs at about 0.05 s after initial transients, the pressure responses after 1 s were used in the calculations. The flow through the valve was described by the ''constant pressure model,''[14] in which an isentropic expansion process is assumed for the flow from the inner cylinder at stagnation condition to the valve throat, whereas an adiabatic expansion process is assumed for the flow from the valve throat into the pipe. Although the pressure drop through the valve is a nonlinear function of the velocity, the nonlinear effect to the total system response is not significant because small values are assigned to the relative cylinder pressures. The boundary condition for the open end of pipe was the time-domain radiation impedance, which can be estimated from the frequency-domain impedance by iterative calculation.[6,15]

In the simulation, the two-load technique[8,16] was adopted as representing the indirect measurement method, which employs two open pipes with different lengths, $L_1=1.0$ or 1.6 m, as loads. The two-load method requires a complex pressure data for each load, i.e., both magnitude and phase as input data. Although the average velocity is different for each load, the data fluctuation associated with the velocity variation is negligible because the difference is typically less than 0.08 M, in which M denotes the Mach number.

In the simulation of the direct measurement method, the source impedance is determined by exciting the source by an external acoustic source.[1,4,5] In the simulation, a constant velocity source was positioned at the end of a pipe, 0.5 m in length, which was attached perpendicularly to the main pipe as shown in Fig. 1(b). At the junction where three pipes join together, the conditions of the continuities of pressure, density and entropy, and the conservation of mass were utilized in the numerical calculations.[14] The pipe lengths, L_1 and L_2, in Fig. 1(b) were 1.0 and 0.5 m, respectively. The maximum amplitude of piston velocity of the external, constant velocity source was set to 0.01 m/s which was chosen to have greater strength than the original source of the reservoir–valve–pipe system. The direct method implements the wave decomposition technique in calculating the input data. In using the linear acoustic theory for frequencies lower than the cutoff frequency of the first cross mode, acoustic pressure and particle velocity can be obtained by addition and subtraction of two plane waves propagating in opposite directions. In using the nonlinear theory, total pressure and velocity cannot be simply obtained as in the linear theory, although it is still true that the acoustic field in the duct consists of two plane waves propagating in opposite directions. When two waves interact with each other, the propagation speed changes continuously and it is not possible to determine a coincident point and time for two waves. Therefore, the sound field due to the superposition of two waves cannot be determined directly. Payri *et al.*[15] showed that the nonlinear wave decomposition is possible for the method of characteristics, which tracks the propagation of a ''wave point'' in time and space. In this study, such a nonlinear wave decomposition technique is used in the simulation.

B. Results of simulation

As well as the difference between direct and load methods, the effects of two major dependent parameters are investigated: cylinder pressure and valve open-area function. The calculated source characteristics from the simulation using the two-load method are shown in Figs. 3 and 4. Figure 3 shows the effect of changing the cylinder pressure P_c on the source parameters, in which the valve open-area profile followed curve 2 in Fig. 2. Here, the first order corresponds to the fundamental frequency, i.e., 40 Hz. The source strength is nearly proportional to the difference in cylinder pressure, P_c-P_0, at low order frequencies. It is noted that the peaks in source resistance, which can be considered antiresonances, are quite sensitive to the variation of cylinder pressure. Figure 4 shows the effect of changing the valve open-area profile when the cylinder pressure is given by $P_c-P_0=10^2$ Pa. The source strength values vary substantially for all orders except orders lower than the third order. The

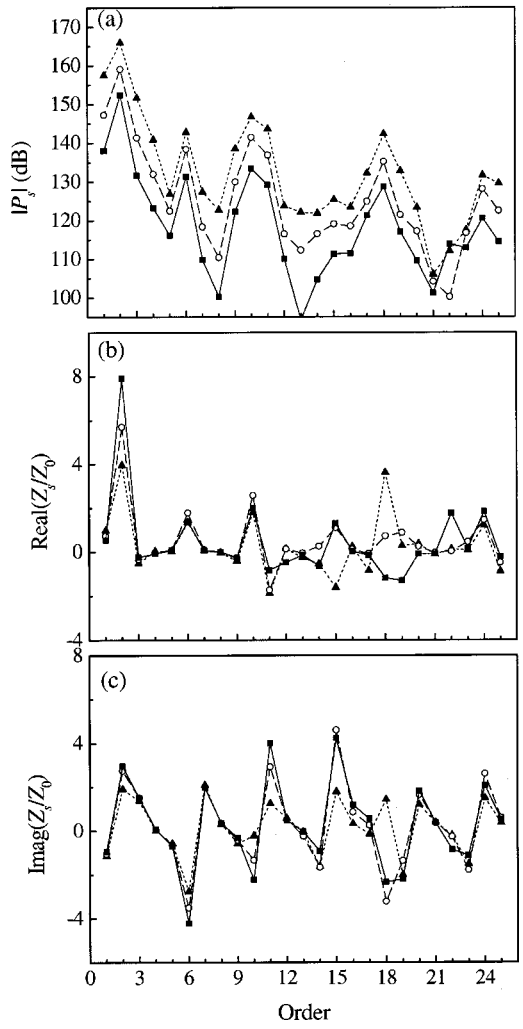


FIG. 3. Calculated source parameters by using the two-load method with changing the cylinder pressure (P_c-P_0), in which curve 2 is used as the valve open-area profile: —■—, 10^2 Pa; --○--, 3×10^2 Pa; ···▲···, 10^3 Pa. (a) Source strength, (b) source resistance, (c) source reactance.

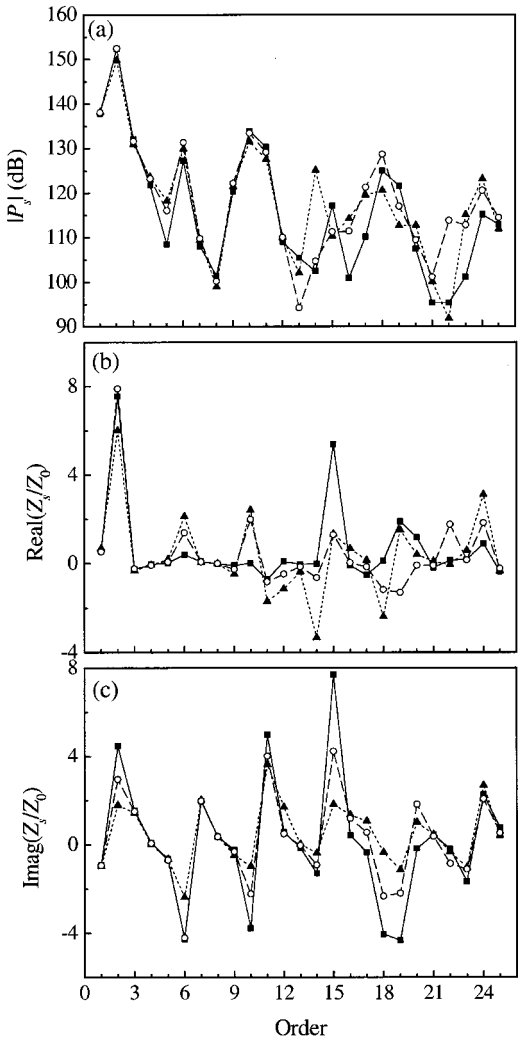


FIG. 4. Calculated source parameters by using the two load method with changing the valve open-area profile ($P_c-P_0=100$ Pa): —■—, curve 1; --○--, curve 2; ···▲···, curve 3. (a) Source strength, (b) source resistance, (c) source reactance.

source impedance is very much sensitive at the peaks in source resistance as the previous case. It is noted that the source resistance is negative at many orders in both of Figs. 3 and 4.

Figures 5 and 6 show the calculated source impedances from the simulation using the direct method. Calculations were carried out by sweeping the frequency. It should be mentioned that the harmonics and subharmonics generated from the excitation of a frequency can be added to the corresponding frequencies due to valve motion. However, the basic concept and premise of direct method is that the strength of external excitation should be much greater than that of the actual source. Therefore, the source impedance could be approximately determined by using the response at the excitation frequency only by neglecting the harmonic and subharmonic components. In Fig. 5, the valve open-area profile also followed curve 2 in Fig. 2 and the result in the case of $P_c-P_0=10^3$ Pa cannot be shown here because an external source having much higher strength is required in this case, which can cause a large nonlinear effect. Instead, a pressure difference $P_c-P_0=10$ Pa is considered in Fig. 5, which represents a very small cylinder pressure as compared with the external source pressure. In Fig. 6, the cylinder pressure is given as $P_c-P_0=10^2$ Pa. As evident from Figs. 5 and 6, the calculated source impedances are quite different from those obtained by the load method. The direct method results in only positive source resistance values for all orders considered. It can be also observed that the direct method is quite robust to the changes of cylinder pressure and valve open-area profile defined in Fig. 2. In Fig. 5, it seems that the strength of external excitation is not sufficiently large in the case of $P_c-P_0=3\times10^2$ Pa, so that the corresponding source impedance at some orders is slightly different from the values obtained for low cylinder pressures.

In Fig. 7, the sound pressure levels predicted by using the source parameters from the two methods at the downstream side of valve aperture are compared with the result from a direct application of the characteristic method, in which an open pipe, 1.3 m in length, was applied as load. For comparison purposes, a constant pressure source (zero impedance) and a constant velocity source (infinite impedance) were additionally considered. The cylinder pressure was given as $P_c-P_0=3\times10^2$ Pa and the valve open-area profile

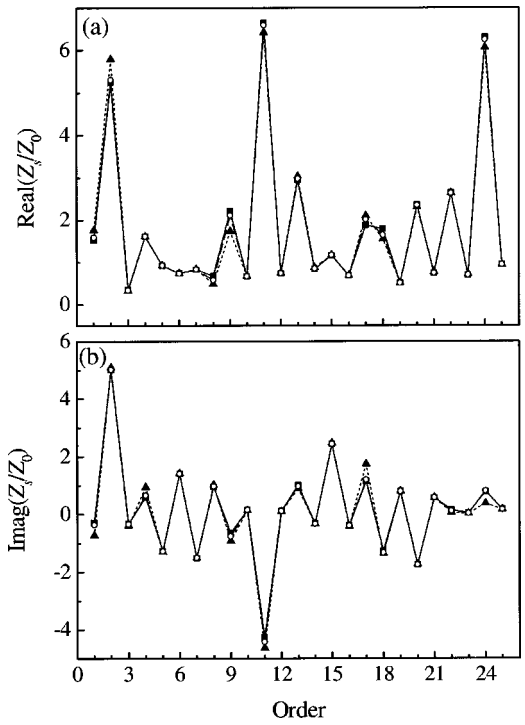


FIG. 5. Calculated source impedances by using the direct method with changing the cylinder pressure ($P_c - P_0$), in which curve 2 is used as the valve open-area profile: —■—, 10 Pa; --○--, 10^2 Pa; ···▲···, 3×10^2 Pa. (a) Source resistance, (b) source reactance.

followed curve 2. Predictions made by the direct method and the constant pressure and velocity source assumptions employed the source strength value which was determined from the pressure data by applying an open pipe load, 1.0 m in length. One can find that the load and direct methods give fairly good predictions in spite of the fact that different

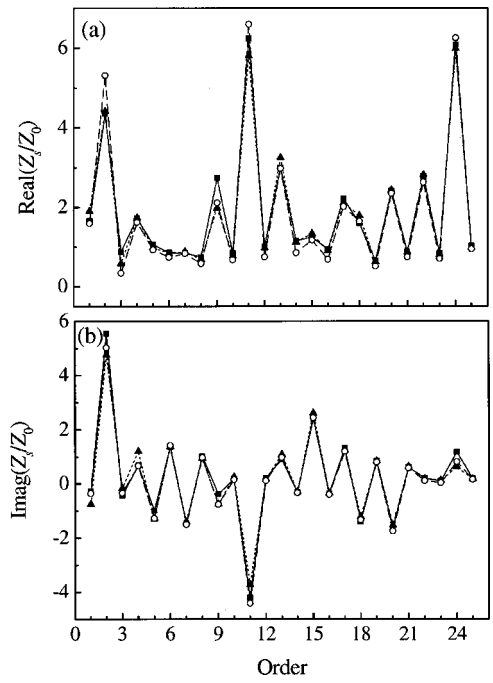


FIG. 6. Calculated source parameters by using the direct method with changing the valve open-area profile ($P_c - P_0 = 100$ Pa): —■—, curve 1; --○--, curve 2; ···▲···, curve 3. (a) Source resistance, (b) source reactance.

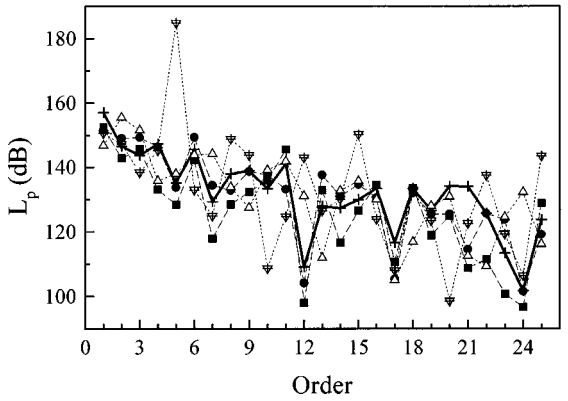


FIG. 7. A comparison of predicted sound pressure levels at the downstream side of the valve aperture by using the estimated source parameters ($P_c - P_0 = 300$ Pa, valve open-area profile=curve 2): --■--, load method [$(L_p)_{\text{overall}} = 154.8$ dB]; --●--, direct method [$(L_p)_{\text{overall}} = 156.8$ dB]; ···△···, constant pressure source assumption [$(L_p)_{\text{overall}} = 161.5$ dB]; ···▽···, constant velocity source assumption [$(L_p)_{\text{overall}} = 184.9$ dB]; -·+·-, calculated by the method of characteristics [$(L_p)_{\text{overall}} = 158.6$ dB].

source parameters obtained by two methods are used in the calculations. However, the constant pressure source model yields relatively large errors at many orders and the constant velocity source model gives very poor prediction. It shows that actual measurement of source parameters is essential for the realistic prediction of the radiated sound pressure level accurately. Overall sound pressure levels in Fig. 7 are 154.8 dB in the load method, 156.8 dB in the direct method, 161.5 dB in constant pressure source assumption, 184.9 dB in constant velocity source assumption, and 158.6 dB in the method of characteristics, while the direct method gives slightly better result than the load method in this work. Such a difference in overall level obtained by the load method would become smaller when the source-load interface is far from the valve aperture and an additional element, e.g., catalytic converter, exists as in many actual measurement situations.[3,6]

III. ELECTRO-ACOUSTIC MODELING FOR FREQUENCY DOMAIN ANALYSIS

In the numerical simulations of the foregoing section, nonlinear flow equations have been used. However, when the cylinder pressure is very low, the whole system can be approximated as a linear system and the flow through the valve is to be expressed by linear time-variant equations. Based on this assumption, the effect of the time-varying nature of the source on the measurement of source parameters is to be investigated. Because it is not possible to obtain the explicit analytical solution of the time-varying systems, only the qualitative analysis is carried out with some restrictions in this section. Although it is an approximate approach, the analysis would be useful in clarifying the causes of the negative source resistances which are often coming out in the measurement of actual fluid machines.

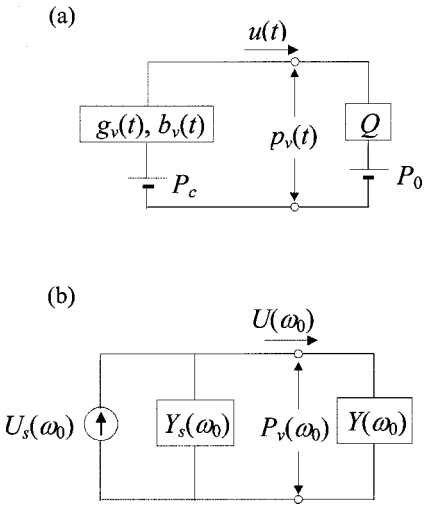


FIG. 8. Equivalent circuits for the simulation of load method. (a) Time domain circuit, (b) source-load representation in frequency domain.

A. Simulation of the load method

By using the electro-acoustic analogy, the simplified acoustic system in Fig. 1(a) representing the measurement condition of the load method can be expressed with an equivalent circuit in Fig. 8(a). The flow through the valve can be modeled as an admittance, that is, it can be described by conductance $g_v(t)$ and susceptance $b_v(t)$. The volume velocity through the valve in the direction of downstream pipe, $u(t)$, is given by

$$u(t)=g_v(t)p(t)+\frac{d}{dt}[b_v(t)p(t)], \tag{10}$$

where $p(t)\equiv P_c-p_v(t)$ and $p_v(t)$ denotes the pressure in the downstream side of valve aperture. It should be mentioned that $g_v(t)$ and $b_v(t)$ are real functions and are subjected to a periodic time-variation because of the periodic valve motion. If Fourier cosine series is used, $g_v(t)$ and $b_v(t)$ are considered as even functions mathematically, whereas, by employing Fourier sine series, they are considered as odd functions. Since the steady state solutions are to be discussed here, the condition at $t<0$ and the initial condition are not important. Therefore, we can use either cosine or sine series or both. Then, each term can be expanded into either Fourier cosine for $t\geqslant 0$ as

$$g_v(t)=\sum_{n=0}^{\infty} G_v(n\omega_0)\cos n\omega_0 t, \tag{11a}$$

$$b_v(t)=\sum_{n=0}^{\infty} B_v(n\omega_0)\cos n\omega_0 t, \tag{11b}$$

where ω_0 is the fundamental angular frequency of the time variation. Even if the Fourier sine series is only used instead of the cosine series, the trend of the results or conclusions would be still same. It is difficult to obtain accurate values of $g_v(t)$ and $b_v(t)$ in actual operating IC-engines, either analytically or experimentally. If the valve is time invariant and the fluctuation of flow is time harmonic, i.e., $e^{j\omega_n t}$, the operator d/dt can be replaced by $j\omega_n$, so that g_v and b_v are related to the real and imaginary parts of the valve admittance in the frequency domain. As shown in Fig. 2, the valve is completely closed over the latter half portion of a cycle, at which time $u(t)$ will be zero and the valve impedance goes to infinity. For this reason, the admittance is the preferable descriptor of the system as compared to the impedance. When using the load admittance $Y(\omega)$ representing a downstream pipe and termination, the acoustic pressure and the volume velocity at the interface are related as

$$u(t)=Q(t)(p_v(t)-P_0)=Q(t)(P_c-p(t)-P_0). \tag{12}$$

Here, $Q(t)$ is a time-varying coefficient relating load pressure and velocity, which is defined by using the load admittance $Y(\omega_n)$ for a specific frequency ω_n as[17]

$$Q(t)\hat{p}(t)=\sum_{n=-\infty}^{\infty} Y(\omega_n)\hat{P}(\omega_n)\exp(j\omega_n t), \tag{13}$$

where

$$\hat{p}(t)=\sum_{n=-\infty}^{\infty} \hat{P}(\omega_n)\exp(j\omega_n t) \quad (n=\text{integer}). \tag{14}$$

Then, Eq. (10) can be rewritten as

$$b_v(t)\frac{dp(t)}{dt}+\left[g_v(t)+\frac{db_v(t)}{dt}+Q(t)\right]p(t) = Y(0)(P_c-P_0). \tag{15}$$

For a system described by Eq. (15) having periodically time-varying coefficients and a constant forcing term, the steady state solution $p(t)$ will be given by[17,18]

$$p(t)=\sum_{n=-\infty}^{\infty} P(n\omega_0)\exp(jn\omega_0 t). \tag{16}$$

Substituting Eq. (16) into Eq. (15) and expanding for each frequency, each coefficient of $e^{jn\omega_0 t}$ can be obtained by comparing both sides of resulting equation. When the zero-frequency is considered, one can obtain

$$Y(0)(P_c-P_0)=\ldots+\tfrac{1}{2}G_v(\omega_0)P(-\omega_0)+[Y(0)+G_v(0)]P(0)+\tfrac{1}{2}G_v(\omega_0)P(\omega_0)+\ldots, \tag{17}$$

and, for the frequency components of $n\omega_0 (n\geqslant 1)$, it follows that

$$0=[Y(n\omega_0)+G_v(0)+jn\omega_0 B_v(0)]P(n\omega_0) +\frac{1}{2}\sum_{m=1}^{\infty}[G_v(m\omega_0)+jn\omega_0 B_v(m\omega_0)] \times[P((n-m)\omega_0)+P((n+m)\omega_0)]. \tag{18}$$

In actual measurements using the load method, a number of loads are applied to the source and then the acoustic pressures due to multiple acoustic loadings are measured to construct an overdetermined equation of the inverse problem. In this manner, one can determine the source parameters based on the response data due to loads with *a priori* known impedance Z or admittance Y. The linear time-invariant model for one-port sources is often expressed in terms of the source

strength $P_s(\omega)$ and the source impedance $Z_s(\omega)$ for the frequency ω being nonzero:

$$P_s(\omega)-Z_s(\omega)U(\omega)=Z(\omega)U(\omega). \tag{19}$$

In terms of the source strength $U_s(\omega)$ and the source admittance $Y_s(\omega)$ as shown in Fig. 8(b), it follows that

$$U_s(\omega)-Y_s(\omega)P_v(\omega)=Y(\omega)P_v(\omega), \tag{20}$$

or

$$-U_s(\omega)-Y_s(\omega)P(\omega)=Y(\omega)P(\omega), \tag{21}$$

where $P_v(\omega)$ is the spectrum of $p_v(t)$, $U_s(\omega)=P_s(\omega)/Z_s(\omega)$, and $Y_s(\omega)=1/Z_s(\omega)$. Equation (21) is obtained from Eq. (20) by using the relation of $p(t)=P_c-p_v(t)$ or $P(\omega)=-P_v(\omega)$.

Because Eqs. (17) and (18) involve infinite number of equations, each with an infinite number of terms, it is not generally possible to obtain the source parameters in explicit form. However, by making some restrictions or approximations, one can derive the source parameters analytically. A very simple one of such restrictions is to specify that the steady component of pressure, i.e., zero-frequency component, is predominant and other components are negligible. Then, the source strength at any harmonic is zero which can be considered a trivial case.

If two frequency components, zero and ω_0, of the pressure are assumed to be dominant compared to other frequency components, the following relation can be obtained by using Eqs. (17), (18) and (21):

$$\begin{aligned}-U_s(\omega_0)-Y_s(\omega_0)P(\omega_0)=Y(\omega_0)P(\omega_0)&=-\tfrac{1}{2}[G_v(\omega_0)+j\omega_0B_v(\omega_0)]P(0)-[G_v(0)+j\omega_0B_v(0)]P(\omega_0)\\&=-\frac{1}{2}\frac{Y(0)(P_c-P_0)[G_v(\omega_0)+j\omega_0B_v(\omega_0)]}{Y(0)+G_v(0)}\\&\quad-\left\{G_v(0)+j\omega_0B_v(0)-\frac{G_v(\omega_0)[G_v(\omega_0)+j\omega_0B_v(\omega_0)]}{4[Y(0)+G_v(0)]}\right\}P(\omega_0).\end{aligned} \tag{22}$$

Expressing the zero-frequency load admittance as $Y(0)=G(0)+jB(0)$, the source parameters can be given by

$$U_s(\omega_0)=\frac{1}{2}\frac{Y(0)(P_c-P_0)[G_v(\omega_0)+j\omega_0B_v(\omega_0)]}{G(0)+G_v(0)+jB(0)}, \tag{23}$$

$$G_s(\omega_0)=G_v(0)-\frac{[G_v(\omega_0)]^2[G(0)+G_v(0)]+\omega_0G_v(\omega_0)B_v(\omega_0)B(0)}{4\{[G(0)+G_v(0)]^2+[B(0)]^2\}}, \tag{24}$$

$$B_s(\omega_0)=\omega_0B_v(0)-\frac{\omega_0G_v(\omega_0)B_v(\omega_0)[G(0)+G_v(0)]-[G_v(\omega_0)]^2B(0)}{4\{[G(0)+G_v(0)]^2+[B(0)]^2\}}. \tag{25}$$

Here, $G_s(\omega_0)$ and $B_s(\omega_0)$ are the real and imaginary parts, respectively, of the source admittance. One can find that the source strength in Eq. (23) is proportional to the difference between the cylinder pressure and the atmospheric pressure. This trend is consistent with the result at low orders in Fig. 3(a). Source parameters depend on the valve impedance at the calculation frequency as well as the valve and load impedances at other frequencies. It is noted that the real parts of the source admittance and its inverse, the source impedance, can be negative when the second term on the right side of Eq. (24) is larger than the first term. It is noted that the negative resistances were obtained in the numerical simulations of the load method as shown in Figs. 3 and 4.

In case that three frequency components, zero, ω_0, and $2\omega_0$, of the pressure are assumed to be dominant, the source parameters are obtained by using Eqs. (17), (18) and (21) as

$$U_s(\omega_0)=\frac{(1-2A_2)}{2(1-A_1A_2)}\frac{Y(0)(P_c-P_0)[G_v(\omega_0)+j\omega_0B_v(\omega_0)]}{Y(0)+G_v(0)}, \tag{26}$$

$$Y_s(\omega_0)=[G_v(0)+j\omega_0B_v(0)]-\frac{(1-2A_2)\ [G_v(\omega_0)-G_v(2\omega_0)A_3][G_v(\omega_0)+j\omega_0B_v(\omega_0)]}{4(1-A_1A_2)\qquad[Y(0)+G_v(0)]}-\frac{A_3}{2}[G_v(\omega_0)+j\omega_0B_v(\omega_0)], \tag{27}$$

where

$$A_1=\frac{G_v(2\omega_0)}{Y(0)+G_v(0)}, \tag{28a}$$

$$A_2=\frac{G_v(2\omega_0)+j2\omega_0 B_v(2\omega_0)}{4[Y(2\omega_0)+G_v(0)+j2\omega_0 B_v(0)]}, \tag{28b}$$

$$A_3=\frac{G_v(\omega_0)+j2\omega_0 B_v(\omega_0)}{2[Y(2\omega_0)+G_v(0)+j2\omega_0 B_v(0)]}. \tag{28c}$$

If additional frequency components of pressure are further included, which might be non-negligible, one can again construct an equivalent circuit having the more separate subcircuits that permits the derivation of the corresponding source characteristics. Source parameters at frequency $n\omega_0$ can be approximately obtained in a similar way. The foregoing source parameters are defined when the source-load interface is at the downstream side of the valve aperture. One can use these source parameters in calculating the source values at any arbitrary source-load interface position in the pipe, which can be done by utilizing the four-pole parameters of the tube between valve aperture and new interface section. However, it should be mentioned that the aforementioned trend of the source resistance remains true at any other arbitrary section in the pipe.

B. Simulation of the direct method

The equivalent circuit of the simple acoustic system for the simulation of the direct method is shown in Fig. 9(a). For the implementation of the direct method, an additional external source is required, which has the strength U_e operating at frequency ω_e. The previous analysis method for simulating the load method can be applied in a similar manner for analysis of the direct method. Here, the direction of velocity $u(t)$ is defined opposite to that in the previous section because the admittance or impedance of the effectively passive source termination is measured. $p(t)$ is defined as $p_v(t)-P_c$ so that Eq. (10) still holds here. Then, an equivalent equation to Eq. (12) can be written as

$$-u(t)=-U_e\exp(j\omega_e t)+Q(t)(p(t)+P_c-P_0). \tag{29}$$

From Eqs. (10) and (29), the following relation can be easily derived:

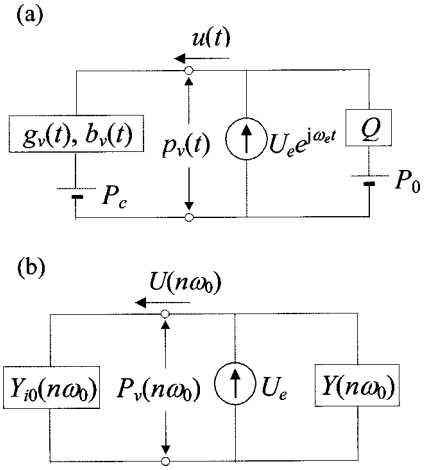


FIG. 9. Equivalent circuits for the simulation of direct method. (a) Time domain circuit, (b) frequency domain circuit.

$$b_v(t)\frac{dp(t)}{dt}+\left[g_v(t)+\frac{db_v(t)}{dt}+Q(t)\right]p(t)$$
$$=U_e\exp(j\omega_e t)-Y(0)(P_c-P_0). \tag{30}$$

Because the external source should be much stronger than the tested source, i.e., $U_e \gg Y(0)(P_c-P_0)$, Eq. (30) represents a system having periodically time-varying coefficients and a sinusoidal forcing function.[17,18] The steady state solution $p(t)$ can be given by

$$p(t)=\sum_{n=-\infty}^{\infty} P(\omega_e+n\omega_0)\exp[j(\omega_e+n\omega_0)t]. \tag{31}$$

Substituting Eq. (31) into Eq. (30) and expanding for each frequency, the following relation can be obtained for a frequency at ω_e:

$$U_e=\ldots+\tfrac{1}{2}[G_v(\omega_0)+j\omega_e B_v(\omega_0)]P(\omega_e-\omega_0)$$
$$+[Y(\omega_e)+G_v(0)+j\omega_e B_v(0)]P(\omega_e)$$
$$+\tfrac{1}{2}[G_v(\omega_0)+j\omega_e B_v(\omega_0)]P(\omega_e+\omega_0)+\ldots. \tag{32}$$

Because the sound field from the external source operating at ω_e in frequency is dominant, the pressure components at other frequencies can be neglected and Eq. (32) can be approximately reduced to

$$U_e\approx[Y(\omega_e)+G_v(0)+j\omega_e B_v(0)]P(\omega_e). \tag{33}$$

Source parameters are to be determined at $n\omega_0$ and, thus, the exciting frequency ω_e will be changed according to the frequency of interest. Then, the equivalent circuit for frequency domain analysis as shown in Fig. 9(b) can be considered and the input admittance Y_{i0} can be regarded as the source admittance as follows:

$$Y_s(n\omega_0)=Y_{i0}(n\omega_0)=\frac{U_e-P_v(n\omega_0)Y(n\omega_0)}{P_v(n\omega_0)}$$
$$=\frac{U_e-P(n\omega_0)Y(n\omega_0)}{P(n\omega_0)}$$
$$\approx G_v(0)+jn\omega_0 B_v(0). \tag{34}$$

The source admittance depends mainly upon the valve admittance at the zeroth order and is not related to the cylinder pressure at all. The direct method gives the results which are less sensitive to the change of harmonic components of open-area valve profile. This trend can be also seen in Sec. II. The source conductance or the source resistance is always positive in Eq. (34), which is also true in Figs. 5 and 6. If the valve is almost time invariant, the source admittances obtained by the load and direct methods in Eqs. (24), (25), (27) and (34) are identical as $Y_s(\omega_0)\approx G_v(0)+j\omega_0 B_v(0)$. This is because $G_v(n\omega_0)\approx 0$, $B_v(n\omega_0)\approx 0$ for $n\geqslant 1$ and $g_v(t)\approx G_v(0)$, $b_v(t)\approx B_v(0)$. In this case, however, the source strength $U_s(\omega_0)$ becomes almost zero in Eqs. (23) and (26), because the pulsating flow can not be generated by such valve motion which can be considered a trivial case.

IV. CONCLUDING REMARKS

Measurement techniques for the in-duct source parameters were simulated with a demonstration example of a simplified exhaust system composed of a reservoir, a valve, and a pipe. The valve included in the system is time varying and the whole system was modeled by both the time domain numerical method and the electro-acoustic analogous circuit. In the numerical simulation, the method of characteristics was employed for calculating the pressure in the pipe, and then the source parameters were calculated by the two-load method and the direct method. In the load method, a negative source resistance is obtained for some frequencies and the source impedance is sensitive to the change of the cylinder pressure and the valve profile at the peaks in the source resistance. On the other hand, the direct method gives positive source resistances and the impedance is found to be very insensitive to the change of cylinder pressure and valve profile. The equivalent circuits for simulation of the two methods were also analyzed by considering the simple exhaust system that represents the linear time-varying source. In the load method, the source strength is proportional to the difference between the cylinder pressure and the atmospheric pressure, while the source parameters are correlated with the valve and load impedances at the initial calculation frequency and other harmonic frequencies. However, the direct method yields the source resistance being always positive and the source admittance depends mainly upon the zeroth order of valve admittance. The trend of present results is consistent with the previous measurement results for various compressor and IC-engine exhaust systems. The whole result is very encouraging for further study because the simulation has been based only on a very simplified source condition.

It should be borne in mind that the acoustic wave propagation through orifices such as the intake and exhaust valves is nonlinear at the high acoustic velocities typical of the gas inflow and outflow of an engine. Such nonlinearities can cause negative resistance at some frequencies in an assumed linear source. The valve motion also causes a real engine source to be time variant, and again this invalidates the entire concept of frequency-dependent source impedance. It has been shown that if an exact, indirect method of source evaluation is applied, presuming a time-invariant source, then negative resistance values can result even for the simplest of linear sources. In this context, it can be concluded that the time-varying nature in the actual sources is probably the most important mechanism that causes the output difference between the load and direct methods, although the effects of nonlinearity have not been fully investigated.

ACKNOWLEDGMENTS

The authors would like to thank Dr. K. Peat for helpful discussions. This work was partially supported by BK21 Project and NRL.

[1] M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley-Interscience, New York, 1987).
[2] H. Bodén, ''On multi-load methods for determination of the source data of acoustic one-port sources,'' J. Sound Vib. **180**, 725–743 (1995).
[3] S.-H. Jang and J.-G. Ih, ''Refined multiload method for measuring acoustical source characteristics of an intake or exhaust system,'' J. Acoust. Soc. Am. **107**, 3217–3225 (2000).
[4] D. F. Ross and M. J. Crocker, ''Measurement of the acoustical internal impedance of an internal combustion engine,'' J. Acoust. Soc. Am. **74**, 18–27 (1983).
[5] M. G. Prasad and M. J. Crocker, ''Acoustical source characterization studies on a multi-cylinder engine exhaust system,'' J. Sound Vib. **90**, 4709–490 (1983).
[6] F. Albertson and H. Bodén, ''Method for prediction of sound generation from the IC-engine exhaust,'' Proceedings of 6th International Congress on Sound and Vibration (1999), pp. 1961–1966.
[7] L. Desmons, J. Hardy, and Y. Auregan, ''Determination of the acoustical source characteristics of an internal combustion engine by using several calibrated loads,'' J. Sound Vib. **179**, 869–878 (1995).
[8] V. H. Gupta and M. L. Munjal, ''On numerical prediction of the acoustic source characteristics of an engine exhaust system,'' J. Acoust. Soc. Am. **92**, 2716–2725 (1992).
[9] H. Bodén, ''The multiple load method for measuring the source characteristics of time variant sources,'' Report TRITA-TAK-8802, Department of Technical Acoustics, Royal Institute of Technology, Stockholm (1986).
[10] F. Albertson, ''On impedances of a simple harmonic oscillator and acoustic impedances in pipes with mean flow,'' Report TRITA-FKT 1999:21, Department of Technical Acoustics, Royal Institute of Technology, Stockholm (1999).
[11] J.-G. Ih and K. S. Peat, ''On the causes of negative impedance in the measurement of intake and exhaust noise sources,'' Appl. Acoust. **63**, 153–171 (2002).
[12] K. S. Peat and J.-G. Ih, ''An analytical investigation of the indirect measurement method of estimating the acoustic impedance of a time-varying source,'' J. Sound Vib. **244**, 821–835 (2001).
[13] M. L. Munjal and A. G. Doige, ''On uniqueness, transfer and combination of acoustic sources in one-dimensional systems,'' J. Sound Vib. **121**, 25–35 (1988).
[14] R. S. Benson, *The Thermodynamics and Gas Dynamics of Internal-Combustion Engines* (Clarendon, Oxford, 1982).
[15] F. Payri, J. M. Desantes, and A. J. Torregrosa, ''Acoustic boundary condition for unsteady one-dimensional flow calculation,'' J. Sound Vib. **188**, 85–110 (1995).
[16] M. L. Kathuriya and M. L. Munjal, ''Experimental evaluation of the aeroacoustic characteristics of a source of pulsating gas flow,'' J. Acoust. Soc. Am. **65**, 240–248 (1979).
[17] D. G. Tucker, *Circuits with Periodically-varying Parameters, Including Modulators and Parametric Amplifiers* (Van Nostrand, Princeton, 1964).
[18] J. A. Richards, *Analysis of Periodically Time-varying Systems* (Springer-Verlag, Berlin, 1983).

Energy radiated by a point acoustic dipole that reverses its uniform velocity along its rectilinear path

G. C. Gaunaurd[a)]
Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

G. C. Everstine[b)]
Naval Surface Warfare Center, Carderock Division, Bethesda, Maryland 20817-5700

(Received 26 January 2002; accepted for publication 4 November 2002)

This work extends a mathematical approach developed recently for monopoles to describe the sound energy radiated by a rectilinearly moving dipole that changes direction along its trajectory. Although the dipole travels with constant speed, it undergoes acceleration by reversing its direction during a finite time interval along its path. This work determines the joint angular and frequency distribution of the radiated energy, its angular distribution, and the total radiated energy output. Results for the radiated energy are systematized by expressing the radiation integrals in terms of hypergeometric functions. This procedure simplifies the evaluations, particularly at low Mach numbers, and permits the comparison of results to the earlier monopole case. [DOI: 10.1121/1.1532031]

PACS numbers: 43.20.Px, 43.20.Rz [MO]

I. INTRODUCTION

There are various situations involving the rectilinear motion of multipole sources that generate still unexamined radiation fields. Those situations in which the multipole undergoes acceleration only through a finite portion of its straight path are of particular interest. Some of these situations were considered in a recent paper[1] for monopole sources. This paper extends the mathematical analysis of a point acoustic dipole that reverses direction along its rectilinear path.

A traditional model for the sound field radiated by a rigid body moving slowly and uniformly through a fluid is that of a moving point dipole with the dipole axis in the direction of motion.[2,3] Today it is known that such a model is too simple[4] and requires correction. The effect of motion is far more complicated, and that effect does not involve just Doppler factors.[5–7] It has been shown that there is amplification in a direction normal to the source motion and that there is an additional omnidirectional term in the sound field. However, analyses that follow the traditional model are still very useful and form a good basis for the investigation of more complicated situations. Such situations can be decomposed into simpler cases such as the point-dipole analysis presented here. There are moving bodies that can reverse their rectilinear direction of motion in very small distances.[1] The next section attempts to provide an analytical prediction for the sound energy radiated during one such turn-around maneuver without dealing with the corrections introduced by the finite size of these bodies.

II. ENERGY RADIATION FROM A POINT DIPOLE WITH A FINITE PERIOD OF UNSTEADY RECTILINEAR MOTION

Consider a dipole source of strength α moving in a rectilinear path, as shown in Fig. 1. The motion starts at $x=-\infty$ when $t=-\infty$. It then moves forward along the x-direction with constant speed v_0. It subsequently slows down to zero speed when $t=0$ at $x=0$, and finally, it turns around and reverses its earlier motion and continues to $x=-\infty$ for $t=\infty$ with the same constant speed v_0. Since the rectilinear motion is in one dimension, this figure shows two (upper and lower) paths of motion for clarity only.

The position $x_s(t)$, speed $v_s(t)$, and acceleration $\dot{v}_s(t)$ of a point dipole which moves along the half-line $-\infty<x<0$ are given, respectively, by

$$x_s(t)=\begin{cases} v_0(t+\tau-2\tau/\pi), & -\infty<t<-\tau, \\ 2v_0(\tau/\pi)[\cos(\pi t/2\tau)-1], & -\tau<t<\tau, \\ v_0(-t+\tau-2\tau/\pi), & \tau<t<\infty, \end{cases} \tag{1}$$

$$v_s(t)=\begin{cases} v_0, & -\infty<t<-\tau, \\ -v_0\sin(\pi t/2\tau), & -\tau<t<\tau, \\ -v_0, & \tau<t<\infty, \end{cases} \tag{2}$$

$$\dot{v}_s(t)=\begin{cases} 0, & -\infty<t<-\tau, \\ -[v_0\pi/(2\tau)]\cos(\pi t/2\tau), & -\tau<t<\tau, \\ 0, & \tau<t<\infty. \end{cases} \tag{3}$$

Sound radiation will occur for $|t|<\tau$, which is the time interval in which the dipole changes direction, and there is a nonvanishing acceleration.

The governing inhomogeneous wave equation in this case is

$$\left(\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\phi(\mathbf{r},t)=-\frac{1}{\rho_0}Q(\mathbf{r},t) =-\frac{\alpha}{\rho_0}\dot{x}_s(t)\frac{\partial}{\partial x}\delta[x-x_s(t)], \tag{4}$$

where c is the sound speed, ρ_0 is the equilibrium density of the medium, and δ is the Dirac delta function (see Appendix

a)Electronic mail: GGaunaurd@arl.army.mil
b)Current affiliation: Consultant, Gaithersburg, MD.

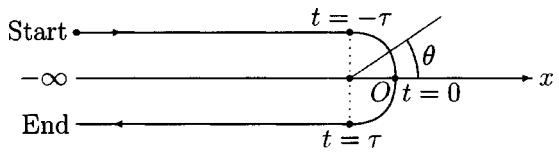


FIG. 1. An acoustic point dipole in rectilinear motion exhibiting a change of direction around $t=0$.

A for more details). The directional distribution of the radiated energy $\tilde{E}(\mathbf{n})$ is known[1] to be

$$\tilde{E}(\mathbf{n})=\int_{-\infty}^{\infty}E(\mathbf{n},\omega)d\omega, \tag{5}$$

where ω is the circular frequency of the dipole, and $E(\mathbf{n},\omega)$ is the directional distribution of the radiated energy spectrum given by[1,8–10]

$$E(\mathbf{n},\omega)=\frac{\omega^2}{2\pi\rho_0 c(4\pi)^2}\left|\int_{-\infty}^{\infty}Q(\mathbf{r},t)e^{i\omega(t-\mathbf{n}\cdot\mathbf{r}/c)}\,d\mathbf{r}\,dt\right|^2. \tag{6}$$

Here $\mathbf{n}$ is an arbitrary unit vector in some direction.

The corresponding radiation field integral in the above equation, henceforth denoted J_R, is obtained by substituting $Q(\mathbf{r},t)$ from Eq. (4):

$$J_R=\int_{-\infty}^{\infty}Q(\mathbf{r},t)e^{i\omega(t-\mathbf{n}\cdot\mathbf{r}/c)}\,d\mathbf{r}\,dt \tag{7}$$

$$=\alpha\int_{-\infty}^{\infty}\dot{x}_s(t)e^{i\omega t}\,dt\int_{-\infty}^{\infty}\frac{\partial}{\partial x}[\delta(x-x_s(t))]e^{-i\omega\mathbf{n}\cdot\mathbf{r}/c}\,d\mathbf{r} \tag{8}$$

$$=i\alpha\frac{\omega}{c}\cos\theta\int_{-\infty}^{\infty}\dot{x}_s(t)e^{i\omega[t-(\cos\theta x_s(t)/c)]}\,dt. \tag{9}$$

This integration can be split into three regimes similar to those in a recent paper,[1] viz.,

$$J_R=v_0\int_{-\infty}^{-\tau}e^{i\omega[t-M\cos\theta(t+\tau-2\tau/\pi)]}\,dt \tag{10}$$

$$-v_0\int_{-\tau}^{\tau}\sin(\pi t/2\tau)e^{i\omega[t-(2M\tau/\pi)\cos\theta(\cos(\pi t/2\tau)-1)]}\,dt \tag{11}$$

$$-v_0\int_{\tau}^{\infty}e^{i\omega[t-M\cos\theta(-t+\tau-2\tau/\pi)]}\,dt, \tag{12}$$

where the Mach number is $M=v_0/c$. The above integrals lead to expressions containing delta functions, which can then be discarded by noticing that $\omega\delta(\omega)=0$. Then, after integrating by parts, we obtain

$$J_R=-\frac{\pi v_0}{2\omega^2\tau}e^{2i(\omega\tau/\pi)M\cos\theta}\times\int_{-\pi/2}^{\pi/2}e^{(2i\omega\tau/\pi)(\sigma-M\cos\theta\cos\sigma)}\times\frac{\sin\sigma+M\cos\theta(1+2\cos^2\sigma)}{(1+M\cos\theta\sin\sigma)^4}\,d\sigma, \tag{13}$$

where the normalized time $\sigma=\pi t/(2\tau)$ has been introduced.

TABLE I. Cases plotted in the figures.

Location	Time t	Right side of Eq. (16)	σ	Figure
Before change in direction	-2τ	$9M^4\cos^4\theta$	$-\pi$	2
At start of change	$-\tau$	$M^2\cos^2\theta/(1-M\cos\theta)^7$	$-\pi/2$	3
Middle of change	0	$9M^4\cos^4\theta$	0	2
End of change	τ	$M^2\cos^2\theta/(1+M\cos\theta)^7$	$\pi/2$	4
After change in direction	2τ	$9M^4\cos^4\theta$	π	2

Substitution of this result into Eq. (6) then yields $E(\mathbf{n},\omega)$ in the form

$$E(\mathbf{n},\omega)=\frac{\alpha^2\cos^2\theta}{2\pi\rho_0 c^3(4\pi)^2}\frac{\pi^2 v_0^2}{4\tau^2}\times\left|\int_{-\pi/2}^{\pi/2}e^{(2i\omega\tau/\pi)(\sigma-M\cos\theta\cos\sigma)}\times\frac{\sin\sigma+M\cos\theta(1+2\cos^2\sigma)}{(1+M\cos\theta\sin\sigma)^4}\,d\sigma\right|^2. \tag{14}$$

Integration over all frequencies eventually yields the angular distribution of the radiated energy from the moving dipole, viz.,

$$\breve{E}(\theta)=\frac{\pi\alpha^2M^2\cos^2\theta}{128\rho_0 c\tau^3}\times\int_{-\pi/2}^{\pi/2}\frac{[\sin\sigma+M\cos\theta(1+2\cos^2\sigma)]^2}{(1+M\cos\theta\sin\sigma)^9}\,d\sigma. \tag{15}$$

The joint angular and temporal distribution of the radiated energy, which could also be called the instantaneous angular distribution of the radiated energy, is denoted $\hat{E}(\theta,\sigma)$ and is given in nondimensional form by

$$\frac{128\rho_0 c\tau^3}{\pi\alpha^2}\hat{E}(\theta,\sigma)=M^2\cos^2\theta\times\frac{[\sin\sigma+M\cos\theta(1+2\cos^2\sigma)]^2}{(1+M\cos\theta\sin\sigma)^9}. \tag{16}$$

Since there is a nonvanishing acceleration only within the interval $-\tau<t<\tau$, integration of Eq. (16) over σ from $-\pi/2$ to $\pi/2$ recovers the result indicated in Eq. (15) for the angular distribution $\breve{E}(\theta)$. Since the evaluation of this general result, as seen below, is quite cumbersome, Table I illustrates some particular relevant cases (also see Fig. 1).

These angular patterns are displayed in Figs. 2, 3, and 4 for Mach numbers from $M=0.10$ to $M=0.25$. Note the different scales used in the figures. Also note that Figs. 2, 3, and 4 are mirror reflections of each other about the plane $\theta=\pi$, which is an indication that the forward (or backward) radiation lobes reverse positions at the start and end of the acceleration-change interval. At their peaks, the lobes in Figs. 3 and 4 have amplitudes about an order of magnitude larger than either before or after the acceleration underwent changes (Fig. 2), in agreement with expectations.

An integration of Eq. (16) over θ yields $\bar{E}(\sigma)$, which has three terms:

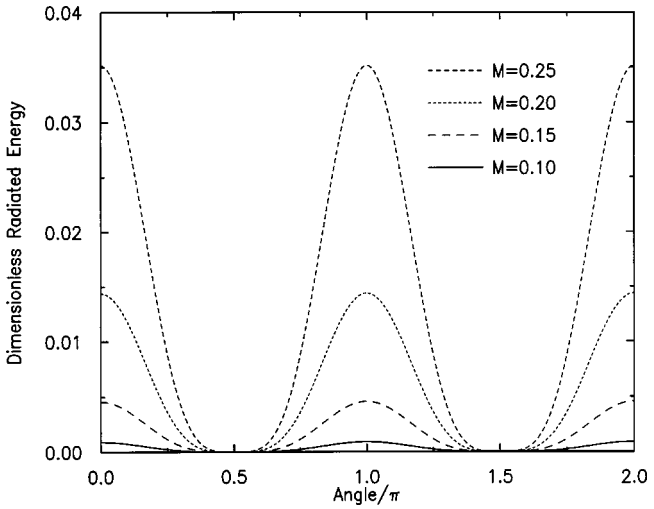


FIG. 2. Angular distribution of the instantaneously radiated energy from the moving dipole at time $t=-2\tau$, 0, and 2τ for Mach numbers from $M=0.10$ to $M=0.25$.

$$\frac{128\rho_0 c\tau^3}{\pi\alpha^2}\overline{E}(\sigma)=2\pi\int_0^{\pi}\frac{128\rho_0 c\tau^3}{\pi\alpha^2}\hat{E}(\theta,\sigma)\sin\theta\,d\theta$$

$$=2\pi M^2\sin^2\sigma\int_0^{\pi}\frac{\cos^2\theta\sin\theta\,d\theta}{(1+M\cos\theta\sin\sigma)^9}$$
$$+4\pi M^3\sin\sigma(3-2\sin^2\sigma)$$
$$\times\int_0^{\pi}\frac{\cos^3\theta\sin\theta\,d\theta}{(1+M\cos\theta\sin\sigma)^9}$$
$$+2\pi M^4(3-2\sin^2\sigma)^2$$
$$\times\int_0^{\pi}\frac{\cos^4\theta\sin\theta\,d\theta}{(1+M\cos\theta\sin\sigma)^9}. \tag{17}$$

The transformation $x=\cos\theta$ and the relabeling $N=-M\sin\sigma$ recasts Eq. (17) in the form

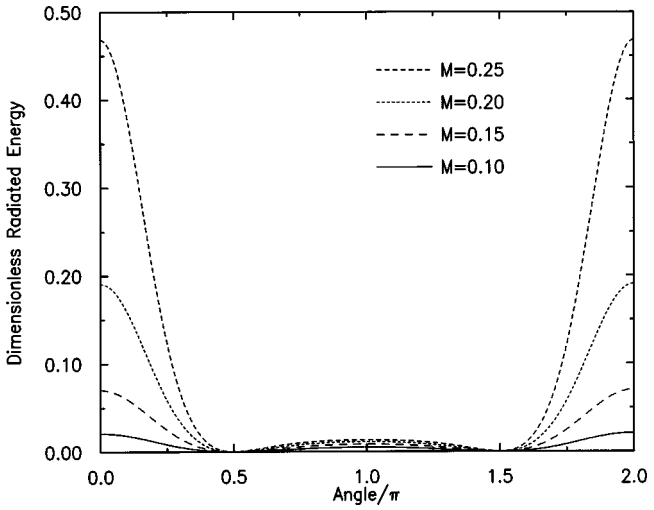


FIG. 3. Angular distribution of the instantaneously radiated energy from the moving dipole at time $t=-\tau$, when the dipole begins its turn-around maneuver, for Mach numbers from $M=0.10$ to $M=0.25$.

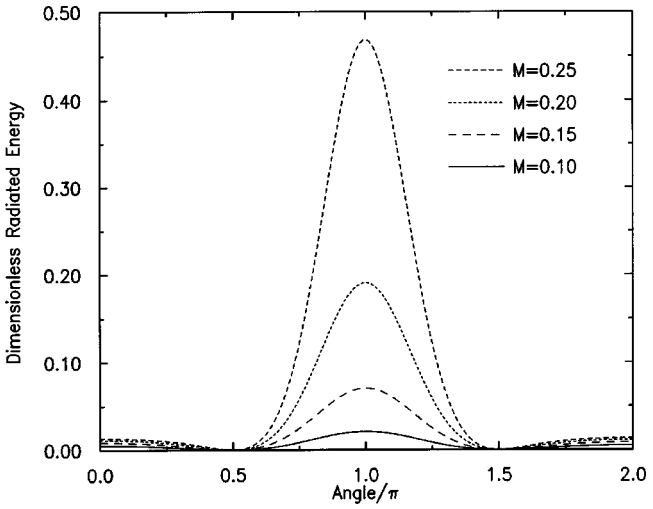


FIG. 4. Angular distribution of the instantaneously radiated energy from the moving dipole at time $t=\tau$, when the dipole ends its turn-around maneuver, for Mach numbers from $M=0.10$ to $M=0.25$.

$$\frac{128\rho_0 c\tau^3}{\pi\alpha^2}\overline{E}(\sigma)=2\pi M^2\sin^2\sigma\int_{-1}^{1}\frac{x^2\,dx}{(1-Nx)^9}$$
$$+4\pi M^3\sin\sigma(3-2\sin^2\sigma)$$
$$\times\int_{-1}^{1}\frac{x^3\,dx}{(1-Nx)^9}+2\pi M^4(3-2\sin^2\sigma)^2$$
$$\times\int_{-1}^{1}\frac{x^4\,dx}{(1-Nx)^9}. \tag{18}$$

These three integrals can be evaluated exactly, but such evaluation leads to quite cumbersome expressions. For example, it has been shown[11] that, for any polynomial $P(x)$ of degree less than n,

$$\int_{-1}^{1}\frac{P(x)dx}{(1-Nx)^n}=\sum_{k=0}^{n-2}A_k\left[\frac{1}{(1-N)^{n-k-1}}-\frac{1}{(1+N)^{n-k-1}}\right]$$
$$+A_{n-1}\ln\left(\frac{1-N}{1+N}\right), \tag{19}$$

where

$$A_k=\frac{(-1)^{k+1}}{N^{k+1}k!}\left[\frac{d^kP(x)}{dx^k}\right]_{x=1/N},\quad k=0,1,2,...,n-1. \tag{20}$$

The application of this result to the last of the integrals in Eq. (18) yields the expression

$$\int_{-1}^{1}\frac{x^4\,dx}{(1-Nx)^9}=\frac{1}{N^5}\left\{\frac{1}{8}\left[\frac{1}{(1-N)^8}-\frac{1}{(1+N)^8}\right]\right.$$
$$-\frac{4}{7}\left[\frac{1}{(1-N)^7}-\frac{1}{(1+N)^7}\right]+\left[\frac{1}{(1-N)^6}\right.$$
$$\left.-\frac{1}{(1+N)^6}\right]-\frac{4}{5}\left[\frac{1}{(1-N)^5}-\frac{1}{(1+N)^5}\right]$$
$$\left.+\frac{1}{4}\left[\frac{1}{(1-N)^4}-\frac{1}{(1+N)^4}\right]\right\}. \tag{21}$$

Similar results could be obtained for the other two integrals. Since these exact expressions are not easily amenable to expansion in powers of N, we take instead a different approach.

Since $N=-M\sin\sigma$, it is clear that $\overline{E}(\sigma)$, in general, is a very complicated function of σ. However, these integrals can also be evaluated in terms of a particular form of the Gauss hypergeometric function

$$\int_0^1 \frac{x^{s-1}\,dx}{(1+\beta x)^n}=\frac{1}{s}F(n,s;1+s;-\beta), \tag{22}$$

where $F(a,b;c;x)$ is the Gauss hypergeometric function.[12] The last of the integrals in Eq. (18) is then

$$\begin{aligned}\int_{-1}^1 \frac{x^4\,dx}{(1-Nx)^9}&=\int_0^1 \frac{x^4\,dx}{(1+Nx)^9}+\int_0^1 \frac{x^4\,dx}{(1-Nx)^9}\\&=\tfrac{1}{5}[F(9,5;6;-N)+F(9,5;6;N)]\\&\approx\tfrac{2}{5}(1+\tfrac{225}{7}N^2+\cdots).\end{aligned} \tag{23}$$

Similarly, the other two integrals reduce to

$$\int_{-1}^1 \frac{x^3\,dx}{(1-Nx)^9}\approx\frac{1}{2}(1+30N^2+\cdots), \tag{24}$$

$$\int_{-1}^1 \frac{x^2\,dx}{(1-Nx)^9}\approx\frac{2}{3}(1+27N^2+\cdots), \tag{25}$$

where the hypergeometric functions have been expanded in terms of the standard hypergeometric series.[12,13]

It then follows from Eq. (18) that

$$\begin{aligned}\frac{128\rho_0 c\tau^3}{\pi\alpha^2}\overline{E}(\sigma)&=\frac{4\pi}{3}M^2\sin^2\sigma(1+27M^2\sin^2\sigma+\cdots)\\&\quad+2\pi M^3\sin\sigma(3-2\sin^2\sigma)\\&\quad\times(1+30M^2\sin^2\sigma+\cdots)+\frac{4\pi}{5}M^4\\&\quad\times(3-2\sin^2\sigma)^2(1+\tfrac{225}{7}M^2\sin^2\sigma+\cdots)\\&\quad+\cdots.\end{aligned} \tag{26}$$

Since $\sigma=\pi t/(2\tau)$, Eq. (26) can also be viewed as describing the complicated temporal distribution or instantaneous energy radiated at time t by the dipole. To verify this interpretation, we note that, for $t=\tau$, for example, one has $\sigma=\pi/2$, in which case Eq. (26) reduces (up to order M^2) to

$$\frac{128\rho_0 c\tau^3}{\pi\alpha^2}\overline{E}\left(\frac{\pi}{2}\right)\approx\frac{4\pi}{3}M^2. \tag{27}$$

The angular distribution entry in Table I for $t=\tau$ is

$$\frac{128\rho_0 c\tau^3}{\pi\alpha^2}E(t=\tau)=\frac{M^2\cos^2\theta}{(1+M\cos\theta)^7}. \tag{28}$$

If we then integrate over all θ and retain only terms of $O(M^2)$, we find that

$$\begin{aligned}\frac{128\rho_0 c\tau^3}{\pi\alpha^2}E(t=\tau)&=2\pi M^2\int_0^\pi\frac{\cos^2\theta\sin\theta\,d\theta}{(1+M\cos\theta)^7}\\&\approx 4\pi M^2/3,\end{aligned} \tag{29}$$

which agrees with Eq. (27). For slow motions (i.e., $M^2\ll 1$), Eq. (26) reduces to its first term, in which case, for any t,

$$\overline{E}\left(\frac{\pi t}{2\tau}\right)=\frac{\pi^2\alpha^2M^2}{96\rho_0 c\tau^3}\sin^2\left(\frac{\pi t}{2\tau}\right). \tag{30}$$

Returning to the direct evaluation of Eq. (15), the evaluation of the total energy output from the motion is the most important result. Its explicit evaluation is possible based on the double integral

$$I=\int_0^\pi\sin\theta\cos^2\theta\,d\theta\int_{-\pi/2}^{\pi/2}\frac{\sin^2\sigma+2M\cos\theta\sin\sigma(3-2\sin^2\sigma)+M^2\cos^2\theta(3-2\sin^2\sigma)^2}{(1+M\cos\theta\sin\sigma)^9}\,d\sigma, \tag{31}$$

which can be split into the three integrals

$$I=I_1+I_2+I_3, \tag{32}$$

where

$$I_1=\int_0^\pi\sin\theta\,d\theta\int_{-\pi/2}^{\pi/2}\frac{\cos^2\theta\sin^2\sigma}{(1+M\cos\theta\sin\sigma)^9}\,d\sigma, \tag{33}$$

$$I_2/(2M)=\int_0^\pi\sin\theta\cos^2\theta\,d\theta\int_{-\pi/2}^{\pi/2}(3-2\sin^2\sigma)\times\frac{\cos\theta\sin\sigma}{(1+M\cos\theta\sin\sigma)^9}\,d\sigma, \tag{34}$$

$$I_3/M^2=\int_0^\pi\sin\theta\cos^4\theta\,d\theta\int_{-\pi/2}^{\pi/2}\frac{(3-2\sin^2\sigma)^2}{(1+M\cos\theta\sin\sigma)^9}\,d\sigma. \tag{35}$$

The last three integrals (I_1, I_2, I_3) can be reduced to the sum of simpler integrals by successive integrations by parts, the details of which are given in Appendix B. The result of this simplification is

$$I = I_1 + I_2 + I_3 = \frac{2\pi}{3}\left(1 + \frac{621}{40}M^2 + \cdots\right), \tag{36}$$

and the total energy radiated by the dipole is

$$\begin{aligned} E_{\text{dip}} &= 2\pi \int_0^{\pi} \breve{E}(\theta)\sin\theta\, d\theta \\ &= \frac{\pi^2 M^2 \alpha^2}{128\rho_0 c\tau^3} I \\ &= \frac{\pi^3 M^2 \alpha^2}{192\rho_0 c\tau^3}\left(1 + \frac{621}{40}M^2 + \cdots\right), \quad M \ll 1. \end{aligned} \tag{37}$$

It would be possible to obtain more terms by keeping them in the expression for the hypergeometric series. For small Mach numbers, the above result suffices. For a very slow motion limit (i.e., $M \to 0$), the above result can be compared to that of a monopole source of strength q undergoing the same motion.[1] This result was

$$E_{\text{mon}} \to \frac{q^2 M^2 \pi}{48\rho_0 c\tau}. \tag{38}$$

Therefore, the ratio of the radiated energies in these two cases is

$$\lim_{M\to 0} \frac{E_{\text{dip}}}{E_{\text{mon}}} = \left(\frac{\pi\alpha}{2q\tau}\right)^2. \tag{39}$$

Thus, the ratio of radiated energies is directly proportional to the square of the ratio of their strengths and inversely proportional to the square of the finite time-span during which both multipoles are changing direction.

Additional comparisons of methods used here are possible now. For a slow motion (i.e., $M \ll 1$), the right-hand side of Eq. (26) reduces to its first term. Integration over σ then yields

$$\frac{128\rho_0 c\tau^3}{\pi\alpha^2} E_{\text{dip}} \underset{M\ll 1}{\to} \frac{4\pi}{3} M^2 \int_{-\pi/2}^{\pi/2} \sin^2\sigma\, d\sigma = 2\pi^2 M^2/3 \tag{40}$$

or

$$E_{\text{dip}} \to \frac{\pi^3 M^2 \alpha^2}{192\rho_0 c\tau^3}, \tag{41}$$

which agrees with the approach that led to Eq. (37) for small M. It was found earlier that, for $t = 0$, Eq. (16) produced the angular pattern

$$\frac{128\rho_0 c\tau^3}{\pi\alpha^2} \hat{E}(\theta, \sigma = 0) = 9M^4 \cos^2\theta, \tag{42}$$

and when this expression is integrated over all θ, the resulting energy is

$$E_{\text{dip}} = 2\pi \int_0^{\pi} 9M^4 \cos^4\theta \sin\theta\, d\theta = 36\pi M^4/5. \tag{43}$$

In this same case, the method that led to Eq. (26) yields

$$E_{\text{dip}} \underset{\sigma\to 0}{\to} 2\pi\left(\tfrac{2}{5}\right)M^4 \cdot 9 = 36\pi M^4/5, \tag{44}$$

which is also in agreement with the previous method above.

III. CONCLUDING REMARKS

The angular distribution of the energy spectrum, the angular distribution of the radiated acoustic energy, and the total radiated energy output of the dipole undergoing the rectilinear motion described in Eq. (1) have been obtained. The evaluation of the integrals is performed in terms of hypergeometric functions. The results are approximated for low Mach number values. A comparison with an analogous result for a monopole undergoing the same ''partially accelerated'' motion is obtained. Some angular distribution patterns of the radiated energy are plotted for some particular cases.

ACKNOWLEDGMENT

The authors gratefully acknowledge H. Levine for his comments.

APPENDIX A

If $\mathbf{n}$ denotes an arbitrarily oriented unit vector and $d\Omega_n$ the element of solid angle about $\mathbf{n}$, then the relations[1,8–11]

$$P(t) = \int Q(\mathbf{r},t)\frac{\partial\phi}{\partial t}\, d\mathbf{r} = \int P(\mathbf{n},t)\, d\Omega_n \tag{A1}$$

and

$$\begin{aligned} P(\mathbf{n},t) &= -\frac{1}{16\rho_0 c\pi^2}\int_{-\infty}^{\infty} Q(\mathbf{r},t)Q(\mathbf{r}',t') \\ &\quad \times \delta''\left(t' - t + \frac{\mathbf{n}}{c}\cdot(\mathbf{r}-\mathbf{r}')\right) d\mathbf{r}\, d\mathbf{r}'\, dt' \end{aligned} \tag{A2}$$

specify the overall instantaneous power $P(t)$ radiated from the source into its surroundings and the amount $P(\mathbf{n},t)$ radiated along the given $\mathbf{n}$ direction. It then follows that the expressions

$$\begin{aligned} E(\mathbf{n}) &= \int_{-\infty}^{\infty} P(\mathbf{n},t)\, dt \\ &= -\frac{1}{16\rho_0 c\pi^2}\int_{-\infty}^{\infty} Q(\mathbf{r},t)Q(\mathbf{r}',t') \\ &\quad \times \delta''\left(t' - t + \frac{\mathbf{n}}{c}\cdot(\mathbf{r}-\mathbf{r}')\right) d\mathbf{r}\, d\mathbf{r}'\, dt\, dt' \\ &= \int_{-\infty}^{\infty} E(\mathbf{n},\omega)\, d\omega = \breve{E}(\theta) = \int_{-\infty}^{\infty} \hat{E}(\theta,\sigma)\, d\sigma \end{aligned} \tag{A3}$$

and

$$E(\mathbf{n},\omega) = \frac{\omega^2}{2\pi\rho_0 c(4\pi)^2}\left|\int_{-\infty}^{\infty} Q(\mathbf{r},t)e^{i\omega(t-\mathbf{n}\cdot\mathbf{r}/c)}\, d\mathbf{r}\, dt\right|^2 \tag{A4}$$

describe the angular distribution and frequency distribution of the radiated energy. The particular source function

$$Q(\mathbf{r},t)=\alpha\dot{x}_s(t)\frac{\partial}{\partial x}\delta[x-x_s(t)] \tag{A5}$$

pertains to a point dipole in motion with colinear axis, as is used in Eq. (4).

APPENDIX B

Here we show the details for the simplification of the integrals in Eqs. (33), (34), and (35). Each of these integrals can be reduced to the sum of simpler integrals by successive integrations by parts. Intermediate results are

$$I_1=\frac{4}{21}\frac{d^2}{dM^2}\left[\int_{-\pi/2}^{\pi/2}\left(\frac{1}{D^6}-\frac{1}{D^5}+\frac{3}{16D^4}\right)d\sigma\right], \tag{B1}$$

where

$$D=1-M^2\sin^2\sigma, \tag{B2}$$

and

$$I_2/(2M)=-\frac{1}{210}\frac{d^2}{dM^2}\left[\left(3M-\frac{2}{M}\right)\int_0^{\pi/2}\left(\frac{80}{D^6}-\frac{48}{D^5}+\frac{3}{D^4}\right)\times d\sigma+\frac{2}{M}\int_0^{\pi/2}\left(\frac{80}{D^5}-\frac{48}{D^4}+\frac{3}{D^3}\right)d\sigma\right], \tag{B3}$$

$$I_3/M^2=\frac{4}{35}\frac{d}{dM}\left\{M\left[63\int_0^{\pi/2}\frac{d\sigma}{D^7}+(234M^2-84)\times\int_0^{\pi/2}\frac{\sin^2\sigma\,d\sigma}{D^7}+(63M^4-312M^2+28)\times\int_0^{\pi/2}\frac{\sin^4\sigma\,d\sigma}{D^7}+(104M^2-84M^4)\times\int_0^{\pi/2}\frac{\sin^6\sigma\,d\sigma}{D^7}+28M^4\int_0^{\pi/2}\frac{\sin^8\sigma\,d\sigma}{D^7}\right]\right\}. \tag{B4}$$

The above integrals all admit representations in terms of hypergeometric functions[13] of argument M^2. The pertinent relation is[14]

$$H_{a,n}(M^2)=\int_0^{\pi/2}\frac{\sin^{2n}\sigma\,d\sigma}{(1-M^2\sin^2\sigma)^a}=\frac{\Gamma(n+\frac{1}{2})\sqrt{\pi}}{2\Gamma(n+1)}F(a,n+1/2;n+1;M^2), \tag{B5}$$

where F is the hypergeometric function, and Γ is the gamma function. The above integrals can thus be reduced to the form

$$I_1=\frac{8}{21}\frac{d^2}{dM^2}\left(H_{6,0}-H_{5,0}+\frac{3}{16}H_{4,0}\right), \tag{B6}$$

$$I_2/M=-\frac{1}{210}\frac{d^2}{dM^2}\left[\left(3M-\frac{2}{M}\right)(80H_{6,0}-48H_{5,0}+3H_{4,0})+\frac{2}{M}(80H_{5,0}-48H_{4,0}+3H_{3,0})\right], \tag{B7}$$

$$I_3/M^2=\frac{4}{35}\frac{d}{dM}\{M[63H_{7,0}+(234M^2-84)H_{7,1}+(63M^4-312M^2+28)H_{7,2}+(104M^2-84M^4)H_{7,3}+28M^4H_{7,4}]\}. \tag{B8}$$

This is a convenient form in which to express the solution, since, using the hypergeometric series[13]

$$F(a,n+1/2;n+1;M^2)=1+\frac{a(n+\frac{1}{2})}{1!(n+1)}M^2+\frac{a(a+1)(n+\frac{3}{2})(n+\frac{1}{2})}{2!(n+1)(n+2)}M^4+\cdots, \tag{B9}$$

all the integrals can be expressed in terms of even powers of the Mach number M. Therefore,

$$I_1=\frac{\pi}{3}\left(1+\frac{81}{4}M^2+\cdots\right), \tag{B10}$$

$$I_2=\frac{\pi}{3}\left(1+\frac{27}{5}M^2+\cdots\right), \tag{B11}$$

$$I_3=\frac{\pi}{3}\left(\frac{27}{5}M^2+\frac{675}{14}M^4+\cdots\right). \tag{B12}$$

[1] H. Levine and G. C. Gaunaurd, "Energy radiation from point (monopole) sources whose duration of accelerated motion is finite," J. Acoust. Soc. Am. **110**, 31–36 (2001).

[2] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[3] J. E. Ffowcs Williams and D. L. Hawkins, "Sound generation by turbulence and surfaces in arbitrary motion," Philos. Trans. R. Soc. London, Ser. A **264**, 321–342 (1969).

[4] A. P. Dowling, "Convective amplification of real simple sources," J. Fluid Mech. **74**, 529–546 (1976).

[5] S. E. Wright, "Sources and observers in motion, Part I: Time variant analysis and its implications for aerodynamic sound," J. Sound Vib. **108**, 361–378 (1986).

[6] S. E. Wright and D. J. Lee, "Sources and observers in motion, Part II: Acoustic radiation from a small rigid body in arbitrary motion," J. Sound Vib. **108**, 379–387 (1986).

[7] S. E. Wright and D. J. Lee, "Sources and observers in motion, Part III: Acoustic radiation from noncompact rigid bodies moving at high speed," J. Sound Vib. **108**, 389–403 (1986).

[8] H. Levine, "The output of acoustic sources," J. Acoust. Soc. Am. **67**, 1935–1946 (1980).

[9] G. C. Gaunaurd and T. J. Eisler, "Classical electrodynamics and acoustics: Sound radiation by moving multipoles I," J. Vibr. Acoust. **119**, 271–282 (1997).

[10] G. C. Gaunaurd and T. J. Eisler, "Classical electrodynamics and acoustics: Sound radiation by moving multipoles II," J. Vibr. Acoust. **121**, 126–130 (1999).

[11] G. C. Gaunaurd, "Contributions to the Theory of Acoustic Radiation by Moving Multipole Sources," Ph.D. thesis, Catholic University of America, Washington, DC, 1971. (Also available with more detail as Naval Ordnance Laboratory Report NOLTR-72-75, 1972, and Technical Note NOLTN-9578, May 1972).

[12] J. Spanier and K. B. Oldham, *An Atlas of Functions* (Hemisphere Publishing, New York, 1987).

[13] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis: An Introduction to the General Theory of Infinite Processes and of Analytic Functions; With an Account of the Principal*, 4th ed. (Cambridge University Press, Cambridge, 1997).

[14] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions*, Vol. I (McGraw-Hill, New York, 1953).

Nonlinear scattering of acoustic waves by natural and artificially generated subsurface bubble layers in sea

Lev A. Ostrovsky
Zel Technologies/NOAA Environmental Technology Laboratory, Boulder, Colorado 80303 and Institute of Applied Physics, Russian Academy of Sciences, Uljanova, 46, Nizhny Novgorod, Russia

Alexander M. Sutin, Irina A. Soustova, Alexander L. Matveyev, and Andrey I. Potapov
Institute of Applied Physics, Russian Academy of Sciences, Uljanova, 46, Nizhny Novgorod, Russia

Zigmund Kluzek
Institute of Oceanology, Polish Academy of Sciences, Powstancow Warszawy 55, 81-712 Sopot, Poland

(Received 18 September 2001; accepted for publication 4 October 2002)

The paper describes nonlinear effects due to a biharmonic acoustic signal scattering from air bubbles in the sea. The results of field experiments in a shallow sea are presented. Two waves radiated at frequencies 30 and 31–37 kHz generated backscattered signals at sum and difference frequencies in a bubble layer. A motorboat propeller was used to generate bubbles with different concentrations at different times, up to the return to the natural subsurface layer. Theoretical consideration is given for these effects. The experimental data are in a reasonably good agreement with theoretical predictions.
 [DOI: 10.1121/1.1526497]

PACS numbers: 43.25.Yw, 43.30.Gv, 43.30.Qd [MFH]

I. INTRODUCTION

Breaking waves are known to produce air bubbles in the oceanic subsurface layer. Bubble layers and clouds formed as a result may strongly affect the propagation and damping of sound in the ocean, the spectra of ambient noise, and the atmosphere-ocean gas exchange. Hence the numerous studies of bubble distribution and its dependence on wind speed (see, for example, Refs. 1–8).

The sea bubble population measurements began in 1970s with the use of optical technique[9,10] and different acoustic methods. In particular, Medwin[11] developed efficient devices based on resonance sound attenuation and conducted measurements under various sea conditions.[1,12,13] An updated version of attenuation technique is described by Farmer and Vagle.[14] Also, remote-sensing acoustic methods were developed based on the linear scattering from bubble layers.[15–20] It was found that bubbles are concentrated in a 5 to 7 m deep upper layer being widely distributed in radii, R, so that their concentration $n(R)$ is proportional to R^{-b}, where b lies in the range of 3 to 4.

However, under some realistic conditions these methods, which are based on linear theory, may be insufficient or ineffective. Such conditions include, for example, the presence of other scattering particles such as plankton, or the closeness of boundaries (surface, bottom), when it is difficult to filter out the signals scattered by bubbles from the signals caused by reflection from other objects. To overcome these difficulties, the use of nonlinear methods of bubble diagnostics can be extremely helpful. Indeed, bubble-related nonlinearity exceeds by several orders that of almost any other object in the ocean with a comparable linear scattering cross section, so that bubbles can be easily selected from the background by their nonlinearity.

The fact that bubbles have a strong acoustical nonlinearity that can be used for bubble diagnostics is well known since at least 1980s. Bubble nonlinearity is manifested in the appearance of higher harmonics and combination frequencies in the spectrum of propagating and scattered signals. Nonlinear methods based on the second harmonic generation have been used to detect bubbles in the ocean,[21,22] in blood,[23] and in pipelines.[24] Similar techniques are widely employed in medical ultrasound imaging, with artificial bubbles used as contrast agents (e.g., Refs. 25–27). The generation of sum- and difference-frequency signals scattered from bubbles was studied theoretically and in laboratory experiments,[28–31] and a similar technique was used for bubble registration in sea.[32–35] Another nonlinear acoustic effect, the subharmonic generation, was also applied for bubble measurements in sea.[36] These nonlinear methods are sensitive enough to register even a single bubble and to distinguish bubbles from other inhomogeneities or reflecting boundaries. Note that bubble population can be measured in a simpler way in case of the incoherent backward nonlinear scattering (nonlinear reverberation). It was demonstrated in a laboratory tank.[31]

Also, the difference-frequency signal generation by two primary (pump) waves with close frequencies is used in parametric acoustic arrays, which provide high radiation directivity at low frequencies for rather small sizes of a pump radiator.[37] The main disadvantage of parametric arrays is their low efficiency. It was suggested that parametric radiation can be enhanced by using bubbly media with high nonlinearity.[38–40] This possibility was experimentally verified in laboratory tanks,[40–43] including the use of resonance in a bubble layer.[44,45]

As regards the experiments with bubbles in real seas, there were relatively few experiments performed (see references above), and in practically all cases nonlinear effects

were measured only locally, at short distances from the transducer (up to about 1 m), and mainly for single bubbles or small bubble groups, whereas it is important to extend such techniques to the remote measurement of bubble layers. Parametric arrays were used to observe linear scattering from bubbles,[18,20] but nonlinear properties of bubbles did not play any significant role in these experiments.

This paper describes some results of our experiments on coherent and incoherent nonlinear scattering from natural and artificial bubble layers performed in Baltic Sea, and the corresponding theoretical treatment of the problem. To the best of our knowledge, it was a first realization of parametric radiation formed by a bubble layer in a real sea. The use of artificial bubbles (generated by a propeller) enabled us to perform measurements at different bubble concentrations. The effects in question could be observed at signal levels characteristic of conventional sonars.

II. EXPERIMENTAL SETUP

The experiments on nonlinear interaction of acoustic waves in a subsurface bubble layer were performed in a shallow area (of about 30 m depth) of the southern part of Baltic Sea, from the board of the Polish research vessel ''Oceania.'' Two echo-sounders were used as transmitters, one at a fixed frequency of 30 kHz, whereas the frequency of the second varied from 31 to 37 kHz. Each transmitter radiated acoustical pulses with axial amplitude of 26 kPa as converted to a distance of 1 m. Pulse duration was 5.5 ms, and time interval between pulses, 200 ms. Backscattered signals were received by a broad-band, piston-type hydrophone. The diameter of each of the two transmitters was 30 cm, and of the hydrophone, 15 cm. The 0.5 level beam width was about 9° for each transmitter and about 17° for the hydrophone. The latter was calibrated by comparison with a standard Bruel&Kjœr 8101 hydrophone. Signals received by the hydrophone were amplified by the Bruel & Kjær Conditioning Amplifier and then sent to a personal computer through the 14-bit analog–digital transformer DATEL SRC-414b2 with a sampling frequency of 250 kHz.

The transmitters were mounted on a steel frame, with centers 40 cm apart and aligned to radiate parallel beams pointed upward to the surface. The hydrophone was placed on the same frame, in the middle between the transmitters, forming a triangle with them. The unit was suspended at depth $H=16$ m (Fig. 1). Most typical wind speed during these experiments was 6 m/s, and sea state 2–3.

To obtain more controllable conditions and observe the effect for different bubble concentrations we used a motorboat propeller to create an artificial bubble cloud. A small motorboat was attached to the vessel's board so that the cavitating jet from the propeller crossed the radiation of the acoustic system placed at 16 m depth (as mentioned) and at a horizontal distance of 7 m from the propeller. The reflected signals were registered beginning 2–3 minutes after the motor was stopped (in order to let large bubbles to be eliminated due to buoyancy). Then the nonlinearly scattered signals were measured at different times. After about 30 minutes the scattered signals ceased to systematically change, so that the measurements at 45 min were considered as those referred to the background situation when the bubble concentration returns to its natural level.

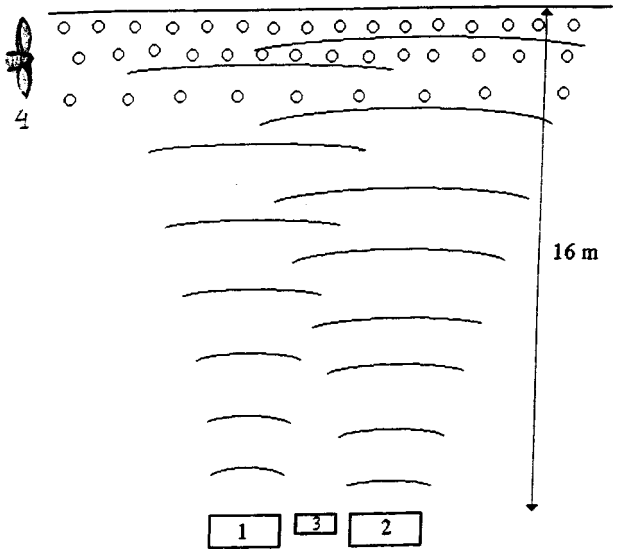


FIG. 1. Schematic of the experiment geometry: (1) transducer at 30 kHz; (2) transducer at 31–37 kHz; (3) hydrophone; (4) motorboat propeller.

III. EXPERIMENTAL RESULTS

A. Nonlinear scattering from a natural bubble layer

The pressure signals backscattered from the bubble layer after 45 min are shown in Fig. 2. As mentioned, at this stage the effect of artificial bubbles could be neglected. Figure 2(a) presents a power spectrum of the total reflected signal averaged over five pulses in the frequency band from 0 to 80 kHz, a bandwidth of each frequency channel being 167 Hz. Along with primary frequencies 29.7 and 35.2 kHz, it shows second harmonics of radiated signals (59.4 and 70.4 kHz), together with the difference frequency (5.5 kHz) and sum-frequency (64.9 kHz) components. To measure these spectral components separately, the received signal was filtered in the 2 kHz bands with the filter bandwidth centers at the selected frequencies. Then, to provide the time dependence for the corresponding amplitudes, Hilbert transform was applied to the received signal, and the modulus of the resulting complex function was used for further processing. The result was averaged over five pulses. It was sufficient for obtaining a large signal/noise ratio. Note that the use of a series longer than few minutes was undesirable because the bubble concentration varies at such time scales.

The envelopes of signals at carrier frequencies are presented in Fig. 2(b). These signals are due to linear propagation through the bubble layer and the subsequent reflection from the water surface. The nonlinearly scattered signals, at difference and sum frequencies respectively, are shown in Fig. 2(c). It is seen that the level of the difference-frequency signal reaches 2.5 Pa. It can be shown that such a pressure level cannot be provided by an incoherent scattering (a summation of the intensities of waves scattered from each bubble). Indeed, the latter can be evaluated using formulas similar to those discussed below in Sec. III C for the sum-

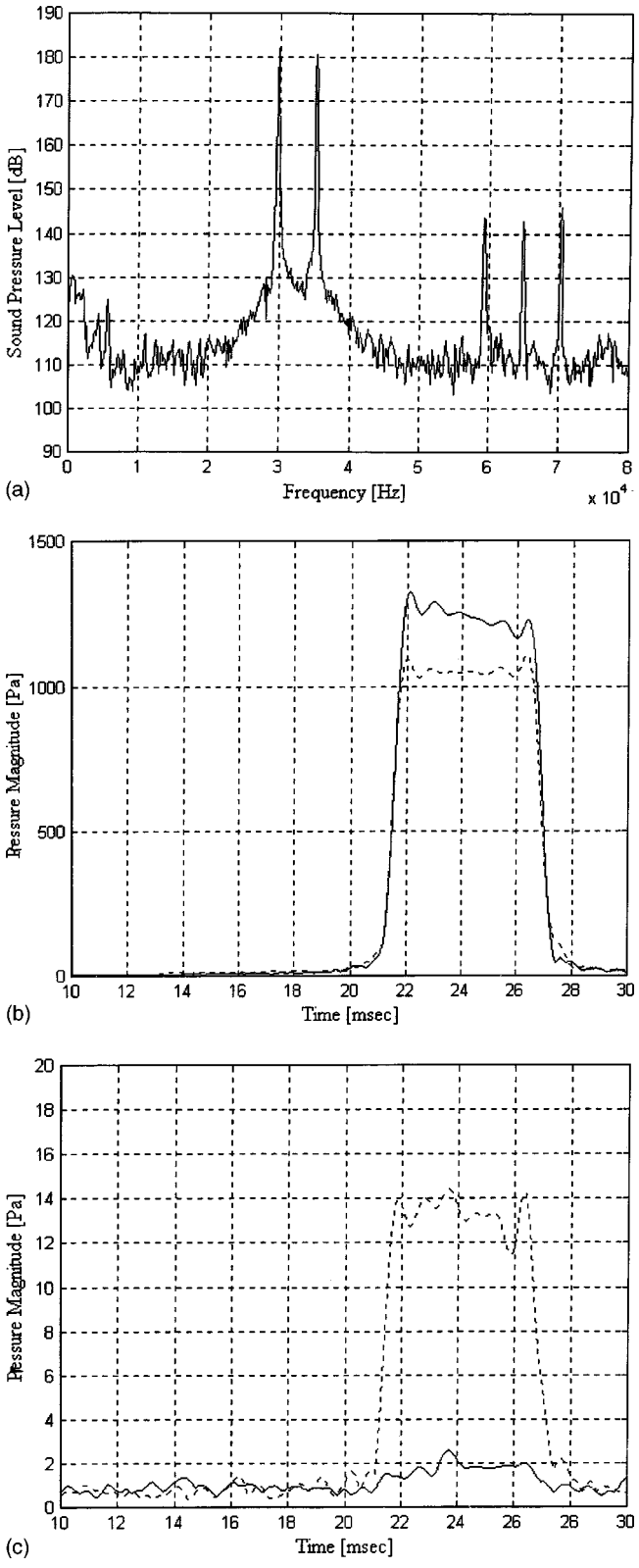


FIG. 2. Signals from the natural bubble layer (45 min after the motorboat propeller had stopped): (a) total signal spectrum; (b) envelopes of signals received at 35.2 kHz (solid line) and 29.7 kHz (dashed line); (c) same for difference-frequency signal (5.5 kHz, solid line) and sum-frequency signal (64.9 kHz, dashed line).

frequency signal, where the factor ω_+^3 in Eq. (15) must be replaced by a much smaller quantity ω_-^3. We interpreted it as a coherent (forward-scattered) low-frequency signal generated in the bubble layer and then reflected from the surface.

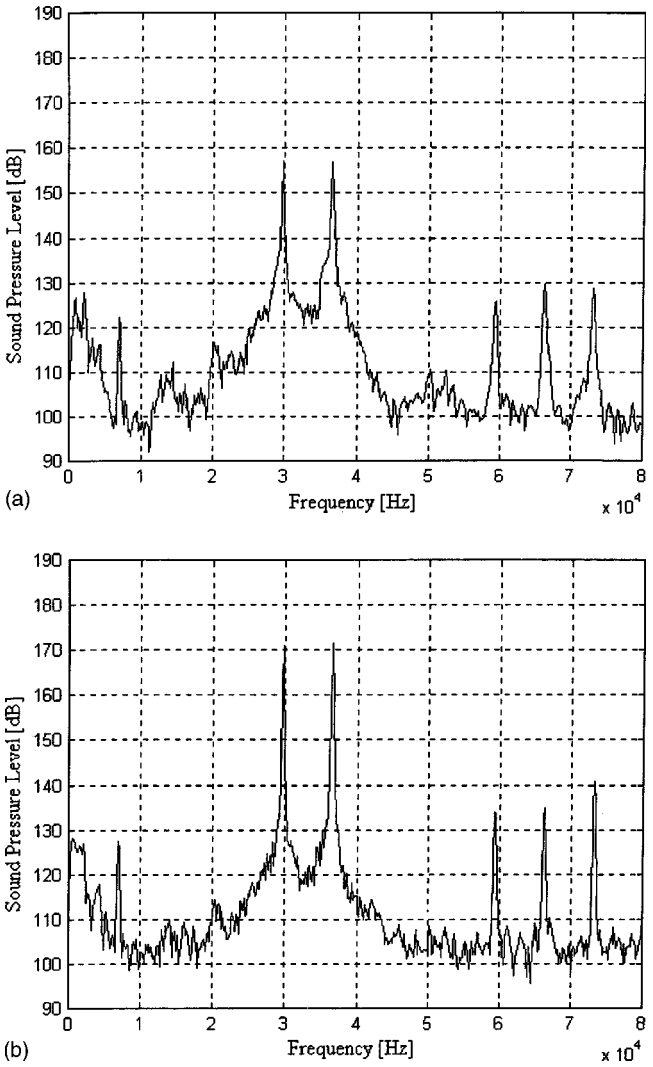


FIG. 3. Total spectrum of signals from the artificial bubble layer after 5 min (a) and 20 min (b) after the motor had stopped.

The recalculation of this signal to the standard 1 m distance from the radiator gives the sound pressure level, SPL=152 dB/μPa m. Note that this is a rather high level as compared to a ''classical'' parametric array having a pump power of only about 0.7 kW for each channel. It is also important that a parametric array using a subsurface bubble layer would not need a long distance for secondary beam formation. Indeed, here the length of an effective source of the difference-frequency wave is limited by the thickness of the bubble layer, whereas the directivity is achieved by the aperture of a radiating bubble area (cf. Ref. 40).

At the same time, the sum-frequency field does not penetrate deeply into the bubble layer due to the attenuation and is formed mainly in the lower part of the layer.

B. Nonlinear scattering from a motorboat wake

At smaller times, the signal was formed by a bubbly wake generated by the propeller. Figures 3(a) and (b) show the full signal spectra for 5 min and 20 min after the motor was stopped, respectively. The time window of spectral analysis was comparable with the pulse duration. As a result, a noticeable part of the energy of the sum-frequency signal

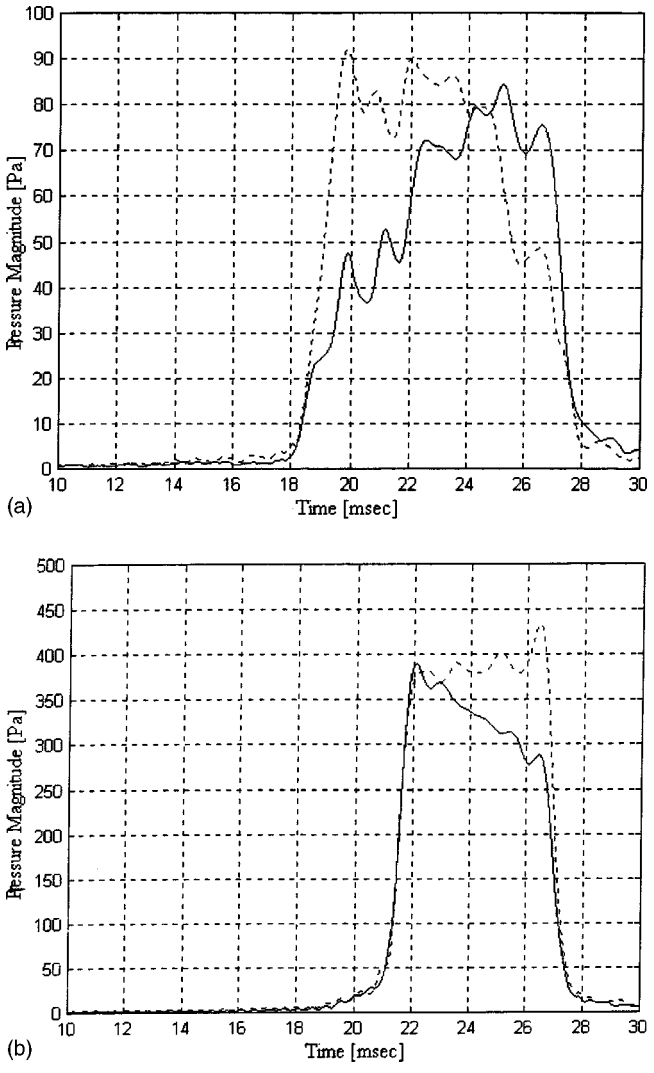


FIG. 4. Envelopes of primary signals received at 36.5 kHz (solid line) and 29.7 kHz (dashed line) from the artificial bubble layer 5 min (a) and 20 min (b) after the motor had stopped.

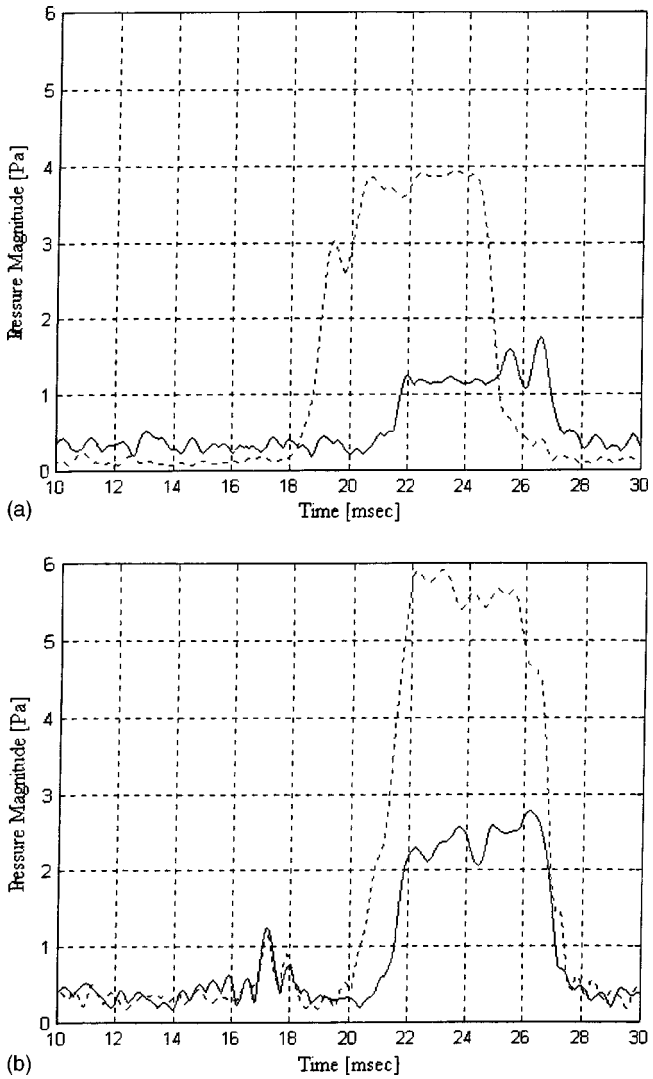


FIG. 5. Envelopes of signals received at difference frequency (6.8 kHz, solid line) and sum frequency (66.2 kHz, dashed line) from the artificial bubble layer 5 min (a) and 20 min (b) after the motorboat had stopped.

(shifted in time with respect to the difference-frequency one) could remain outside the window, thus decreasing the spectral level. The better quantitative results follow from Fig. 5 below showing the pulse envelopes. Figures 4(a) and (b) present the envelopes of both primary waves (29.7 and 36.5 kHz). Signals at the difference (6.8 kHz) and sum (66.2 kHz) frequencies are presented in Figs. 5(a) and (b).

As seen from the curves plotted in Fig. 5(a), the sum-frequency pulse arrives about 2.6 ms earlier than the difference-frequency signal. This implies that the sum-frequency signal arises due to nonlinear reverberation in the subsurface bubble layer, whereas the difference-frequency signal is generated coherently in the course of propagation of primary waves and reflected from the wave surface. Note that in the total echo signal [see Fig. 4(a)] it is practically impossible to distinguish between the parts corresponding to reverberation and reflection. The thickness of the bubble layer, l, could be evaluated from the delay between sum-frequency signal that starts to be reflected from its lower edge, and a difference-frequency signal reflected from the sea surface. For the time of 5 min after the motor was stopped, the value of l was about 1.9 m.

Comparison with the case of natural bubbles (Fig. 2) shows that dependence of the received nonlinear signal levels on the bubble concentration is not straightforward: their level in Fig. 2(a) exceeds that in Figs. 3(a,b) in spite of the decrease of bubble concentration with time. The same is seen from comparison of Figs. 5(a) and 5(b). The presumable cause of that is the decrease of losses and the corresponding growth of the primary signals, evident from Figs. 2–4.

IV. THEORY

A. Average field

Theoretical considerations of nonlinear sound propagation in liquid with bubbles have been performed more than once (see Ref. 46 and references therein). For our purposes, these results must be significantly modified to be able to take into account the signals reflected from the surface and access nonlinear reverberation at the sum frequency. In what follows we consider the corresponding models and compare them with the experimental data.

Because the bubbles can move and they are randomly distributed in space, the scattered sound contains, in general, a coherent part (average field) and an incoherent one, i.e., a sum of powers of waves scattered with random phases. The relationship between these two parts depends on a number of factors, such as average distance between bubbles, their motion in subsurface turbulent flows and due to buoyancy, wave scattering from the wavy surface, and the way of the received signal processing. In our case, at one impulse duration (5.5 ms) the bubbles can be considered ''frozen'' in space. However, the bubble positions may vary from pulse to pulse, so that, as mentioned, a group of five pulses (with total duration of about 1 s) was processed in each measurement.

If the amount of bubbles per wavelength is large, the liquid–gas mixture can be considered as a continuous medium with effective parameters depending on the bubble content and wave frequency (for small bubbles, only their monopole pulsations should be taken into account). Thus, the average acoustic pressure amplitude P_s at a frequency ω_s can be found from the wave equation

$$\Delta P_s + (K_s^2 + 2iK_s\alpha_s)P_s = q_s, \tag{1}$$

where $K_s = \omega_s/c$ is the wave number at frequency ω_s, c is the sound speed, α_s is the damping rate in linear approximation, and q_s is nonlinear (virtual) source proportional to the products or squares of the primary wave amplitudes. The nonlinearly generated wave amplitudes are always small compared to those of the primary waves (with frequencies ω_1 and ω_2), so that the primary waves can be considered in linear approximation, $q_{1,2}=0$.

The bubble size distribution in real seas, characterized by a function $n(R)$, is usually broad, and if the Q factor for a single bubble is high enough (for small bubbles, $Q \approx 10$), the medium response is determined primarily by resonant bubbles. In this case, according to the well-known formula relating the attenuation coefficient α to the bubble concentration for linear waves (Ref. 1, Sec. 6.4.2):

$$\alpha = \pi n(R_\omega)\lambda R_\omega^2/2. \tag{2}$$

Here λ is wavelength and R_ω is the resonant-bubble radius related to the sound frequency as (with the surface tension neglected)

$$R_\omega = \frac{1}{\omega}\sqrt{\frac{3\gamma p_0}{\rho_0}},$$

where γ is the polythropic ratio of the gas in a bubble, p_0 is static pressure, and ρ_0 is the medium density. In particular, the bubbles resonant at primary frequencies have a radius close to 100 microns.

Comparing the scattered impulse amplitudes at primary frequencies after 5 min and 45 min after the boat motor was stopped and supposing that after 45 min the bubble concentration has returned to its natural value, the evaluated value of α was 0.72 m^{-1}, and the total wave attenuation in the layer due to artificial bubbles, 12.5 dB. From the formula (2) this gives the resonant bubble concentration $n(R_\omega)$ generated by the propeller of about 700 $\mu\mathrm{m}^{-1}\,\mathrm{m}^{-3}$ at $\omega = 30$ kHz. Supposing that n is proportional to $R^{-3.5}$ (i.e., between the powers 3 and 4 mentioned above), we obtain that $n = 7\times10^{-6}$ $R^{-3.5}\,\mathrm{m}^{-4}$. This is confirmed by the direct measurement. Indeed, as seen from Fig. 4(a), the level of the reflected primary signal was about 80 Pa. The calculation for the lossless case gives a value of about 1600 Pa. From here one can again estimate the bubble concentration at the pump frequency as 700 $\mu\mathrm{m}^{-1}\,\mathrm{m}^{-3}$.

Note that, due to the smallness of bubble volume ratio and especially due to their broad distribution in radii, the bubble effect on the sound speed is small. This follows from the general formula for sound speed $c(\omega)$ (Ref. 46, p. 229):

$$c(\omega) = c_0\left[1 + \frac{4\pi c_0^2}{\omega^2}\int_{R_{\min}}^{R_{\max}} \frac{\xi n(R) R\, dR}{\xi^2 + Q^{-2}}\right]^{-1/2}, \tag{3}$$

where c_0 is the sound speed in water without bubbles, and $\xi = (R_\omega/R)^2 - 1$. Integrating in any reasonable limits, even from $R_{\min} = 1$ μm to $R_{\max} = 1$ mm gives, with the above $n(R)$ and $Q = 10$, the change the sound speed due to bubbles to be of order 10^{-3} (void fraction in these limits is of order 10^{-6}). This is confirmed by the data of sound-speed variations in shallow seas which are presumably due to the bubbles (Ref. 1, Fig. 6.3.3) showing the values from 10^{-3} to 10^{-2} for comparable winds. Hence, in what follows we consider the medium nondispersive ($c = c_0$), whereas the bubble effect on attenuation and nonlinearity is crucial.

Now consider the nonlinear effects of frequency transformation. The nonlinear sources q_s in (1) created by the linear primary waves are evidently quadratic in their amplitudes. If the primary wave frequencies are $\omega_{1,2}$, the respective values of q_s for difference-frequency ($\omega_- = \omega_1 - \omega_2$) and sum-frequency ($\omega_+ = \omega_1 + \omega_2$) waves can be written as

$$q_- = \frac{\epsilon_-\omega_-^2}{\rho_0 c_0^4} P_1 P_2^*, \tag{4}$$

$$q_+ = \frac{\epsilon_+\omega_+^2}{\rho_0 c_0^4} P_1 P_2^*. \tag{5}$$

Here ϵ_+ and ϵ_- are the nonlinearity parameters that differ for different types of interaction. The specifics of bubbles is in the dependence of these parameters on frequencies of interacting waves and on the distribution of bubbles over their sizes.

The nonlinearity parameter for the difference-frequency wave generation can be represented in the form (Ref. 46, p. 233):

$$\epsilon_- = i\pi^2 c_0^4\left[\frac{(3\gamma+2)n(R_\omega)}{\omega^3(\omega_- - i\omega/Q)} - \frac{n(R_\Omega)}{\omega_-^2\omega^2}\right]. \tag{6}$$

Here $\gamma \approx 1.4$ is the polytropic ratio for air. In the case considered, the first term in this formula dominates. Indeed, for $n \sim R^{-3.5}$, the ratio of the second term to the first term is less than $(\omega_-/\omega)^{2.5} \ll 1$. For $n(R_\omega) = 700$ $\mu\mathrm{m}^{-1}\,\mathrm{m}^{-3}$ and $Q = 10$ as estimated above for a primary wave at frequency of 30 kHz, from (6) it follows that $|\epsilon_-| \approx 1800$. This value is evidently huge as compared with water without bubbles. Note that ϵ_- is complex so that the phase of nonlinearly generated low-frequency signal is shifted with respect to the primary wave.

For the sum-frequency signal generation, the corresponding formula is

$$\epsilon_+ \simeq \frac{\pi c^4}{3\omega^4} n(R_{2\omega}). \tag{7}$$

It is readily seen that the equivalent nonlinearity parameter for a coherently generated sum-frequency signal is much weaker than that for the difference-frequency signal (it is worth noting that the opposite is true for scattering from a single bubble[46]). The reason for that is compensation of harmonics generated slightly below and slightly above resonance with opposite phases whereas the difference frequency is comparable with the resonance width for a single bubble. Simple estimates show that (7) gives a value of few units, which is comparable with the nonlinearity of water without bubbles.

B. Difference-frequency signal radiation from a bubble layer near a surface

Consider now the generation of a difference-frequency wave by a bubble layer under the action of a biharmonic pump wave. For simplicity we suppose that the layer is homogeneous, of thickness l, and located immediately beneath the surface. A primary wave source of radius a is situated at a depth $H>l$, and a hydrophone at the same depth, as in Fig. 1. The wave propagating upward is then reflected from the surface and travels back to the hydrophone.

To describe the difference-frequency field P_-, we use the equation (1) with the nonlinearity parameter defined by (6). The average density and velocity of sound in the bubble layer and surrounding water are assumed to be the same. As already mentioned, the first of these assumptions is valid because of a small volume content of the bubbles (of order 10^{-7} for bubbles smaller than 0.5 mm), whereas the second is satisfied because the main contribution is due to the resonant bubbles.

We consider a bubble layer located in the far-field zone of the pump source. Since the pump frequencies are close to each other, the directivity is almost the same for them both. Thus, spatial distribution of the pump pressure amplitude in the far field zone at either frequency at a distance r from the radiator has the standard form

$$P_{1,2}(r) = \frac{A_0 R_f D(\theta,\varphi)}{r} \exp(-iK_{1,2}r - \alpha r), \tag{8}$$

where A_0 is the source amplitude factor, $R_f = \pi a^2/\lambda$ is the Fraunghoffer length for the radiator, D is the directivity diagram in polar coordinates (θ,φ); $K_{1,2} = \omega_{1,2}/c_0$ are wave numbers of the pump waves, and $\lambda_{1,2} = 2\pi/K_{1,2}$ are the corresponding wave lengths close to each other.

The difference-frequency field is defined by the solution of (1):

$$P_- = \int \frac{q_- \exp(-iK_- r)\, d\mathbf{r}}{4\pi r}, \tag{9}$$

where $K_- = \omega_-/c_0$, and integration volume is essentially formed by the part of the bubble layer of an effective thickness l lying inside the acoustic beam. Substituting (4) into (9) and integrating over the bubble volume where the primary field is nonzero, we obtain the distribution of the low-frequency signal amplitude. This integral has a rather clumsy form, but for a situation when the bubbles are located in the far field zone of both the radiator and the receiver, it can be readily found using the stationary phase method. For the case when the radiator and receiver are situated at the same distance from the layer, the difference-frequency acoustic pressure received after reflection from the surface is

$$\begin{aligned} P_-(H) &= -\frac{\epsilon_- A_0^2 R_f^2 \omega_- l m}{32\rho c^3 H^2} D^2, \\ m &= \frac{(1-\exp[-(2\alpha-\alpha_-)l])}{(\alpha_- - 2\alpha)l} \\ &\quad \times \exp(-2\alpha_- l)[\beta_- + \beta^2 \exp(-2\alpha l)]. \end{aligned} \tag{10}$$

Here α_- is the damping rate for the difference-frequency wave in the bubble layer, β_- and β are surface reflection coefficients for the difference-frequency and primary waves, respectively.

Note that for an unrealistic case of low attenuation ($\alpha l \ll 1$) and full reflection of acoustic waves from a flat surface ($\beta_- = \beta = -1$), the expression in the last parentheses tends to zero, and only the change of β due to sea roughness would provide the backscattered signal. This effect is associated with phase reversal as a result of free surface reflection.[46]

Actually, however, attenuation of primary waves is quite significant, as it was seen above, and the low-frequency pressure amplitude is finite. If one still neglects the damping at difference frequency letting $\alpha_- l = 0$ [the concentration of bubbles resonant at ω_- is too small to use the formula (2)], the factor m defining the dependence in (10) on losses at the main frequencies, reduces to

$$m = \frac{(1-\epsilon^{-2\alpha l})^2}{2\alpha l} \tag{11}$$

and reaches its maximum, $m_{\max} \simeq 0.41$, at $\alpha l \simeq 0.62$. Hence, for a smooth reflecting surface, an optimal, in this sense, case is when the primary wave amplitude decreases by about 45% upon passing the bubble layer in both directions. In our experiments, $\alpha l \simeq 1.43$ so that $m \simeq 0.33$. In this case only about 25% of pressure amplitude (1/16 of its energy) incident to the layer reaches the surface and about 6% of amplitude comes back out of it.

Evaluation of the effect of sea surface roughness on wave reflection sea surface is generally a complex problem. In our case, the surface can be considered frozen during the pulse duration, a few milliseconds, so that the temporal coherence is preserved. However, the reflected wave can be scattered in a finite angle which should be compared with the angular width of radiator and receiver diagrams. A rough estimate can be made from evaluation of the Rayleigh parameter, $\mathrm{Ra} = 2K_-\sigma$, where σ is the rms displacement of the surface (e.g., Ref. 47, Chap. 7). For large Ra, the acoustic beam is deflected at the angles of order of the rms of the wave slope, $g \sim \sigma/\Lambda$, where Λ is the characteristic surface wave length. By using the known Piersson–Moskowitz spectrum for deep-sea wind waves with wave numbers less than

K_- (the effect of shorter waves is small), we found that at the observed sea state (6 m/s wind), g is of order 0.05. For a shallow sea, this quantity should be even smaller (long-wave part of the surface spectrum is suppressed), albeit not much smaller in the considered range of wavelengths. Hence, the difference-frequency sound is scattered at 2°–3° to each side, and it remains within the diagram width of the receiver (17°). Hence, we do not take the sea roughness into account.

A convenient expression follows from (10) for the field at the beam axis ($D=1$) for the case of relatively high attenuation of primary wave, when, according to (11), $m \approx 1/2\alpha l$ (which is reasonably good in our case, $\alpha l = 1.43$, giving $m \approx 0.35$ instead of 0.33). Namely,

$$P_-(H) = \frac{\epsilon_- A_0^2 R_f^2 \omega_-}{64 \rho c^3 \alpha H^2}. \tag{12}$$

It is noteworthy that the nonlinearity parameter ϵ_- and attenuation coefficient α are both proportional to the bubble concentration,[21]

$$\epsilon_- / \alpha = 5.3 \times 10^4 \lambda Q. \tag{13}$$

Consequently, the nonlinearly scattered field (12) does not depend on bubble concentration in this case. In particular, if one uses a subsurface bubble layer to improve the efficiently of a parametric array, the field of such an array will not depend on bubble concentration. In this approximation random variations of bubble concentration do not affect the array's work. This result has an obvious restriction: the nonlinearity of the medium with bubbles must remain much larger than that of water, because the latter has been neglected in all previous calculations. Note that for smaller losses, $P_-(H)$ preserves its dependence on n, which can be used to determine the bubble concentration.

From (12), for the 5-min-old wake, the level of difference-frequency signal is estimated to be about 2.8 Pa. The experimental result is about 1.5 Pa [Fig. 5(a)]. Considering that such factors as the attenuation of the difference-frequency wave have not been taken into account in this calculation, and that the expression for ϵ_- is a rather rough approximation when ω_- is comparable with the bubble resonance curve width, the agreement can be considered reasonably good.

C. Nonlinear reverberation at sum frequency

Coherent field at sum frequency can be calculated from Eq. (1) in a way similar to that used for the difference-frequency field, with substitution of ϵ_+ from (7). As already mentioned, the nonlinearity parameter for the coherent sum-frequency signal is estimated to be of order of that in water. Besides, damping of a sum-frequency signal is even stronger than for the primary waves [due to the increase of n in (2) with the decrease of resonance radius]. As a result, the sum-frequency signal created at the receiver due to bubbles would be even weaker than that generated due to the nonlinearity of water, which does not explain the experimental results.

More adequate is a model of incoherent scattering when the intensities of signals scattered by separate bubbles in the direction of the receiver are summed. In this case, one can speak of nonlinear reverberation in a bubble layer.

Both linear and nonlinear reverberation can be characterized by a scattering coefficient β, defined as an effective scattering cross section per unit volume of a medium. In these terms, the intensity I of the field scattered by a medium volume v is[1,2]

$$I = \frac{1}{8\pi\rho c} \int_v \frac{\beta |P_1 P_2|}{r^2} d\mathbf{r}. \tag{14}$$

Here again, the scattering coefficient is determined primarily by resonant bubbles. To find the value of β_+ at the sum frequency, one should integrate over spherical waves scattered from each bubble. The corresponding analysis for the sum-frequency scattering was performed in Ref. 22 to give

$$\beta_+ = \frac{\pi^2 \gamma^2 \omega_3^4 Q^3 n(R_\omega) |P_1 P_2|}{2\rho_0^2 \omega^8 R_\omega} = \mu_+ |P_1 P_2|. \tag{15}$$

Here the coefficient μ_+ is proportional to bubble concentration at primary frequencies. In the considered frequency range, the quality factor Q is only moderately dependent on frequency,[48] and here it is supposed to be 10 again.

The geometry of the problem is the same as before, with a horizontally homogeneous layer of thickness l, and the transmitters and the receiver situated at the same depth H and directed upwards, as in Fig. 1. The acoustical pressure produced by the radiator in the far field is defined by Eq. (8), with both pump waves having the same amplitude A_0.

In our case the length of the radiated pulse, $c_0 \tau \approx 8$ m (τ is the pulse duration) exceeds both the layer thickness (1.9 m), and the characteristic length of wave damping, $\alpha^{-1} \approx 1.4$ m at primary frequencies. Hence, the result will correspond to a continuous signal. The signal received at the receiver is produced by a scattering volume having the length of order α^{-1} and a transverse size determined by the radiator's directivity diagram. Substituting (8) and (15) into (14), after simple transformations we find the average intensity of scattered field at the sum frequency,

$$I_+ = \frac{A_0^4 R_f^4 \mu_+ \Psi_+}{64\pi\rho_0 H^4 c(2\alpha + \alpha_+)}, \tag{16}$$

where

$$\Psi_+ = \int_0^{4\pi} D^4(\theta, \varphi) d\Omega, \tag{17}$$

and α_+ is the damping rate at sum frequency. To evaluate the latter, we again use the dependence $n(R) \sim R^{-3.5}$. Then, according to the formula (2), $\alpha_+ = \alpha(2\omega) \approx \sqrt{2}\alpha(\omega)$. For the value of $\alpha(\omega) = 0.72\ \mathrm{m}^{-1}$ given above, we have $\alpha_+ = 1.02\ \mathrm{m}^{-1}$.

The main features of the nonlinear reverberation that distinguish it from its linear prototype are that its efficiency grows with the source power, and the scattered signal intensity decreases with distance from the transmitters/receiver as H^{-4}, more rapidly than in the linear case.

Using the formula (16), we made estimates for a plane acoustical radiator with the radius of 15 cm, radiation fre-

quency 30 kHz and acoustical power 0.7 kW used in our experiments. The calculation for the nonlinearly scattered signal pressure made according to (16), gives the value of 6 Pa. This again slightly exceeds the observed value of 4 Pa [see Fig. 5(a)], but the agreement is still reasonably good.

Note that the above result does not depend on the bubble concentration. Indeed, according to (2) and (15), the ratio μ_+/α in (16) is independent of n. However, for short impulses when their length, $c_0\tau$, is smaller than the thickness of the scattering layer, the sum-frequency signal amplitude does depend on $n(R_\omega)$ and, consequently, short impulses can be used to measure resonant bubble concentration from sum frequency generation.

V. CONCLUSIONS

In this paper we presented the data of experiments with the bubble layer created by a propeller and with a natural bubble layer. For the first time the nonlinear backscatter of a bi-frequency signal at a difference and sum frequencies was measured in a remote-sensing configuration. The subsequent theoretical analysis permitted us to calculate the levels of a scattered signal which in both cases was in a reasonably good agreement with the experiment. Also, the bubble layer's effective thickness and concentration of resonant bubbles in it were determined. The validity of the theory is confirmed by comparison with the data of linear wave attenuation.

ACKNOWLEDGMENTS

The authors are grateful to I. M. Fuks for valuable discussions. This work was partially supported by the US Office of Naval Research and the Russian Foundation of Basic Research (Grant No. 00-05-64252).

[1] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977).
[2] R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
[3] D. M. Farmer and S. Vagle, "Waveguide propagation of ambient sound the ocean-surface bubble layer," J. Acoust. Soc. Am. **86**, 1897–1908 (1989).
[4] M. J. Buckingham, "Sound speed and void fraction profiles in the sea surface," Appl. Acoust. **51**, 225–250 (1997).
[5] N. Q. Lu, A. Prosperetti, and S. W. Yoon, "Underwater noise emission from bubble clouds," IEEE J. Ocean. Eng. **15**, 275–281 (1990).
[6] S. A. Thorpe, "On the clouds of bubbles formed by breaking waves in deep water and their role in air-sea gas transfer," Philos. Trans. R. Soc. London, Ser. A **304**, 155–210 (1982).
[7] D. K. Woolf and S. A. Thorpe, "Bubbles and the air-sea exchange of gases in near saturation conditions," J. Mar. Res. **49**, 235–466 (1991).
[8] E. C. Monahan, "Occurrence and evolution of acoustically relevant subsurface bubble plumes and their associated, remotely monitorable, surface whitecaps," in *Natural Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht, 1993), pp. 503–517.
[9] B. D. Johnson and R. C. Cooke, "Bubble populations and spectra in coastal waters: A photographic approach," J. Geophys. Res., C: Oceans Atmos. **84**, 3761–3766 (1979).
[10] P. A. Kolobaev, "Investigation of the concentration and statistical size distribution of wind-produced bubbles in the near-surface ocean layer," Oceanology **15**, 659–661 (1976).
[11] H. Medwin, "*In situ* acoustic measurements of bubble population in coastal ocean waters," J. Geophys. Res. **75**, 599–611 (1970).
[12] H. Medwin, "Acoustical determination of bubble-size spectra," J. Acoust. Soc. Am. **62**, 1041–1044 (1977).
[13] N. Breitz and H. Medwin, "Instrumentation for in situ acoustical measurements of bubble spectra under breaking waves," J. Acoust. Soc. Am. **86**, 739–743 (1989).
[14] D. M. Farmer and S. Vagle, "Bubble measurements using a resonator system," in *Natural Physical Processes Associated with Sea Surface Sound*, edited by T. G. Leighton (University of Southampton, Highland, UK, 1997), pp. 155–162.
[15] A. Lövik, "Acoustic measurements of gas bubble spectrum in water," in *Cavitation and Inhomogeneities in Underwater Acoustics*, edited by W. Lauterborn (Springer-Verlag, Berlin–Heidelberg–New York, 1980), pp. 211–218.
[16] I. R. Schippers, "Density of air-bubbles below the sea surface, theory and experiments," in Ref. 15, pp. 205–210.
[17] B. McGarthey and B. McBary, "Echo-sounding on probable gas bubbles from the bottom of Saanich Inlet, British Columbia," Deep-Sea Res. **12**, 285–294 (1965).
[18] V. Akulichev, V. Bulanov, and S. Klenin, "Acoustic probing of gas bubbles in the sea," Sov. Phys. Acoust. **32**, 177–180 (1986).
[19] S. O. McConnell, "Acoustic measurements of bubble densities at 15–50 kHz," in *Natural Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht, 1993), pp. 237–252.
[20] M. Gensane, "Bubble population measurements with a parametric array," J. Acoust. Soc. Am. **95**, 3183–3190 (1994).
[21] L. A. Ostrovsky and A. M. Sutin, "Nonlinear acoustic diagnostics of discrete inhomogeneities in liquids and solids," *Proceedings of the 11th International Congress on Acoustics*, Paris, 1983, Vol. 2, pp. 137–140.
[22] L. A. Ostrovsky and A. M. Sutin, "Nonlinear sound scattering from subsurface bubble layer," in *Natural Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer Academic, Dordrecht, 1993), pp. 363–370.
[23] F. N. Fenlon and J. W. Wohn, "On the Amplification of Modulated Acoustic Waves in Gas-Liquid Mixtures," in *Cavitation and Inhomogeneities in Underwater Acoustics*, edited by W. Lauterborn (Springer-Verlag, Berlin–Heidelberg–New York, 1980), pp. 141–150.
[24] D. L. Miller, "Ultrasonic detection of resonance cavitation bubbles in a flow tube by their second-harmonic emissions," Ultrasonics **19**, 217–224 (1981).
[25] P. J. A. Frinking, A. Bouakaz, J. Kirkhorn, F. Tencate, and Nico de Jong, "Ultrasound contrast imaging: Current and new potential," Ultrasound Med. Biol. **26**, 965–975 (2000).
[26] P. H. Chang, K. K. Shung, and H. B. Levene, "Quantitative measurements of second harmonic Doppler using ultrasound contrast agents," Ultrasound Med. Biol. **22**, 1205–1214 (1996).
[27] S. Krishnan and M. O'Donnel, "Transmit aperture processing for nonlinear contrast agent imaging," Ultrason. Imaging **18**, 77–105 (1996).
[28] G. Gimenez, M. Chamant, and J. P. Farnand, "Non-linear response of a single bubble driven by a two-components exciting wave," *Proceedings of the 10th International Symposium on Nonlinear Acoustics*, Kobe, Japan, 1984, pp. 83–87.
[29] V. L. Newhouse and P. M. Shankar, "Bubble size measurement using the nonlinear mixing of two frequencies," J. Acoust. Soc. Am. **75**, 1473–1477 (1984).
[30] J. Y. Chapelon, P. M. Shankar, and V. L. Newhouse, "Ultrasonic measurement of bubble cloud size profiles," J. Acoust. Soc. Am. **78**, 196–201 (1985).
[31] A. M. Sutin, S. W. Yoon, E. J. Kim, and I. N. Didenkulov, "Nonlinear acoustic method for bubble density measurements in water," J. Acoust. Soc. Am. **103**, 2377–2384 (1998).
[32] B. M. Sandler, D. A. Selivanovsky, and A. Yu. Sokolov, "New results regarding bubble concentration with radii from 6 to 20 m at sea," Sov. Phys. Tech. Phys. **27**, 1038–1039 (1982).
[33] D. Phelps and T. G. Leighton, "Oceanic bubble population measurements using a buoy-deployed combination frequency technique," IEEE J. Ocean. Eng. **23**, 4, 400–410 (1998).
[34] D. Phelps, D. G. Ramble, and T. G. Leighton, "The use of a combination frequency technique to measure the surf zone bubble population," J. Acoust. Soc. Am. **101**, 1981–1989 (1997).
[35] A. M. Sutin, "Experimental investigations of nonlinear coherent and incoherent scattering from bubble clouds," in *Natural Physical Processes Associated with Sea Surface Sound*, edited by T. G. Leighton (University of Southampton, Highfield, UK, 1997), pp. 211–218.
[36] D. Phelps and T. G. Leighton, "High resolution bubble sizing through detection of the subharmonic response with a two frequency technique," J. Acoust. Soc. Am. **99**, 1985–1992 (1996).

[37] B. K. Novikov, O. V. Rudenko, and V. I. Timoshenko, *Nonlinear Hydroacoustics* (AIP Press, New York, 1987).
[38] H. C. Woodsum, ''Enhancement of parametric efficiency by saturation suppression,'' J. Sound Vib. **69**, 27–33 (1980).
[39] A. M. Lerner and A. M. Sutin, ''Influence of gas bubbles on the field of a parametric sound radiator,'' Sov. Phys. Acoust. **29**, 388–390 (1983).
[40] L. M. Kustov, V. E. Nazarov, L. A. Ostrovsky, A. M. Sutin, and A. S. Zamolin, ''Parametric acoustic radiator with a bubble layer,'' Acoust. Lett. **6**, 15–17 (1982).
[41] V. E. Nazarov and A. M. Sutin, ''Far-field characteristics of a parametric sound radiator with a bubble layer,'' Sov. Phys. Acoust. **30**, 477–479 (1984).
[42] L. M. Kustov, V. E. Nazarov, and A. M. Sutin, ''Nonlinear sound scattering by a bubble layer,'' Sov. Phys. Acoust. **32**, 500–503 (1986).
[43] T. Asada and Y. Watanabe, ''Experiments of parametric amplification using nonlinear vibration of bubbles under water,'' in *Frontiers of Nonlinear Acoustics: Proceedings of 12th ISNA*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier Science, London, 1990), pp. 85–490.
[44] O. A. Druzhinin, L. A. Ostrovsky, and A. Prosperetti, ''Low-frequency acoustic wave generation in a resonant bubble-layer,'' J. Acoust. Soc. Am. **100**, 3570–3580 (1996).
[45] L. A. Ostrovsky, A. M. Sutin, I. A. Soustova, A. I. Matveev, and A. I. Potapov, ''Nonlinear, low-frequency sound generation in a bubble layer: theory and laboratory experiment,'' J. Acoust. Soc. Am. **104**, 722–726 (1998).
[46] K. A. Naugolnykh and L. A. Ostrovsky, *Nonlinear Processes in Acoustics* (Cambridge University Press, New York–Cambridge, 1998).
[47] F. G. Bass and I. M. Fuks, *Wave Scattering from Statistically Rough Surfaces* (Pergamon, New York, 1979).
[48] Ch. Devin, Jr., ''Survey of thermal, radiation, and viscous damping of pulsating air bubbles in water,'' J. Acoust. Soc. Am. **31**, 1654–1657 (1959).

The sound-speed gradient and refraction in the near-ground atmosphere

D. Keith Wilson[a)]
U.S. Army Research Laboratory, AMSRL-CI-EE, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

(Received 28 December 2000; revised 22 October 2002; accepted 31 October 2002)

A systematic description of sound refraction in the near-ground atmosphere is developed by modeling the effective sound-speed gradient with Monin–Obukhov similarity theory. The resulting gradient equation can be recast in a form involving just three nondimensional variables. The first is the ratio of a sound-speed scale (representing the strength of the turbulent fluctuations in the sound speed) to the friction velocity. The second is the ratio of the actual height to a transitional height where contributions from the near-ground wind-speed gradients and the adiabatic lapse rate are roughly balanced. The third is simply the cosine of the angle between the propagation direction and mean wind direction. When the magnitude of the sound-speed scale/friction velocity ratio is large, refraction is unconditionally upward or downward, depending on sign of the ratio. A small value for this ratio indicates nearly neutral atmospheric stratification, for which refraction is determined by the wind direction for small values of the nondimensional height and is upward for larger values. The contribution to refraction from air humidity is determined as a function of the Bowen ratio and found to be significant over wet surfaces. Weather conditions appropriate for measurement of sound pressure levels are also discussed.
[DOI: 10.1121/1.1532028]

PACS numbers: 43.28.Fp, 43.50.Vt [LCS]

I. INTRODUCTION

The atmospheric surface layer (roughly the 50–100 m of the atmosphere nearest the ground) is characterized by sharp vertical gradients in the wind speed, temperature, and humidity. These gradients refract sound, creating substantial diminishment or enhancement of sound levels relative to what would be observed in a homogeneous atmosphere. The Monin–Obukhov turbulence similarity theory is now a well-established method for modeling gradients in thermally stratified flows such as the atmospheric surface layer.[1–3] Several previous authors, including Klug[4] and L'Espérance *et al.*,[5] have applied Monin–Obukhov similarity on a case-by-case basis to refraction calculations in the atmosphere.

The purpose of this paper is to systematically analyze the implications of Monin–Obukhov similarity for the surface-layer effective sound-speed gradient, including the effects of temperature, humidity, and wind speed. By reducing the problem to a minimal number of dimensionless parameters, the analysis provides very general insights into the dependence of sound refraction on near-ground atmospheric wind and stability conditions. In particular, a set of distinctive refraction regimes, whose boundaries have a quantifiable dependence on the dimensionless parameters, can be identified. A nondimensional equation for the effective sound-speed gradient is developed in Sec. II. This section includes an analysis of the significance of the humidity contribution to the gradient. The behavior of the nondimensional gradient equation is discussed and illustrated with ray traces in Sec. III. Atmospheric conditions that produce a nearly constant or zero gradient are also discussed.

II. THEORY

Refraction of sound rays in a moving medium such as the atmosphere is described by[6]

$$\frac{d\mathbf{n}}{dt} = -\nabla_\perp c - \sum_{i=1}^{3} n_i \nabla_\perp u_i , \tag{1}$$

where $\mathbf{n}$ is the wave front normal, t is time, c is the sound speed, $\mathbf{u}$ is the medium velocity, and $\nabla_\perp = \nabla - \mathbf{n}(\mathbf{n}\cdot\nabla)$ is the gradient parallel to the wave fronts. Considering only refraction of nominally horizontal rays by the mean vertical gradients of sound speed and wind, we may neglect the $i=3$ term in Eq. (1) and move n_1 and n_2 under the differentiation in the $i=1$ and $i=2$ terms. The equation can then be written

$$\frac{d\mathbf{n}}{dt} = -\nabla_\perp c_{\text{eff}} , \tag{2}$$

in which

$$c_{\text{eff}}(z,\alpha) = \langle c(z)\rangle + \langle U(z)\rangle \cos\,\alpha \tag{3}$$

is called the effective sound speed,[7] z is the height above the ground, U is the horizontal wind speed, and α is the angle between the source-receiver path and the wind vector. The angular brackets indicate ensemble means, which are assumed in this paper to depend only on height. Equation (2) shows that refraction of the sound is determined primarily by the gradient of the effective sound speed; gradients such that $\partial c_{\text{eff}}/\partial z < 0$ result in upward refraction of sound waves, whereas $\partial c_{\text{eff}}/\partial z > 0$ results in downward refraction. There-

[a)]Present address: U.S. Army Cold Regions Research Laboratory, CEERD-RC, 72 Lyme Rd., Hanover, NH 03755-1290.

fore, the behavior of this gradient controls the basic nature of the sound propagation. The main purpose of this section is to develop a general equation for $\partial c_{\text{eff}}/\partial z$ in the near-ground atmosphere.

The sound speed in air is given by[8]

$$c=\sqrt{\gamma_a R_a T(1+\eta q)}, \tag{4}$$

where $\gamma=C_p/C_v$ is the ratio in air of the specific heats for constant pressure and constant volume, R is the gas constant, the subscript a indicates dry air, $\eta=0.511$, T is temperature, and q is the water vapor mixing ratio (mass of vapor divided by mass of dry air in a sample). Let us write $T=T_0+\Delta T$ and $q=q_0+\Delta q$, where T_0 and q_0 are constant reference values for the temperature and humidity (for example, the mean values at the standard atmospheric observation height of 2 m), and ΔT and Δq are small perturbations whose mean values generally depend on height. Substituting these expansions into Eq. (4) and keeping only the first-order contributions in ΔT, Δq, and q_0, one has

$$c=c_0\left[1+\frac{1}{2}\left(\frac{\Delta T}{T_0}+\eta\Delta q\right)\right], \tag{5}$$

where $c_0=\sqrt{\gamma_a R_a T_0(1+\eta q_0)}$. The result of substituting Eq. (5) into Eq. (3) and differentiating with respect to height is

$$\frac{\partial c_{\text{eff}}}{\partial z}=\frac{c_0}{2T_0}\frac{\partial\langle\Delta T\rangle}{\partial z}+\frac{c_0\eta}{2}\frac{\partial\langle\Delta q\rangle}{\partial z}+\cos\alpha\frac{\partial\langle U\rangle}{\partial z}. \tag{6}$$

The quantities $\partial\langle\Delta T\rangle/\partial z$ and $\partial\langle\Delta q\rangle/\partial z$ in Eq. (6) can be replaced by $\partial\langle T\rangle/\partial z$ and $\partial\langle q\rangle/\partial z$, since T_0 and q_0 are height-independent.

For simplicity, let us initially consider only the contributions from the temperature and wind gradients to $\partial c_{\text{eff}}/\partial z$. According to the Monin–Obukhov similarity theory,[1–3] the gradients of temperature and wind can be written in the following forms:

$$\frac{kz}{P_t T_*}\left(\frac{\partial\langle T\rangle}{\partial z}+\Gamma_d\right)=\phi_h(\zeta), \tag{7}$$

and

$$\frac{kz}{u_*}\frac{\partial\langle U\rangle}{\partial z}=\phi_m(\zeta), \tag{8}$$

where $k=0.40$ is von Kármán's constant, $P_t=0.95$ is the turbulent Prandtl number in neutral stratification, $\Gamma_d=g/C_p=0.0098$ K/m is the dry adiabatic lapse rate (accounting for the decrease of temperature with height due to compression in the air column), g is gravitational acceleration, u_* is the friction velocity, $T_*=-\langle w'T'\rangle_s/u_*$ is a temperature scale, and $\langle w'T'\rangle_s$ is the covariance of vertical velocity (w) and temperature at the surface. (The primes indicate the fluctuation of a quantity about the mean value at that height, e.g., $T'=T-\langle T\rangle$.) The ϕ's are heat and mass transfer functions, dependent on the dimensionless height ratio $\zeta=z/L_o$, where $L_o=-u_*^3 T_0/kg\langle w'T'\rangle_s$ is the Obukhov length. These functions are expected to have a universal dependence on ζ for any well-mixed turbulent layer over a reasonably homogeneous fetch (flat ground with uniform roughness elements). The following forms for ϕ_h and ϕ_m are recommended based on Högström[9] and Wilson:[10]

$$\phi_{h,m}(\zeta)=\begin{cases}(1+a_{h,m}|\zeta|^{2/3})^{-1/2}, & \zeta<0,\\ 1+b_{h,m}\zeta, & \zeta\geqslant 0,\end{cases} \tag{9}$$

where the a's and b's are constants with the values $a_h=7.9$, $a_m=3.6$, $b_h=8.4$, and $b_m=5.3$. The case $\zeta<0$ corresponds to buoyantly unstable conditions, which typically occur when the sun heats the ground. Buoyantly stable conditions ($\zeta>0$) typically occur when the ground cools at night. In the limit $|\zeta|\to 0$, the gradient functions equal 1 and the wind and temperature profiles take on their familiar logarithmic forms.[3] Of course, the gradient functions can be integrated to determine the actual height dependence of the wind speed and temperature. The interested reader may refer to Refs. 2, 3, and 10 for the customary procedure. The result is

$$\langle T(z)\rangle=\langle T(z_r)\rangle-\Gamma_d(z-z_r)+\frac{P_t T_*}{k}\left[\ln\frac{z}{z_r}-\Psi_h\left(\frac{z}{L_o}\right)+\Psi_h\left(\frac{z_r}{L_o}\right)\right] \tag{10}$$

and

$$\langle U(z)\rangle=\frac{u_*}{k}\left[\ln\frac{z}{z_0}-\Psi_m\left(\frac{z}{L_o}\right)+\Psi_m\left(\frac{z_0}{L_o}\right)\right], \tag{11}$$

where z_r is a reference height at which the temperature is known, z_0 is the surface roughness length, and

$$\Psi_{h,m}(\zeta)=\begin{cases}2\ln[(1+\phi_{h,m}^{-1})/2], & \zeta<0,\\ -b_{h,m}\zeta, & \zeta\geqslant 0.\end{cases} \tag{12}$$

Substituting Eqs. (7) and (8) into Eq. (6), we can write the vertical dependence of the effective sound-speed gradient in the following dimensionless form:

$$\frac{kz}{u_*}\frac{\partial c_{\text{eff}}}{\partial z}=\frac{c_0 P_t T_*}{2T_0 u_*}\phi_h(\zeta)-\frac{\Gamma_d c_0 kz}{2T_0 u_*}+\cos(\alpha)\phi_m(\zeta). \tag{13}$$

Using the thermodynamic relation $C_p=C_v+R$ and sound-speed equation for an ideal gas, $c_0^2=\gamma RT_0$, we find $\Gamma_d=gT_0(\gamma-1)/c_0^2$. This relationship allows Eq. (13) to be rewritten as

$$\frac{kz}{u_*}\frac{\partial c_{\text{eff}}}{\partial z}=P_t A\phi_h(\zeta)-\frac{\gamma-1}{4A}\zeta+\cos(\alpha)\phi_m(\zeta), \tag{14}$$

where

$$A=\frac{c_0 T_*}{2T_0 u_*}. \tag{15}$$

Alternatively, we could divide Eq. (14) through by A, and then have an expression for the effective sound-speed gradient nondimensionalized by $2T_0 kz/c_0 T_*$. In either case the dimensionless gradient depends on the variables A, ζ, and $\cos\alpha$. The first term, the contribution from the potential temperature gradient, is positive in stable conditions ($\zeta, T_*>0$) and negative in unstable conditions ($\zeta, T_*<0$). The second term, originating from the adiabatic temperature

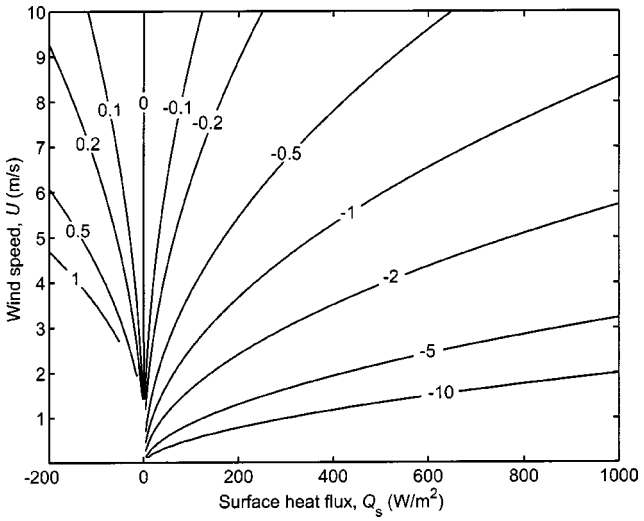


FIG. 1. Dependence of the parameter A (ratio of the sound-speed scale, c_*, to the friction velocity, u_*) on the surface heat flux and wind speed at 2 m height. The calculation is for a roughness length $z_0=0.01$ m.

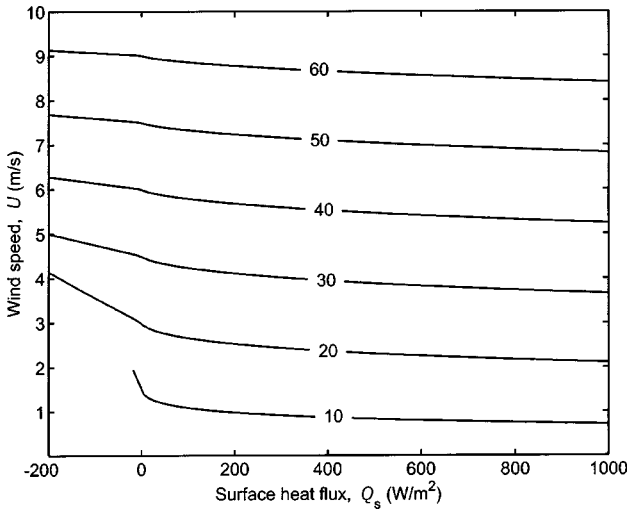


FIG. 2. Dependence of the length scale L_g on the surface heat flux and wind speed at 2 m height. This scale indicates the height at which contributions to the effective sound-speed gradient from wind gradients and the adiabatic lapse rate are approximately in balance. The calculation is for a roughness length $z_0=0.01$ m. Contour values are in meters.

lapse, is always negative since A and ζ have the same sign. The third term, from the wind gradient, is positive for downwind propagation and negative for upwind propagation. The ratio A can be expressed in a useful alternative form by defining a sound-speed scale as $c_*=-\langle w'c'\rangle_s/u_*$. From Eq. (5) (without the humidity term) we have $c'=(c_0/2T_0)T'$, leading to $c_*=(c_0/2T_0)T_*$. Therefore, equivalently to Eq. (15), one has

$$A=\frac{c_*}{u_*}. \tag{16}$$

Hence A is simply the ratio of the sound-speed scale to the friction velocity; physically, it can be thought of as representing the strength of the fluctuations in the actual sound speed divided by the strength of the fluctuations in the wind speed. The dependence of A on the underlying atmospheric forcings of wind speed and sensible heat flux, $Q_s=\rho_0 C_p\langle w'T'\rangle_s=-\rho_0 C_p u_* T_*$, is illustrated in Fig. 1 by contours of constant values of A for a range of these forcing parameters. These calculations were performed by first solving for u_* from $\langle U(z)\rangle$ in Eq. (11). Although the equation is nonlinear in u_* (recall that L_o depends on u_*), it can be solved by a simple iterative method.[5] A wind-speed observation height $z=2$ m and surface roughness $z_0=0.01$ m (characteristic of fairly level grass plains[2]) were used. Once u_* is determined, the ratio A follows from Q_s. In fair, summer weather at midlatitudes, the heat flux varies roughly over the range -50 W/m$^2<Q_s<600$ W/m^2 during a diurnal cycle.[2]

Although the nondimensionalization in Eq. (14) is a natural one from the standpoint of the Monin–Obukhov theory, it is somewhat cumbersome to apply systematically to sound refraction because the normalizing factor, kz/u_*, depends on height. An alternative formulation follows by defining a new nondimensional height

$$\bar{z}=\frac{z}{L_g},\quad L_g=\frac{c_0 u_*}{kg}. \tag{17}$$

The quantity $\bar{z}$ indicates the relative importance of wind gradients and the adiabatic temperature lapse rate for sound refraction. For $\bar{z}\ll 1$, wind gradients are dominant. As $\bar{z}$ increases, the wind gradients weaken and eventually, when $\bar{z}\gg 1$, the adiabatic temperature lapse rate controls refraction. The length scale L_g can be thought of as a transition height where the contributions to $\partial c_{\rm eff}/\partial z$ from wind gradients and the adiabatic temperature lapse are roughly in balance. Contours of constant values of L_g, as a function of the wind speed and sensible heat flux, are plotted in Fig. 2.

Dividing Eq. (14) by $\bar{z}$, an equation is produced in which the gradient is normalized by the constant factor c_0/g:

$$\frac{c_0}{g}\frac{\partial c_{\rm eff}}{\partial z}=\frac{P_t A}{\bar{z}}\phi_h(\zeta)-\frac{\gamma-1}{2}+\frac{\cos\alpha}{\bar{z}}\phi_m(\zeta). \tag{18}$$

In deriving Eq. (18), use has been made of the relationship $\zeta=2A\bar{z}$, which follows from the definition of ζ and Eqs. (15) and (17).

Equation (18) is the main result of this paper. It shows that the effective sound-speed gradient can be determined from just three nondimensional parameters, A, $\bar{z}$, and $\cos\alpha$, that depend on the conditions of the near-ground atmosphere. A useful application of the equation is determination of the angle α_0 at which the gradient vanishes (i.e., at which there is no net refraction). This angle, which depends on the height from the ground, is

$$\alpha_0=\cos^{-1}\left[\frac{(\gamma-1)\bar{z}-2P_t A\phi_h(\zeta)}{2\phi_m(\zeta)}\right]. \tag{19}$$

Furthermore, for a given propagation direction, it is possible in some circumstances to solve Eq. (18) for a height $\bar{z}$ at which the gradient vanishes. For the particularly simple case of neutral conditions ($A=0$ and $\phi_h=\phi_m=1$), one has $\bar{z}=2(\cos\alpha)/(\gamma-1)$. In downwind propagation ($\alpha=0$), we therefore have $\bar{z}=5$, implying that a ray launched horizon-

tally from a source at this height would continue to travel horizontally. But $\bar{z}=5$ turns out to be very high above the ground in most atmospheric conditions: for a typical value of $u_*=0.3$ m/s, $\bar{z}=5$ corresponds to 130 m.

Let us now consider how the presence of water vapor modifies the preceding analysis. The Monin–Obukhov form of the humidity gradient is $\partial q/\partial z=(P_q q_*/kz)\phi_w(\zeta)$, where $q_*=-\langle w'q'\rangle_s/u_*$. The constant P_q and the function ϕ_w are usually assumed to be the same as P_t and ϕ_h.[3] The temperature gradient is still given by Eq. (7) in humid air, although the actual temperature should strictly be replaced by the virtual temperature $T_v=T(1+\mu q)$ (including replacing $\langle w'T'\rangle_s$ with $\langle w'T_v'\rangle_s$ in the definition of the Obukhov length), where $\mu=0.607$. [See, for example, Eqs. (2.24), (2.32), (3.25) in Garratt.[3] The purpose of the virtual temperature is to account for the effect of water vapor (which is less dense than dry air) on buoyancy.] However, if we assume $\phi_w=\phi_h$, Eq. (7) can be applied to either the virtual or actual temperature. The net result is that Eqs. (18) and (19) apply to humid as well as dry air, although A must be defined from Eq. (16) with

$$c_*=\frac{c_0}{2}\left(\frac{T_*}{T_0}+\eta q_*\right). \tag{20}$$

Furthermore, ζ is no longer exactly $2A\bar{z}$, but rather

$$\zeta=2A\bar{z}\frac{1+\mu q_* T_0/T_*}{1+\eta q_* T_0/T_*}. \tag{21}$$

The quantity q_*/T_* is proportional to the Bowen ratio β, which is defined as the ratio of the sensible to the latent heat flux at the surface. (The role of the Bowen ratio in determining the relative contributions of temperature and humidity to the acoustic index-of-refraction was previously pointed out by Wesely.[11]) Specifically,

$$\beta=\frac{\rho_s C_p\langle w'T'\rangle_s}{\rho_s L_v\langle w'q'\rangle_s}=\frac{C_p T_*}{L_v q_*}, \tag{22}$$

where ρ_s is the density of air at the surface and L_v is the latent heat of vaporization for water.[2] Substituting for q_*/T_* and replacing the constants with numerical values, Eqs. (20) and (21) become

$$c_*=\frac{c_0 T_*}{2T_0}\left(1+\frac{0.061}{\beta}\right), \tag{23}$$

and

$$\zeta=2A\bar{z}\frac{0.073+\beta}{0.061+\beta}. \tag{24}$$

Stull[2] (p. 273) indicates that the typical range of values for β is from 0.1 over water to 5 over semi-arid regions. Small values produce the largest changes in c_* and ζ in comparison to dry air. One finds that when $\beta=0.5$, a typical value for grassland, c_* exceeds its dry value of $(c_0/2T_0)T_*$ by 12% whereas ζ exceeds its dry value of $2A\bar{z}$ by only 2%. For a more extreme, $\beta=0.2$ (characteristic of saturated ground), the corrections amount to 30% and 5%, respectively. Therefore the contribution from the humidity gradient to the sound-speed gradient can be significant, particularly over a wet surface, although the contribution from the temperature gradient is still almost always larger. The humidity correction to ζ can be neglected.

If the air is saturated (as it would be in fog or in a cloud), the dry adiabatic lapse rate in Eq. (7) should no longer be used. The adiabatic lapse rate in saturated air varies between roughly 0.004 and 0.007 K/m, depending on the air temperature.[12]

The validity of the analysis in this section is of course limited to situations where the Monin–Obukhov theory is valid. The theory requires that the near-ground atmosphere can be modeled as a constant vertical-flux layer; that is, the covariances of vertical velocity with quantities such as temperature and horizontal velocity must all be nearly equal to their surface values throughout the layer. As a result, the predictions are restricted to heights $z<0.1z_i$, where z_i is the thickness of the overall atmospheric boundary layer (about 500–2000 m). Furthermore, applications of the Monin–Obukhov similarity theory is tenuous when $\zeta\gtrsim 1$ due to weak turbulent mixing and stratification in the surface layer during buoyantly stable conditions (Garratt,[3] p. 50). Unfortunately, there is no known surface-layer similarity theory for this situation. It should also be pointed out that many alternative parametrizations are possible besides the one developed in this paper. For example, u_* could be replaced by c_* in the definition for $\bar{z}$, Eq. (17). The resulting ratio would represent the relative contributions of the near-ground sound-speed gradient and the temperature lapse to $\partial c_{\text{eff}}/\partial z$.

III. RESULTS AND DISCUSSION

Figure 3 shows the sector of positive effective sound-speed gradient (and hence downward refraction) as a function of $\bar{z}$ and A, determined from Eq. (19). A contour value of δ degrees indicates downward refraction for $-\delta/2<\alpha<\delta/2$ and upward refraction for other angles. Figure 4 shows the gradient calculated from Eq. (18) for downwind ($\alpha=0°$), crosswind ($\alpha=\pm 90°$), and upwind ($\alpha=180°$) propagation. The reference temperature T_0 was set to 293 K in the figures. (The particular value of T_0 has very little effect on the results.) The calculations all assume the relationship $\zeta=2A\bar{z}$; as discussed in Sec. II, this relationship is exact for dry air but is still a good approximation in humid air. The dashed lines in Figs. 3 and 4 are placed at $\zeta=1$. Recall that Monin–Obukhov similarity is tenuous for $\zeta\gtrsim 1$ (near and above the dashed line), although the predictions should at least be qualitatively correct.

On the basis of Fig. 3, the behavior of the gradient can be categorized into three distinct regimes depending on the value of A.

(1) $A\lesssim -1$: Refraction of sound is upward for all propagation directions relative to the wind. This regime is dominated by a strong negative sound-speed (temperature and/or humidity) gradient, typically resulting from intense solar heating of the ground during the daytime or possibly from a very moist ground. Figure 5, a ray trace for $A=-2$, illustrates the refractive characteristics of this regime. (The methodology by which this ray trace and the others in this paper were created is discussed in the Appendix.)

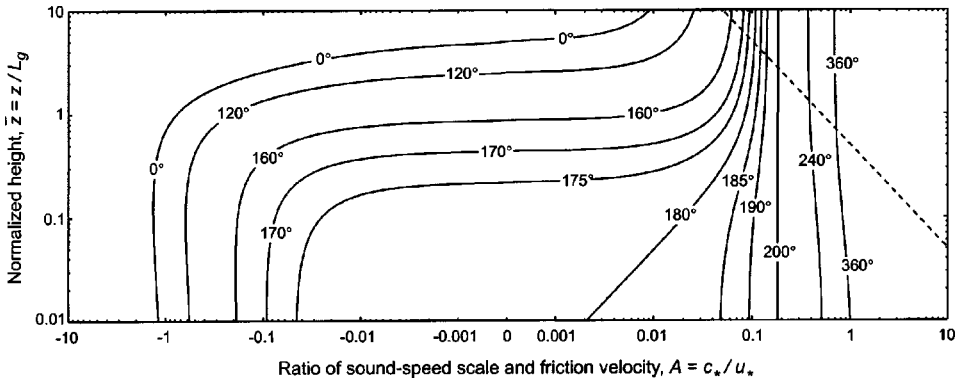


FIG. 3. Angular extent of the sector (in degrees) for which the effective sound-speed gradient is positive. A contour value of δ degrees indicates positive gradient (downward refraction) for $-\delta/2<\alpha<\delta/2$ and negative gradient (upward refraction) for other angles, where α is the angle between the propagation and wind directions. (Therefore the contour $\delta=0°$ implies upward refraction for all propagation directions and $\delta=360°$ implies downward refraction for all propagation directions.) The dashed line is placed at $\zeta=z/L_o=1$, above which application of the Monin–Obukhov similarity theory is tenuous.

(2) $-0.1\lesssim A\lesssim 0.1$: Refraction by wind gradients dominates near the ground ($\bar{z}\lesssim 0.1$). The effective sound-speed gradient is positive for downwind propagation and negative for upwind propagation (to within about $\pm 10°$). However, when $\bar{z}\gtrsim 1$, upward refraction prevails for all propagation directions, as a result of the adiabatic lapse in temperature. A ray trace for $A=0$ is shown in Fig. 6.

(3) $A\gtrsim 1$: Refraction of sound is downward for all propagation directions relative to the wind. This regime is dominated by a strong positive sound-speed (temperature) gradient, which frequently occurs on clear nights with light winds. A ray trace for $A=1$ is shown in Fig. 7.

Values such that $0.1\lesssim|A|\lesssim 1$ represent transitions between these regimes.

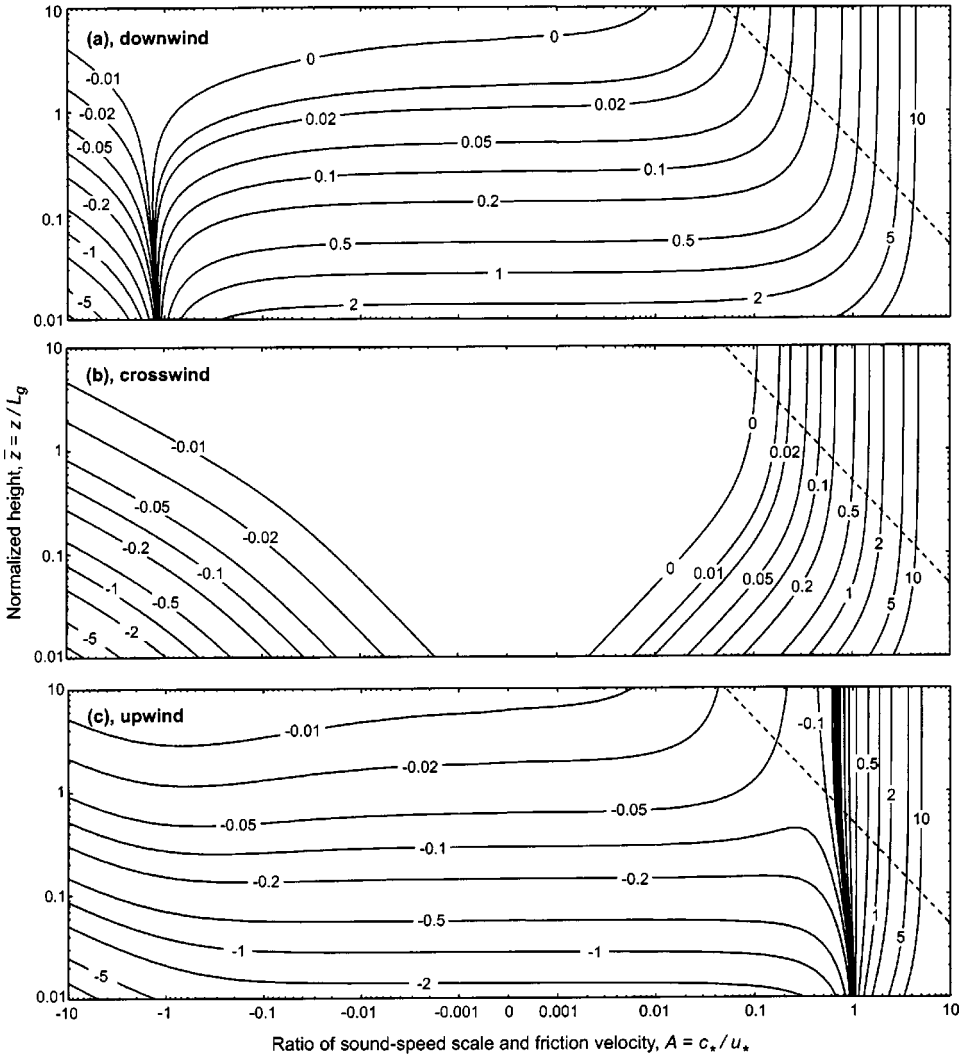


FIG. 4. Effective sound-speed gradient (s^{-1}) determined from Monin–Obukhov similarity. Axes are the same as Fig. 1. (a) Downwind propagation, $\alpha=0°$. (b) Crosswind propagation, $\alpha=\pm 90°$. (c) Upwind propagation, $\alpha=180°$.

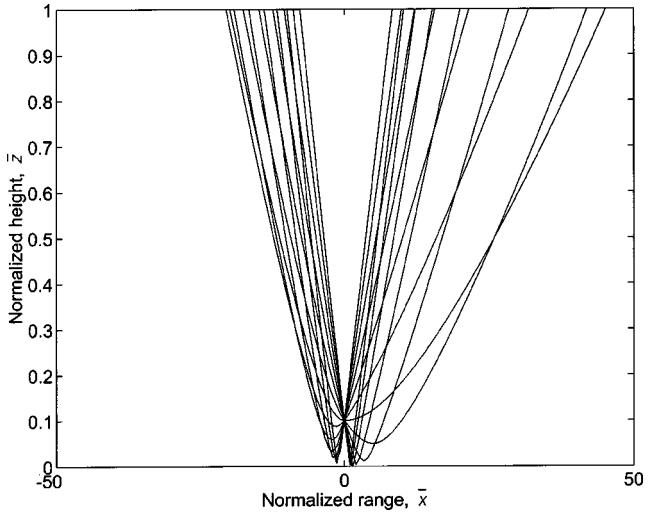


FIG. 5. Ray trace for $A=-2$. This trace illustrates the refractive regime $A\lesssim -1$, in which propagation is controlled primarily by a negative sound-speed gradient. The vertical cross section shows propagation upwind (to the left of the source) and downwind (to the right). The source height is $\bar{z}=0.1$ and rays launched between $\pm 6°$ at 1° increments are shown. The normalized coordinates are converted to dimensional ones by multiplication with the length scale L_g, which takes on a value of about 10 m in light wind conditions and 50 m in high wind conditions.

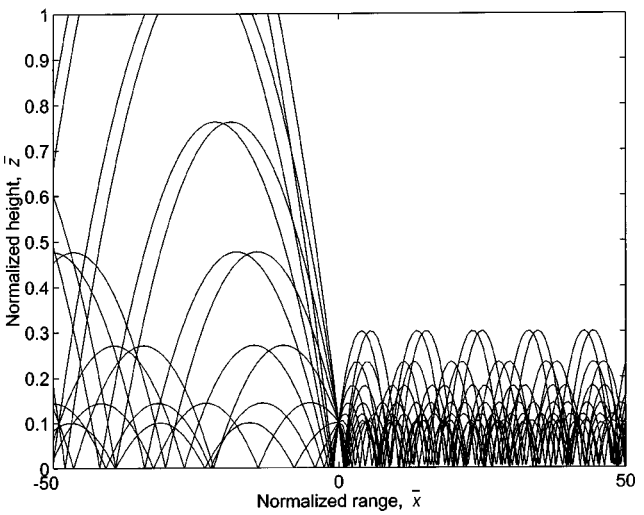


FIG. 7. Ray trace for $A=1$. This trace illustrates the refractive regime $A\gtrsim 1$, in which propagation is controlled primarily by a positive sound-speed gradient. As in Fig. 5, upwind and downwind propagation are shown.

Figures 3 and 4 provide insight into the applicability of certain modeling assumptions that are commonly made for the effective sound-speed gradient. For example, many propagation calculations are based on the simplifying assumption that the effective sound-speed gradient $\partial c_{\text{eff}}/\partial z$ is constant (i.e., the effective sound speed has a linear height dependence). Such constant-gradient cases appear in Fig. 4 as vertically oriented contours. We observe that the atmospheric surface layer has an approximately constant gradient only (1) for very stable conditions ($A\gtrsim 1$), (2) for the downward direction in a very narrow region around $A\simeq -1$, and (3) for the crosswind direction in neutral conditions ($A\simeq 0$). More typically (for $A\lesssim 1$ in all three propagation directions shown in Fig. 4) the gradient has a large absolute value near the ground and becomes small and negative when $\bar{z}\gtrsim 1$. The even more restrictive assumption that $\partial c_{\text{eff}}/\partial z\simeq 0$ (a nonrefractive atmosphere) holds only for very specific conditions: (1) for the upwind direction in a very narrow region around $A\simeq 1$, (2) for the downwind direction in a very narrow region around $A\simeq -1$, and (3) for the crosswind direction in neutral conditions ($A\simeq 0$). *In no atmospheric condition are refractive effects small for all propagation directions*. Thus, the constant or zero effective sound-speed gradient and corresponding approximations, although useful for deriving analytical solutions of propagation phenomena, rarely approximate realistic conditions in the atmospheric surface layer.

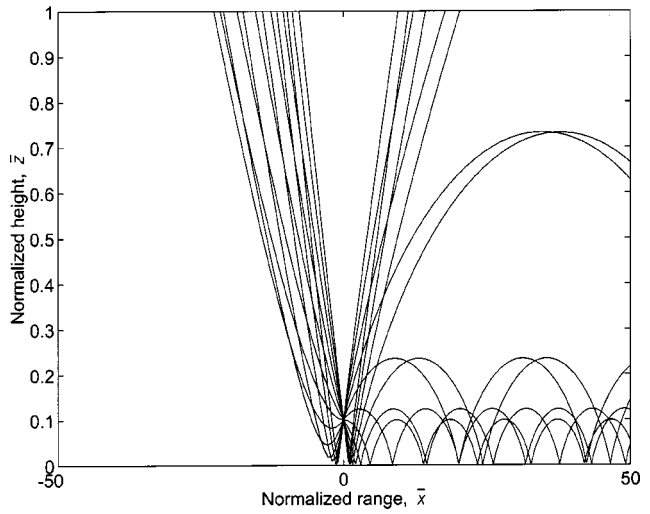


FIG. 6. Ray trace for $A=0$. This trace illustrates the refractive regime $-0.1\lesssim A\lesssim 0.1$, in which propagation is controlled primarily by wind shear. As in Fig. 5, upwind and downwind propagation are shown.

The figures also provide some insight as to the appropriate meteorological conditions for performing outdoor sound pressure level (SPL) measurements that are in accordance with the ANSI standard S12.18-1994.[13] The general measurement method (method 1) in the ANSI standard is designed mainly to avoid conditions of upward refraction, which would lower the observed SPL. Strong winds are also not allowed during measurements because of the wind noise and turbulent scattering accompanying this situation. The general method specifies that the wind speed shall not exceed 5 m/s as measured at 2 m height, and that during the daytime SPL measurements must be made within an angle of $\pm 45°$ from the downwind direction (corresponding to the 90° contour in Fig. 3). Based on Fig. 3, the condition $A\gtrsim -0.5$ ensures downward refraction near the ground within the $\pm 45°$ sector. A value $A\gtrsim 0$ (as is typically the case at nighttime) would enable measurements within a $\pm 90°$ sector (corresponding to the 180° contour in Fig. 3). The standard also permits measurements made at any direction relative to the wind if there is a well-developed ground-based temperature inversion. Figure 3 shows that the inversion will be sufficiently well developed to ensure downward refraction provided that $A\gtrsim 1$.

During daytime conditions, when A is typically negative,

the desire to make measurements when $|A|$ is small conflicts to some extent with the desire to make measurements when the wind speed is low. This is because u_* (which is in the denominator of A) increases with increasing wind speed. The meteorological conditions for which both A and the wind speed are in an appropriate range can be assessed from Fig. 1. In order to satisfy $A \gtrsim -0.5$ with a wind speed $\langle U(z=2\ \text{m})\rangle \leqslant 5$ m/s (the criterion for the general method), the heat flux must be less than about 150 W/m^2, which is often the case on cloudy days and/or when the solar elevation angle is low. But for smaller wind speeds (less than about 2 m/s), the heat flux must be less than 30 W/m^2, a value generally possible during the daytime only when a very thick, stratus cloud cover is present. Therefore, to avoid upward refraction during low-wind conditions, measurements must typically be made either at night or when there is a very thick cloud cover.

IV. CONCLUSION

The Monin–Obukhov similarity theory provides a ''universal'' characterization of the effective sound-speed gradient in the atmospheric surface layer. The predictions are universal in the sense that they should apply to propagation over reasonably flat ground with uniform roughness elements. A simple rewriting of the Monin–Obukhov result for the gradient leads to a refraction parametrization involving three nondimensional quantities that are dependent on the atmospheric conditions and propagation geometry. The first of these quantities (A) is the ratio of the sound-speed scale (c_*) to the friction velocity (u_*) and represents the relative strengths of the turbulent fluctuations in sound speed and wind speed. The second ($\overline{z}$) is the ratio of the actual height above the ground to a transitional height where wind gradients and the adiabatic lapse rate contribute roughly equally to the effective sound speed gradient. The third quantity is the cosine of the angle between the propagation direction and the wind direction. Using these parameters, three distinct propagation regimes can be identified: one ($A \lesssim -1$) where refraction is upward at all propagation angles relative to the mean wind, a second ($A \gtrsim 1$) where it is downward for all propagation angles, and a third ($|A| \lesssim 0.1$) where refraction is determined by the wind direction when $\overline{z} \lesssim 0.1$ and is upward for $\overline{z} \gtrsim 1$. Atmospheric conditions (as parametrized by the value of A) supporting a small value of the near-ground effective sound-speed gradient appear to be very rare, and in no situation does a small gradient occur in all propagation directions simultaneously. Constant effective sound-speed gradients (linear profiles) are also rare in any propagation direction. Downward refraction conditions rarely exist during the daytime, even in the downwind direction, except when the wind speed is greater than about 5 m/s and there is a cloud cover, or if the wind speed is greater than about 10 m/s in clear conditions. The contribution of temperature fluctuations to the sound-speed scale c_* (and therefore to the effective sound-speed gradient) is generally more important than that from humidity fluctuations, although the humidity contribution appears to be significant (around 30%) over wet surfaces.

ACKNOWLEDGMENTS

I thank G. A. Daigle (National Research Council Canada), V. E. Ostashev (NOAA Environmental Technologies Laboratory), D. W. Thomson (Pennsylvania State University), the reviewers, and the associate editor for their helpful comments that have led to improvement of this paper.

APPENDIX: NONDIMENSIONAL RAY TRACE EQUATIONS

In this section, the procedure used to create the nondimensional ray traces, Figs. 5–7, is described. The starting point is to define a nondimensional time, $\overline{t}=(g/c_0)t$, so Eq. (2) becomes

$$\frac{d\mathbf{n}}{d\overline{t}} = (n_1 n_3, n_2 n_3, n_3^2 - 1)\left(\frac{c_0}{g}\frac{\partial c_{\text{eff}}}{\partial z}\right). \tag{A1}$$

This result allows Eq. (18) to be used for the gradient. To perform the ray tracing, Eq. (1) must be supplemented with the following equation for the velocity of the ray:[6]

$$\mathbf{v}_{\text{ray}} = \frac{d\mathbf{x}}{dt} = c\mathbf{n} + \mathbf{u}, \tag{A2}$$

where $\mathbf{x}$ is the position of the ray. For the present purpose of developing qualitatively realistic ray traces, the right side may be approximated as $c_0\mathbf{n}$. Setting $\overline{\mathbf{x}}=\mathbf{x}/L_g$ for consistency with Eq. (17), we have

$$\frac{d\overline{\mathbf{x}}}{d\overline{t}} = \frac{kc_0}{u_*}\mathbf{n}. \tag{A3}$$

Therefore, to perform the nondimensional ray traces, u_*/c_0 must be specified in addition to A and $\overline{z}$. A value of $u_*/c_0 = 0.1/340$ was used for the ray traces in this paper.

A remaining practical issue is that Eq. (18) predicts infinite gradients at the ground. The Monin–Obukhov similarity functions, Eq. (9), are invalid in the vicinity of the surface roughness elements. Therefore, for Figs. 5–7, $\partial c_{\text{eff}}/\partial z$ was set to zero below the normalized roughness length, $\overline{z_0} = z_0/L_g$. The value for $\overline{z_0}$ was 0.001. With values set for $\overline{z_0}$ and u_*/c_0, it is then straightforward to integrate Eqs. (A1) and (A3) in time by standard numerical techniques.

[1] A. S. Monin and A. M. Obukhov, ''Basic laws of turbulent mixing in the ground layer of the atmosphere,'' Trans. Geophys. Inst. Akad. Nauk USSR **151**, 163–187 (1954).

[2] R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer, Dordrecht, 1988).

[3] J. R. Garratt, *The Atmospheric Boundary Layer* (Cambridge University Press, Cambridge, 1992).

[4] H. Klug, ''Sound speed profiles determined from outdoor sound propagation measurements,'' J. Acoust. Soc. Am. **90**, 475–481 (1991).

[5] A. L'Espérance, J. Nicolas, D. K. Wilson, D. W. Thomson, Y. Gabillet, and G. Daigle, ''Sound propagation in the atmospheric surface layer: Comparison of experiment with FFP predictions,'' Appl. Acoust. **40**, 325–346 (1993).

[6] A. D. Pierce, ''Wave equation for sound in fluids with unsteady inhomogeneous flow,'' J. Acoust. Soc. Am. **87**, 2292–2299 (1990).

[7] The reader is referred to Ostashev (Ref. 8) for more detailed discussions of the effective sound speed concept and its limitations.

[8] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN Spon, London, 1997).

[9] U. Högström, ''Review of some basic characteristics of the atmospheric surface layer,'' Boundary-Layer Meteorol. **78**, 215–246 (1996).
[10] D. K. Wilson, ''An alternative function for the wind and temperature gradients in unstable surface layers,'' Boundary-Layer Meteorol. **99**, 151–158 (2001).
[11] M. L. Wesely, ''The combined effect of temperature and humidity fluctuations on refractive index,'' J. Appl. Meteorol. **15**, 43–49 (1976).
[12] J. M. Wallace and P. V. Hobbs, *Atmospheric Science: An Introductory Survey* (Academic, New York, 1977).
[13] ANSI S12.18-1994, ''American National Standard Procedures for Outdoor Measurement of Sound Pressure Level'' (Acoustical Society of America, New York, 1994).

Complex reflection phase gradient as an inversion parameter for the prediction of shallow water propagation and the characterization of sea-bottoms

Phillip Joseph
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, England

(Received 9 April 2001; revised 18 February 2002; accepted 28 October 2002)

In this paper a quantity is proposed, referred to as the complex reflection phase gradient, whose use in a matched field inversion procedure allows for the rapid extraction of first order geo-acoustic information about the sea-bottom. It is based on the observation that at low grazing angles the reflection phase and bottom loss for a wide range of sea-bottom types commonly exhibits an approximate linear relationship to the vertical component of the acoustic wave number at the seabed. The real part of this quantity specifies the rate at which the reflection phase varies with vertical acoustic wave number while the imaginary part quantifies the rate of change of bottom loss. Despite being defined with just two real parameters it is shown that it provides an accurate prediction of the sound field for a wide range of bottom types. In addition, its measurement permits an estimate to be made for the input impedance to the seabed in the zero grazing angle limit and, in the case of a homogeneous elastic half-space of known density, the compressional and shear wave speed. The main advantage of the two-parameter seabottom representation is that each parameter is readily inverted from comparatively few acoustic pressure measurements. The usefulness of the technique is illustrated by the results from computer simulated acoustic pressure measurements made at just eleven sensors in a simple shallow water channel, and results from a 10 cm deep laboratory channel at frequencies between 10 kHz and 75 kHz.
[DOI: 10.1121/1.1532003]

PACS numbers: 43.30.Ma, 43.30.Pc, 43.30.Bp [DLB]

I. INTRODUCTION

The measurement of the geo-acoustic seabottom parameters necessary for making accurate predictions of long-range acoustic propagation in a shallow water waveguide is generally very difficult. The main difficulty arises from the repeated interaction of sound between the sea surface and sea-bottom. The influence of the latter is therefore prevented from easily being isolated, and hence accurately quantified. Inverse techniques potentially overcome this difficulty. However, a complete specification of the sea-bottom involves a large number of parameters. Their inversion therefore requires extensive and accurate acoustic pressure measurements as well as considerable computational effort. In many cases the inversion is ill-conditioned and the results susceptible to significant error. Even the simplest homogeneous half-space sea-bottom model requires five real parameters to fully describe it acoustically (mass density and compressional and shear complex wave speeds). Each of these parameters is difficult to measure *in situ*, and yet error in any one will introduce some degree of error into the acoustic field prediction. The structure of the sea-bottom is generally more complicated than this simple idealization and so the number of parameters required to characterize it, and the difficulties of making their direct measurement, are correspondingly greater.

In this paper a simple representation of the sea-bottom reflectivity is proposed as the basis for a pragmatic inversion procedure. The sea-bottom model is defined by just two real parameters whose inversion allows for the rapid collection of geo-acoustic data from which acoustic propagation may be predicted. Furthermore, since only two parameters are required, the inversion may be performed accurately with comparatively few acoustic pressure measurements, for example, by a typical sonar receiver array. This paper will also show that under certain conditions the two parameters of the model also allows for:

(i) the estimation of the low-grazing input impedance to the seabed;
(ii) an estimate to be made for the compressional and shear wave speed in a homogeneous sea-bottom of known density, and the shear wave speed of a solid sediment layer overlaying a solid substrate.

The basis for the sea-bottom model is the recognition that long-range propagation is due to interaction of sound with the seabed at small grazing angles. At these angles the reflection-phase and bottom loss (in dB) is approximately linearly proportional to the grazing angle. This property is observed in the reflectivity of a wide range of idealized and realistic sea-bottom types. For example, in semi-infinite homogenous fluid and solid sea-bottoms,[1,2] in the presence of a solid sediment layer overlaying a solid substrate,[3] and in the more general case where there the compressional and shear wave speeds vary continually with depth.[4]

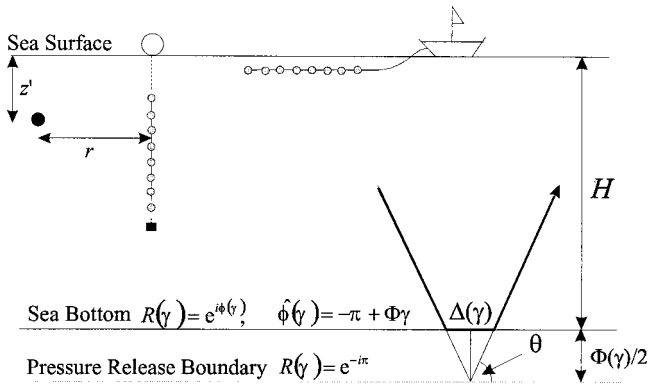


FIG. 1. Approximate representation of an arbitrary impedance boundary by a plane wave reflection coefficient with linear complex phase of gradient Φ. Either a horizontal or vertical sensor array is used to measure the sound field for its inversion.

II. COMPLEX REFLECTION PHASE GRADIENT

A. Basic principles

A horizontally stratified shallow water waveguide of depth H overlaying a plane boundary with complex plane wave reflection coefficient R is shown in Fig. 1. Expressing the complex plane wave reflection coefficient in the exponential form $R(\gamma)=e^{i\phi(\gamma)}$ defines the complex phase as $\phi(\gamma)=-i\ln R(\gamma)$, where γ is the vertical component of the acoustical wave number at the seabed. A series expansion of $\phi(\gamma)$ about $\gamma=0$, to first order, is of the form

$$\phi(\gamma)=\phi(0)+\gamma\left.\frac{\partial\phi(\gamma)}{\partial\gamma}\right|_{\gamma=0}+\cdots. \tag{1}$$

This expansion is useful as a geo-acoustic representation of the sea-bottom for two reasons. One is that the first term takes the constant value $\phi(0)=-\pi$; the phase change occurring at any angle upon reflection at a plane pressure release boundary. Second, is that terms of order two and higher can usually be neglected in predictions of long-range shallow water propagation. This linear approximation is central to the inversion principle proposed here since it significantly reduces the number of independent parameters that need to be determined. It is justified because (i), the complex reflection phase usually varies linearly with γ over a large wave number range and (ii), the low order, least attenuated dominant modes, interact at the seabed with small grazing angle, and hence small γ.

For the purpose of predicting long-range sound propagation, therefore, the expansion of Eq. (1) may be truncated, retaining only the first order term. From hereon this process will be referred to as the linear (complex) phase approximation. Defining the (complex) coefficient in γ by $\Phi(\gamma)=\partial\phi(\gamma)/\partial\gamma$, recalling that $\phi(\gamma)=-i\ln R(\gamma)$, and using the asymptotic property $R(\gamma)_{\lim\gamma\to0}\to-1$ gives

$$\Phi(0)=\left.\frac{\partial\phi(\gamma)}{\partial\gamma}\right|_{\gamma\to0}=i\left.\frac{\partial R(\gamma)}{\partial\gamma}\right|_{\gamma\to0}. \tag{2}$$

A linear complex phase approximation $\hat{\phi}$ of the seabed at sufficiently small γ is therefore of the form

$$\hat{\phi}(\gamma)=-\pi+\Phi(0)\gamma. \tag{3}$$

B. Relationship between complex reflection phase gradient and complex effective depth

As illustrated in Fig. 1, the coefficient $\Phi(0)$ in Eq. (3) has an interpretation as twice the (complex) depth of a fictitious pressure release boundary located below the physical seabed.[3] The linear (complex) phase approximation is therefore closely related to the complex effective depth approximation proposed by Zhang and Tindle[5] for the purpose of approximating the acoustic field in iso-speed channels. The definition of complex reflection phase gradient, however, relates entirely to the reflectivity of the seabottom and is therefore valid for arbitrary sound speed profile. It is therefore more general than the effective depth representation.

Expressing the linear phase approximation to the reflection coefficient as $\hat{R}(\gamma)=|\hat{R}(\gamma)|e^{i\,\mathrm{Re}\,\hat{\phi}(\gamma)}$, the complex phase may be written as $\hat{\phi}(\gamma)=\mathrm{Re}\,\hat{\phi}(\gamma)-i\ln|\hat{R}(\gamma)|$. Comparison with Eq. (3) yields

$$\mathrm{Re}\,\hat{\phi}(\gamma)=-\pi+\Phi_{\mathbf{R}}(0)\gamma, \tag{4a}$$

$$|\hat{R}(\gamma)|=\exp[-\Phi_{\mathbf{I}}(0)\gamma], \tag{4b}$$

where $\Phi_{\mathbf{R}}(\gamma)=\mathrm{Re}\{\Phi(\gamma)\}$ and $\Phi_{\mathbf{I}}(\gamma)=\mathrm{Im}\{\Phi(\gamma)\}$. Equation (4a) is the linear phase assumption implicit in the conventional effective depth approximation, while Eq. (4b) predicts that the magnitude of the reflection coefficient approximates locally at zero grazing angle to an exponentially decaying function of γ. The expression for the bottom loss, $BL(\gamma)=-20\log_{10}|R(\gamma)|$ (dB), applied to Eq. (4b) therefore becomes

$$\widehat{BL}(\gamma)=8.7\Phi_{\mathbf{I}}(0)\gamma\ \ (\mathrm{dB}). \tag{5}$$

III. PROCESSORS FOR THE INVERSION OF THE COMPLEX REFLECTION PHASE GRADIENT

The Bartlett and Maximum Likelihood Matched Field processors[6] are now compared for their ability to invert, with least ambiguity, the parameter Φ from just 11 complex acoustic pressure measurements made in a shallow water iso-speed waveguide. Eleven sensors were chosen to illustrate the utility of the technique with a typical, readily available small array. The use of more than 11 sensors in the examples that follow was not found to produce significant improvements in inversion accuracy, whereas the use of fewer sensors was found to cause significant deterioration in accuracy. No meaningful results were obtained with the use of either processor when five sensors or less were used. Optimal values of $\hat{\Phi}_0$ are sought that maximizes the match between the measured acoustic pressures at N measurement locations expressed by the column vector $\mathbf{p}=[p(r_1,z_1)\times p(r_2,z_2)\cdots p(r_N,z_N)]^T$, and a model of the acoustic pressures represented by the vector $\hat{\mathbf{p}}(\hat{\Phi})=[\hat{p}(r_1,z_1,\hat{\Phi})\times\hat{p}(r_2,z_2,\hat{\Phi})\cdots\hat{p}(r_N,z_N,\hat{\Phi})]^T$. The replica, or trial solution $\hat{\mathbf{p}}$, is obtained as a solution to the wave equation with a seabed reflection coefficient approximated by $\hat{\phi}(\gamma)=-\pi+\hat{\Phi}_{\mathbf{R}}\gamma+i\hat{\Phi}_{\mathbf{I}}\gamma$. Thus, only two parameters are sought and these may be found by exhaustive search over the range of likely values, $0\leqslant\hat{\Phi}_{\mathbf{R}}/\lambda\leqslant2$, $0\leqslant\hat{\Phi}_{\mathbf{I}}/\lambda\leqslant0.5$.

TABLE I. Summary of seabottom parameters used in the Chapman *et al.* seabed investigation (Ref. 2). Note that critical angles are based on a sound speed in water of 1484 m/s.

Seabed type	Density (kgm^{-3})	Compressional wave speed (ms^{-1})	Shear wave speed (ms^{-1})	Critical angle (degs)
A	1600	1550	125	16.8
B	1700	1700	200	29.2
C	1800	1850	300	36.7
D	1900	2000	450	42.1
E	2000	2150	650	46.4
F	2100	2300	850	49.8

The output of the Bartlett, linear processor $P_B(\hat{\Phi})$ is obtained by weighting the pressure measurements made by the receiver array by a normalized trial solution $\mathbf{w}(\hat{\Phi}) = \hat{\mathbf{p}}(\hat{\Phi})/\|\hat{\mathbf{p}}(\hat{\Phi})\|$, where $\|\cdot\|$ is L_2 norm, summing and squaring. The result is the quadratic form,

$$P_B(\hat{\Phi}) = \mathbf{w}^+(\hat{\Phi})\mathbf{K}\mathbf{w}(\hat{\Phi}), \tag{6}$$

where $\mathbf{K}$ is the measured covariance matrix $\mathbf{K} = E\{\mathbf{p}\mathbf{p}^+\}$ with elements $K_{ij} = E\{p_i^* p_j\}$, the superscript "+" denotes the Hermitian transpose operator and $E\{\}$ averaging of the stochastic time varying acoustic pressure fluctuations in a narrow frequency band. In general, $\mathbf{K}$ will comprise contributions from both signal and noise, which may "overlap" as they often have the same correlation structure across the array since both propagate through the same ocean environment. This processor is the most widely used, and is most robust to error in the replica field and the presence of measurement noise. It has the disadvantage that it suffers from poor sidelobe characteristics and is therefore susceptible to spurious outputs and generally poor signal to noise rejection properties.

The other most commonly used processor is the Maximum Likelihood (ML) beamformer. This processor adaptively adjusts its weighting vector to minimize its mean square output whilst constrained to pass a signal from the "look" direction $\mathbf{w}(\hat{\Phi})$ with unity gain. It offers superior resolution to the linear processor at the expense of a greater sensitivity to mismatch in the replica field and the presence of measurement noise. The ML beamformer output is calculated from[6]

$$S_{\mathrm{ML}}(\hat{\Phi}) = (\mathbf{w}^+(\hat{\Phi})\mathbf{K}^{-1}\mathbf{w}(\hat{\Phi}))^{-1}. \tag{7}$$

IV. APPLICATION TO A HOMOGENOUS, ELASTIC HALF-SPACE

A. Propagation models and waveguide properties

Many of the features and advantages of the proposed inversion scheme can be illustrated by the results of its application to an iso-speed water column overlaying a semi-infinite homogeneous, elastic half-space. For consistency with previous work on the effective depth approximation, the Bartlett and maximum likelihood processors are used to invert for the complex reflection phase gradient of five semi-infinite, homogenous, elastic bottoms with real wave speeds and densities (i.e., no *P* or *S*-wave attenuation) suggested by Chapman *et al.*[2] These are listed in Table I.

For each of the sea-bottoms *A* to *F* the "measured" acoustic pressure $p(r,z)$ was computed at 100 Hz in a 50 m deep waveguide using a Fast Field propagation model. A high precision adaptive integration routine was used to evaluate the Hankel Transform of g, $p(r,z) = \int_0^K g(z,z',\gamma) J_0(k_r r) k_r \, dk_r$, where $g(z,z',\gamma)$ is the analytic Green function for an iso-speed waveguide overlaying an impedance boundary.[7] The upper wave number limit of integration K in this Fast Field model was chosen such that $K \gg \omega/c$ to ensure solution convergence, and to ensure that the acoustic field is inclusive of the nonmodal near field.

For reasons of computational expediency the replica "trial" field was calculated from the summation over normal modes

$$w(r,z,\Phi) = N_n^{-1} \sum_{n=1}^{N} \hat{P}_n \sin(\mathrm{Re}\,\hat{\gamma}_n z) \sin(\mathrm{Re}\,\hat{\gamma}_n z') H_0^{(1)}(\hat{k}_n r), \tag{8}$$

where N_n is the normalization factor introduced such that $(L_r^{-1} \int_{r_0}^{r_0+L_r} |w|^2 r \, dr)^{1/2} = 1$, $H_0^{(1)}$ denotes the zeroth order Hankel function of the first kind, r_0 is the horizontal distance of the sensor closest to the source, and L_r is the length of the horizontal array. The terms $\hat{P}_n$ and $\hat{\gamma}_n$ are the modal amplitudes and eigenvalues, which in the iso-speed channel assumed here, were approximated from $\hat{\Phi}$ by defining the effective depth $\tilde{H} = H + \frac{1}{2}\hat{\Phi}$, and using the following results from effective depth theory[5]

$$\hat{P}_n(\hat{\Phi}) = 2\tilde{H}^{-1}, \tag{9a}$$

$$\hat{\gamma}_n = n\pi/\tilde{H}, \tag{9b}$$

and

$$\hat{k}_n = \sqrt{k^2 - \hat{\gamma}_n^2}, \tag{9c}$$

where $k = \omega/c$. By comparison with the number of propagating modes in a waveguide with upper and lower pressure release boundaries, the number of modes N included in the modal sum was calculated from $N = \mathrm{int}[2\tilde{H}/\lambda]$.

B. Horizontal receiver array example

The first example investigated here of Φ-inversion uses eleven pressure measurements made at a depth of 15 m in a 50 m deep waveguide, uniformly located in range between 2 km and 2.2 km due to a 10 m deep source transmitting at 100 Hz. The Bartlett and Maximum Likelihood processor outputs versus $\hat{\Phi}_{\mathbf{R}}$ and $\hat{\Phi}_{\mathbf{I}}$ for sea-bottom *A* are shown in Figs. 2(a) and (b).

The linear processor output in Fig. 2(a) exhibits significant ambiguity and hence poor localization caused by a broad main "beam" and poor sidelobe rejection owing to the few number of sensors used in this example. Nevertheless, a distinct optimum value is observed at $\hat{\Phi}_0/\lambda = (1.49 \pm 0.10) + (0.20 \pm 0.010)i$, where the uncertainty bounds are calculated from the values at which the correlation falls below 95%. Significantly less ambiguity is observed for the output of the maximum likelihood processor in Fig. 2(b), for which

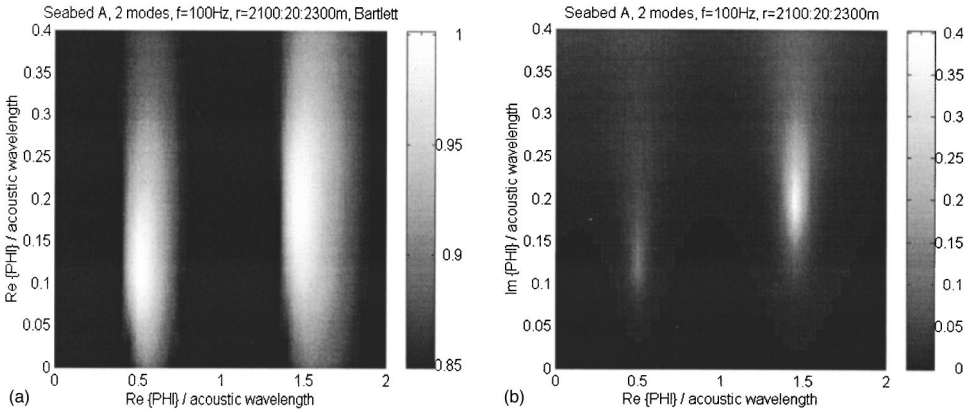


FIG. 2. (a) Bartlett processor output versus {Re Φ,Im Φ}/λ for seabed A. (b) Maximum Likelihood processor output versus {Re Φ,Im Φ}/λ for seabed A.

$\hat{\Phi}_0/\lambda = (1.49 \pm 0.01) + (0.21 \pm 0.002)i$, with the uncertainty bounds now being deduced, arbitrarily, from the size of the ''bright spot.'' These inverted values represent a significant underestimate by 19% of the real part of the theoretical value $\Phi(0)/\lambda = 1.85 + 0.004i$ calculated from Eq. (16) below, whereas its imaginary part is over-predicted by an order of magnitude.

Part of the reason for this discrepancy in this low-loss example is because one of the two dominant modes present is not a proper mode, but a highly attenuated leaky whose angle of interaction with the seabottom exceeds the low 16.8° critical grazing angle of the seabed. The conventional (complex) effective depth approximation is therefore unable to predict this behavior. Another reason for this error is due to departures of the complex phase variation from the assumed functional form of Eqs. (4a), (4b).

The processor outputs for the higher wave speed example of sea-bottom *C* are plotted in Figs. 3(a) and (b), respectively. The linear processor output in this example, as for seabed type *A*, while less ambiguous also exhibits poor resolution, particularly for Φ_1. By contrast, the ML processor output localizes both real and imaginary parts extremely well to yield $\hat{\Phi}_0/\lambda = (0.75 \pm 0.01) + (0.023 \pm 0.005)i$. The real part of this inverted value represents a 9% underpredication of the real part of the theoretical value of $\Phi(0)/\lambda = 0.82 + 0.067i$ calculated from Eq. (16) below, while its imaginary part represents an underestimate by a factor of 3. In any case, since the bottom loss in this example is small (due to low shear wave speed), this difference in the imaginary part is shown below to have only a small effect on the transmission loss.

The processor outputs for the highest loss seabottom, example *F*, are presented in Figs. 4(a) and (b). The performance of both the linear and maximum likelihood processor outputs is poorer in this high-loss example than in previous examples. This is because the simple Pekeris mode shape functions assumed for the replica field in this high loss seabottom example may not be an accurate representation of the exact field, and also because the modes are highly attenuated such that only the lowest order, least attenuated, mode is of significant amplitude at the receiver. Nevertheless, the ML output provides a sufficiently unambiguous inversion result of $\hat{\Phi}_0/\lambda = [(0.11 \pm 0.01) + (0.44 \pm 0.03)i]$. The corresponding theoretical value for this example is $\Phi(0)/\lambda = 0.10 + 0.41i$, representing an approximate 10% error in both the real and imaginary parts of Φ. Thus, an error of about 10% appears to be inherent in the inversion scheme. It arises primarily because of small departures in the exact variation of complex reflection phase with grazing angle from the assumed form of Eqs. (4a), (4b). Nevertheless, it is demonstrated below that $\hat{\Phi}_0$ used in propagation model gives excellent predictions of transmission loss compared with exact calculations (Fig. 6), and with experimental measurements (Fig. 11).

Note that generally poorer processor performance is observed with pressure measurements made at much closer, and at much longer range, from the source. At receiver locations less than a few tens of wavelengths, the near field makes a significant contribution to the total acoustic field. The near field interacts at the seabed at high grazing angles where the linear phase assumption is likely to be invalid. The field at close range is therefore poorly correlated with the modal

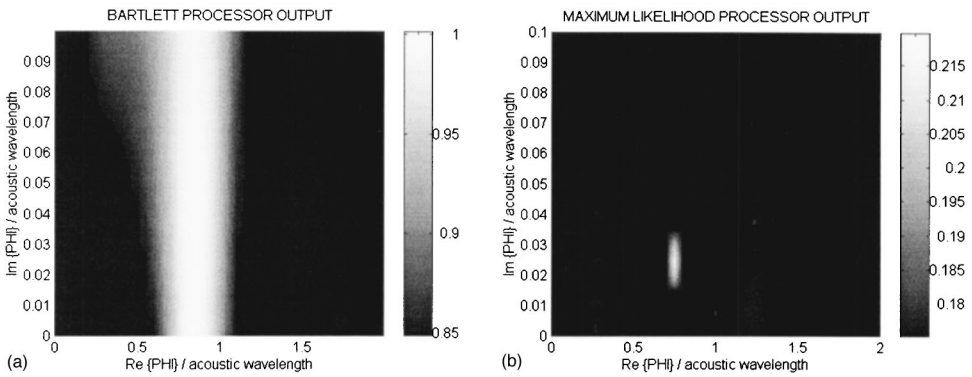


FIG. 3. (a) Bartlett processor output versus {Re Φ,Im Φ}/λ for seabed C. (b) Maximum Likelihood processor output versus {Re Φ,Im Φ}/λ for seabed C.

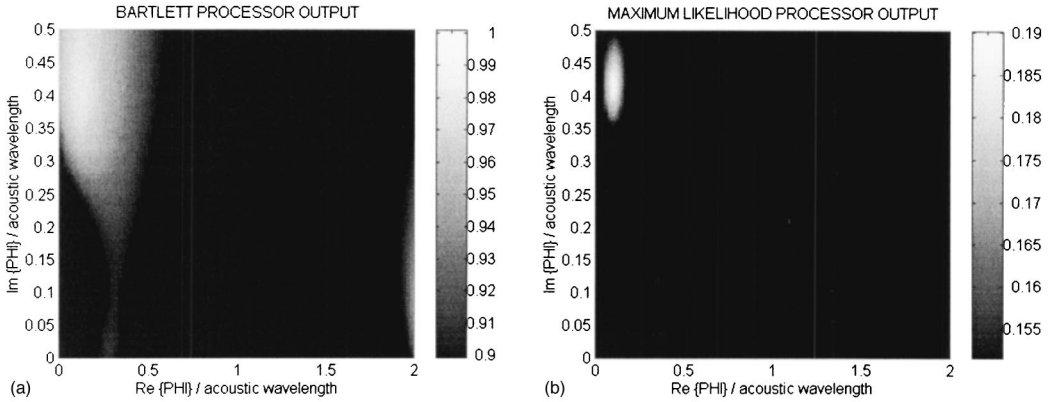


FIG. 4. (a) Bartlett processor output versus {Re Φ,Im Φ}/λ for seabed F. (b) Maximum Likelihood processor output versus {Re Φ,Im Φ}/λ for seabed F.

field computed from the linear phase assumption. Poor processor performance is also observed at very long range where highly attenuated, high order, modes may decay to insignificant levels at the receiver leading to loss of information, and hence poor localization performance. Processor performance generally improves with increasing frequency as the number of propagating modes increases.

C. Vertical receiver array

Figure 5 depicts the ML output versus $\Phi_{\mathbf{I}}$ and $\Phi_{\mathbf{R}}$ for seabed *C* applied to acoustic pressure data obtained by a VLA with 11 sensors uniformly separated between the sea surface and sea-bottom. The array is at 2 km range from the source of 100 Hz frequency. The value $\hat{\Phi}_0$ of maximum processor output is in good agreement with Fig. 4(b) obtained with the horizontal sensor array. The main difference is the significantly poorer localization of $\Phi_{\mathbf{I}}$ and the appearance of spurious outputs well away from the theoretical value.

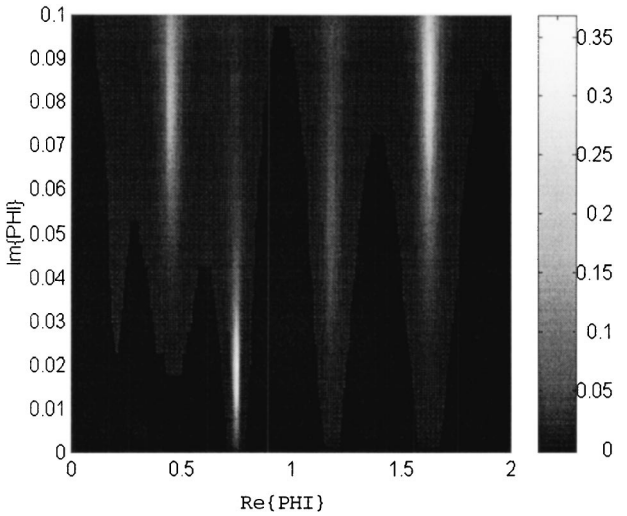


FIG. 5. Maximum Likelihood processor output versus {Re Φ,Im Φ}/λ for seabed C obtained with a VLA at 11 depth positions every 5 m from the surface to the bottom, 2 km in range from the source.

D. Comparison of exact transmission loss with that predicted using measured complex phase gradient

The optimum $\hat{\Phi}_0$ values obtained in Sec. B from just 11 pressure measurements were substituted into the expression for $p(r,z)$ and a calculation made of the transmission loss between 0 and 5 km at 15 m depth. A comparison of these results with the transmission loss calculated using the exact reflection coefficient is presented in Fig. 6 for seabeds *A* to *F*. Good agreement is obtained between the transmission loss predictions in all cases, including those for which the inverted and theoretical Φ values were in poor agreement. Nonphysical features present in the simulated data at 2.4 and 4.8 km are due to occasional instabilities in our Fast Field program, caused by the adaptive integration routine.

V. A PHYSICAL INTERPRETATION OF THE COMPLEX PHASE GRADIENT AND ITS RELATIONSHIP TO THE PARAMETERS OF A SEMI-INFINITE ELASTIC HALF-SPACE

A. Relationship to the seabed input impedance

An approximate relationship is now derived between $\Phi(\gamma)$ and the normalized seabed input impedance $Z(\gamma)$ that

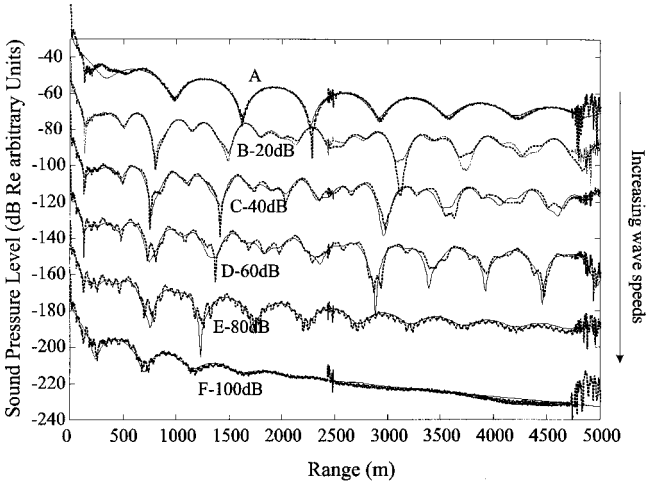


FIG. 6. Comparison between exact theoretical Transmission Loss (dashed curve) and that predicted using optimal Φ values in the linear complex phase model (solid curve) for the seabottom types A–F.

is valid at low grazing angles, and which becomes exact in the zero grazing angle limit. The acoustic pressure p and the vertical component of particle velocity u_z at the seabed, due to a plane wave reflected at an impedance boundary with change of (complex) phase given by $-\pi+\Phi\gamma$, may be expressed as the sum of incident and reflected waves,

$$p=p_0(e^{-i\gamma z}+e^{i(-\pi+\Phi z)}e^{i\gamma z}), \tag{10a}$$

$$u_z=\frac{p_0\gamma}{\rho ck}(e^{-i\gamma z}-e^{-i\pi}e^{i\gamma(z+\Phi)}). \tag{10b}$$

If $\Phi(\gamma)$ is chosen to produce identical normalized input impedance $Z(\gamma)=p/\rho cu_z$ to that produced by the physical boundary, the following relation is derived,

$$\frac{k}{\gamma}\frac{e^{-i\gamma z}-e^{i(\gamma z+\Phi)}}{e^{-i\gamma z}+e^{i(\gamma z+\Phi)}}=Z(\gamma). \tag{11}$$

Evaluating this expression at the seabed $z=0$ and solving for $\Phi(\gamma)$ yields

$$\Phi(\gamma)=\frac{-i}{\gamma}\ln\frac{1-(\gamma/k)Z(\gamma)}{1+(\gamma/k)Z(\gamma)}, \tag{12}$$

where $\gamma/k=\sin\theta$. For sufficiently small grazing angles satisfying $\gamma/k\ll Z^{-1}(\gamma)$, the logarithm term is closely approximated by the leading term in its series expansion of the form

$$\ln\frac{1-(\gamma/k)Z(\gamma)}{1+(\gamma/k)Z(\gamma)}\approx-2(\gamma/k)Z(\gamma),\quad \gamma/k\ll Z^{-1}(\gamma). \tag{13}$$

Equation (13) substituted into Eq. (12) indicates that at small grazing angles, the quantities $\Phi(\gamma)$ and $Z(\gamma)$ are approximately proportional. Unlike the reflection coefficient they are slowly varying with γ. Separating the result into its real and imaginary parts yields $\mathrm{Re}\{Z(\gamma)\}\approx(k/2)\Phi_{\mathbf{I}}(\gamma)$ and $\mathrm{Im}\{Z(\gamma)\}\approx-(k/2)\Phi_{\mathbf{R}}(\gamma)$. At the grazing angles of validity, $\Phi_{\mathbf{R}}$ is determined by the input reactance $\mathrm{Im}\{Z(\gamma)\}$, while $\Phi_{\mathbf{I}}$ is determined by the input resistance $\mathrm{Re}\{Z(\gamma)\}$ to the seabed. The latter describes the reflection loss due to the various dissipative processes present at the ocean bottom. In the limit $\gamma\to0$,

$$2ik^{-1}\lim_{\gamma\to0}Z(\gamma)=\Phi(0). \tag{14}$$

The reflection coefficient $R(\gamma)$, on the other hand, converges exactly to -1 as $\gamma\to0$ and is therefore an unsuitable parameter for characterizing the low grazing angle sea-bottom behavior. The use of Eq. (14) provides a much simpler method of evaluating $\Phi(0)$ through the evaluation of the input impedance to the seabed.

B. Use of Φ for determining the wave speeds in a homogenous elastic half-space

The seabottom often closely approximates to a homogenous elastic half-space. The input impedance to this idealised bottom is given by[8] $Z(\theta)=Z_p\sin^2 2\theta_s+Z_s\cos^2 2\theta_s$, where $Z_p=\rho c_p/\sin\theta_p$, $Z_s=\rho_b c_s/\sin\theta_s$, ρ_b, is the density and c_p and c_s are the compressional and shear wave speed, respectively. The angles θ_p and θ_s are the grazing angles of transmission into the solid half-space and are related to the grazing angle θ in water by $c/\cos\theta=c_p/\cos\theta_p=c_s/\cos\theta_s$. Evaluating the input impedance Z defined above in the zero grazing angle limit, $\gamma\to0$, $\theta\to0$, making use of Snell's law above yields

$$\lim_{\gamma\to0}Z(\gamma)=n_p m^{-1}(2n_s^2-1)^2(1-n_p^2)^{-1/2}+4n_s^3m^{-1}(1-n_s^2)^{1/2}, \tag{15}$$

where $n_p=c_p/c$, $n_s=c_s/c$, and $m=\rho/\rho_b$. The real and imaginary parts of the complex phase gradient in the zero grazing angle limit for an idealized homogenous solid half-space where $n_s<1$ for both fast bottoms $n_p>1$, and slow bottoms $n_p<1$, may now be summarized thus:

$n_p>1$; Fast Bottoms

$$\Phi_{\mathbf{R}}(0)=\frac{2n_p(1-2n_s^2)^2}{km(n_p^2-1)^{1/2}}, \tag{16a}$$

$$\Phi_{\mathbf{I}}(0)=\frac{8n_s^3(1-n_s^2)^{1/2}}{km}; \tag{16b}$$

$n_p\leqslant1$; Slow Bottoms

$$\Phi_{\mathbf{R}}(0)=0, \tag{17a}$$

$$\Phi_{\mathbf{I}}(0)=\frac{2n_p(2n_s^2-1)^2}{km(1-n_p^2)^{1/2}}+\frac{8n_s^3(1-n_s^2)^{1/2}}{km}. \tag{17b}$$

Equation (16a) is twice the real part of the extra depth derived previously by Chapman *et al.*, following a different approach.[2] Equation (17) confirms that upon reflection at a slow sea-bottom there is no phase change but significant reflection loss.

C. Low grazing angle behavior of the Bottom Loss for a homogenous elastic half-space

The theoretical expression of Eq. (16b) for $\Phi(0)$ due to an elastic half-space will now be used in conjunction with $\widehat{BL}(\gamma)\approx8.7\gamma\Phi_{\mathbf{I}}(0)$ of Eq. (5) to obtain a simple expression for the bottom loss as a function of grazing angle, density and the shear wave speed. A series expansion of Eq. (16b) in the shear wave speed ratio n_s is of the form,

$$km\Phi_{\mathbf{I}}(0)\approx8n_s^3-4n_s^5+\ldots,\quad n_s<0.5. \tag{18}$$

To leading term, an approximation to the bottom loss may therefore be written as

$$\widehat{BL}(\theta)=70m^{-1}n_s^3\sin\theta(\mathrm{dB}),\quad n_s<0.5,\ \theta<20^\circ. \tag{19}$$

D. Use of Φ_I and Φ_R for estimating effective shear and compressional wave speeds in a sea-bottom of known density

This section will demonstrate the use of the inversion measurements of $\Phi_{\mathbf{R}}$ and $\Phi_{\mathbf{I}}$, together with the results of Eqs. (16), for obtaining estimates of the effective shear and compressional wave speed in a fast seabottom of known density. Here, it is assumed that the sea-bottom approximates to a homogenous elastic half-space and that excitation of shear waves in the sea-bottom is the dominant loss mechanism. The inversion results of complex Φ/λ for seabottoms A to F are listed in the second and third columns of Table II. Error

TABLE II. Values of ''measured'' complex phase gradient and a comparison between actual and inferred shear and compressional wave speeds.

Seabed	$m(=\rho_w/\rho)$	''Measured'' $\Phi_{\mathbf{R}}/\lambda$	''Measured'' $\Phi_{\mathbf{I}}/\lambda$	Actual c_s (m/s)	Estimated c_s (m/s)	Actual c_p (m/s)	Estimated c_p (m/s)
A	0.62	1.49±0.01	0.21±0.002	125	694±33[a]	1550	1575±2
B	0.59	0.92±0.01	0.015±0.004	200	280±27	1700	1770±8
C	0.55	0.75±0.01	0.023±0.005	300	325±23	1850	1940±20
D	0.53	0.63±0.01	0.047±0.008	450	399±23	2000	1921±21
E	0.50	0.36±0.01	0.190±0.010	650	625±11	2150	2020±46
F	0.48	0.11±0.01	0.440±0.030	850	815±19	2300	2116±218

[a]Poor agreement in this low wave speed example is due to the presence of a leaky mode, which is not predicted by the linear phase model.

bounds are included, deduced subjectively from the size of the ''bright spot'' obtained using the maximum likelihood processor. These values assume $\lambda=14.84$ m corresponding to 100 Hz and a sound speed of 1484 ms^{-1}, and the density ratio $m=\rho/\rho_b$ listed in the second column.

The shear wave speed for fast bottoms (identified by a nonzero $\Phi_{\mathbf{R}}$ measurement) may be exactly determined from the solution $\Phi_{\mathbf{I}}$ of Eq. (16b). Expanding this equation yields, $64n_s^8-64n_s^6+(km\Phi_{\mathbf{I}}(0))^2=0$. In most shallow water sediments,[8] however, $c_s/c<0.5$, so that the first term may be neglected to yield an approximate solution for the shear wave speed as

$$c_s/c\approx\tfrac{1}{2}(km\Phi_{\mathbf{I}}(0))^{1/3},\quad c_s/c<0.5. \tag{20}$$

Thus, by assuming that $\hat{\Phi}_0=\Phi(0)$, the use of Eq. (20) provides an estimate for the shear wave speed. A comparison of the exact shear wave speed with the approximation of Eq. (20) applied to the inversion ''measurements'' in column 4 are compared in columns 5 and 6, respectively. Acceptable agreement is observed, except for seabed *A* where the estimated shear wave speed is five times greater than the correct value. The finding can be explained by the presence of a dominant virtual mode in this example due to low compressional wave speed and hence the low critical grazing angle of 16.8°. Equation (20) assumes energy loss through shear wave excitation, whereas in this example, energy is lost to compressional waves through the interaction of leaky modes with the seabed at grazing angles exceeding the critical grazing angle. This method for the estimation of the shear wave speed is therefore inappropriate in this low wave speed example since it is based upon incorrect assumptions about the loss mechanism. The value obtained is therefore an ''effective'' shear wave speed.

With the shear wave speed determined from Eq. (20), the compressional wave speed ratio follows exactly from Eq. (16a) as,

$$\frac{c_p}{c}=\frac{km\Phi_{\mathbf{R}}(0)}{\sqrt{(km\Phi_{\mathbf{R}}(0))^2-4(1-2n_s^2)^4}}. \tag{21}$$

The exact compressional wave speed and the value estimated from Eq. (21) using the estimated shear wave speed in column 6 of Table II are compared in columns 7 and 8. Agreement is generally good. Note that a high $\Phi_{\mathbf{R}}$ appears to be characteristic of low compressional wave speeds (for a fixed density), while a high $\Phi_{\mathbf{I}}$ value is characteristic of high shear wave speed seabottoms. A single complex value of Φ, by measurement or otherwise, may therefore be used for seabottom classification.

VI. A LABORATORY EXAMPLE OF Φ INVERSION FOR AN ELASTIC SEDIMENT LAYER OVERLAYING A HOMOGENEOUS ELASTIC HALF-SPACE

A. Laboratory tank and bottom composition

Model-scale acoustic pressure measurements were made in a 10 cm deep shallow water waveguide in the frequency range 10 kHz to 7.5 kHz ($4<kH<30$). The bottom comprised a 0.07 m thick sandy layer overlaying a concrete base approximately 2 m thick. The experiment was performed in the ISVR's water tank facility of dimensions 8×8 m^2. Two Brüel & Kjær 8103 miniature hydrophones of 5 mm diameter were used to make the acoustic pressure measurements in the water column; one acting as the source and the other as receiver. The sound field was measured across range and depth by the traverse of the receiver hydrophone by two computer controlled stepper motors accurate to within 1/40 mm. The measurement grid was taken over 0.022(0.004)1.216 m in range from the source and over 0.005(0.005)0.1 m in height above the bottom. The source hydrophone was at 0.043 m depth below the surface and driven with a rectangular pulse of 10 μs duration. The source gain was set so to ensure sufficient signal to noise ratio, but not so high as to introduce nonlinear distortion. Acquisition of the received signal was performed at 250 kHz sampling

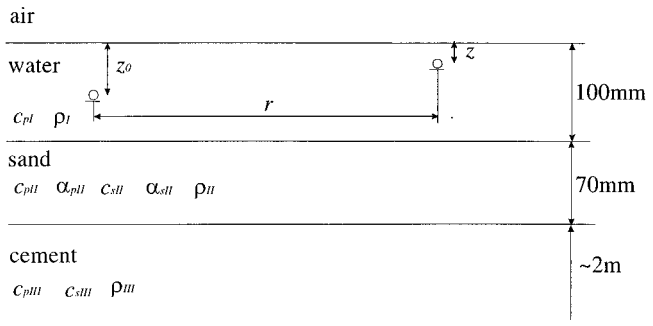


FIG. 7. Laboratory shallow water channel and its bottom composition (not to scale). The water, sand and cement layers have compressional wave speeds c_{pI}, c_{pII}, c_{pIII}, densities ρ_I, ρ_{II}, ρ_{III} and shear wave speeds of c_{sII}, c_{sIII} in the sand and cement layers, respectively. The sediment compressional and shear waves have respective attenuation rates of α_{pII}, α_{sII}, while those in the cement base are assumed to be negligible.

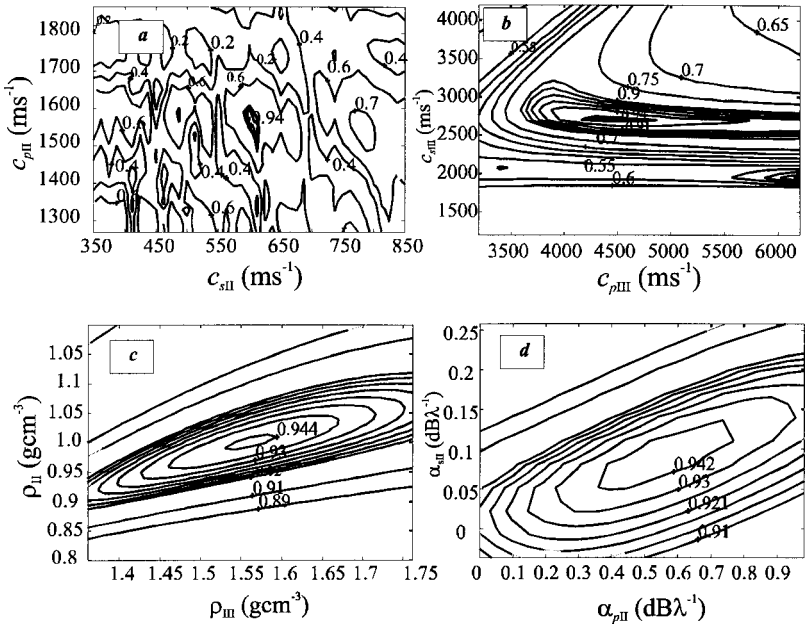


FIG. 8. Variations in Bartlett processor output about their optimal values in (a), sediment compressional and shear wave speeds, (b) basement compressional and shear wave speeds (c), densities, and (d), rates of attenuation in sediment wave speeds. Results based on 26 pressure measurements made in a 10 cm laboratory tank at 35 kHz, uniformly located between 0 and 1.216 m from the source. All values are held at their optimum values unless being varied.

frequency and time-synchronized to the excitation signal thus ensuring phase coherence between the source and receiver pressure signals. Fourier Transforming of the received time histories revealed levels of signal-to-noise ratio in excess of 50 dB in the frequency range of between 8 and 75 kHz, at all measurement positions. The measurement acquisition time period was limited to 4 ms so as to exclude from the received signal reflections from the tank walls. This constraint imposes a frequency resolution of 250 Hz in the measured data.

A sketch of the experimental waveguide and its bottom composition is presented in Fig. 7. The sediment layer of 0.07 m depth in relation to the 0.1 m depth of the laboratory water channel is unrealistically thick compared with the percentage sediment depth of most coastal regions. The sediment depth was chosen in order that the differences between the *S*-wave resonance frequencies that occur in the sediment layer are much greater than the 250 Hz measurement frequency resolution. Twenty-six pressure measurements uniformly positioned between 0 and 1.216 m at 30 mm depth at 35 kHz were used in to invert for the compressional wave speeds c_{pII}, c_{pIII}, shear wave speeds c_{sII}, c_{sIII}, their rates of attenuation α_{pII}, α_{sII}, and the densities ρ_{II}, ρ_{III}. The replica trial field was again computed from Eq. (8) with the reflection coefficient computed from the layered media theory due to Brekhovskikh and Lysanov.[8] The cement layer was assumed to be of infinite depth. Four ''slices'' through the eight-dimensional ambiguity-surface are shown in Fig. 8, and the optimal bottom parameter values obtained by this exhaustive search procedure tabulated in Table III.

Figure 8 illustrates the difficulties commonly encountered in a full field inversion procedure. While the wave speeds [Figs. 8(a) and (b)], are localized to reasonable accuracy, the densities and wave speed attenuations [Figs. 8(c) and (d)], are not. Large changes in their trial values give only small changes in the processor output. This is because their influence on the sound field is weak. Consequently, measurement errors in input parameter, such as water depth, will significantly affect the inversion results of these quantities. Moreover, this inversion required extensive computation time, despite the use of a simple Pekeris type solution for the trial sound field.

The procedure described above was repeated with the bottom reflectivity computed from the complex reflection phase gradient. Variations in the Bartlett and Maximum Likelihood correlators versus $\Phi_{\mathbf{R}}$ and $\Phi_{\mathbf{I}}$ are plotted in Figs. 9(a) and (b), respectively for the same pressure data used to obtain Fig. 8. The maximum likelihood output has been normalized to give a maximum output equal to unity.

Both figures show a distinct and unambiguous maximum value at $\Phi_0/\lambda = 0.15 + 0.18i$. Here, the Bartlett processor output is 0.917, which is only slightly less than the maximum output of 0.944 obtained by the use of a complete wave model of the bottom in terms of wave speeds and densities. The accuracy of the two-parameter reflection coefficient model at this frequency is to be expected since there are just two dominant modes. The linear complex phase model with two adjustable parameters must therefore fit exactly to the reflection coefficient at their two mode-angles. However, further confirmation of the validity of the sea-bottom model is provided in the next section where equally accurate sound field predictions are presented at higher frequencies (Fig. 10), which involve a larger number of dominant modes. The Φ_0 estimate in these high frequency examples will comprise

TABLE III. Summary of optimal parameters obtained by matched field inversion in the laboratory 0.1 m deep channel at 35 kHz from 26 pressure measurements uniformly located between 0 and 1.216 m from the source.

c_{pII}	c_{sII}	α_{pII}	α_{sII}	ρ_{II}	c_{pIII}	c_{sIII}	ρ_{III}
1570 ms^{-1}	595 ms^{-1}	0.55 dB λ^{-1}	0.09 dB λ^{-1}	1000 kgm^{-3}	4490 ms^{-1}	2710 ms^{-1}	1560 kgm^{-3}

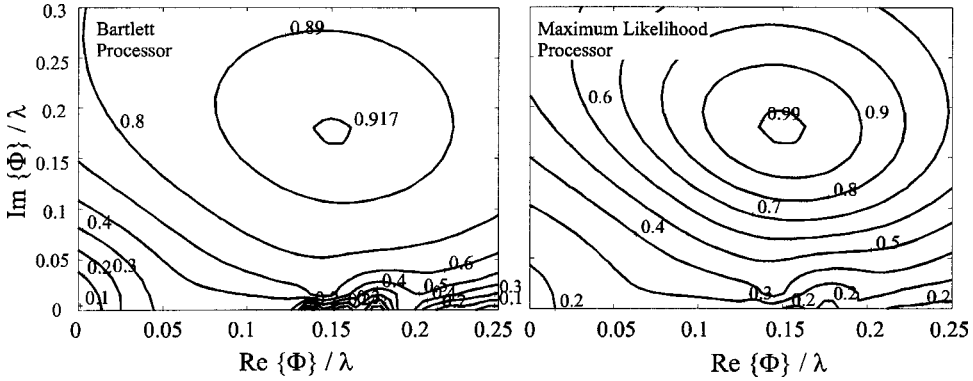


FIG. 9. Variation in Bartlett and Maximum Likelihood processor output respectively for varying real and imaginary Φ values at 35 kHz using 26 pressure measurements uniformly located between 0 and 1.216 m from the source.

of an average of all the modes weighted by their comparative contribution to the measured acoustic field.

B. Inversion of Φ versus frequency

The inversion results depicted in Fig. 9 at 35 kHz was repeated over the frequency range of between 5 kHz and 50 kHz in increments of the measurement FFT resolution of 250 Hz. The inverted values of Re{Φ} and Im{Φ} versus frequency are plotted as circles in Fig. 10. To illustrate the advantage of Φ as a robust inversion parameter, the inversion was performed using pressure measurements made at just 11 sensors of 0.03 m depth between the ranges of between 0.42 and 1.22 m in 0.08 m increments. For comparison is the predicted value of Re{Φ} and Im{Φ} obtained from the gradient of the complex reflection coefficient in the zero grazing angle limit, predicted from the wave model whose inversion parameters are listed in Table III. Also shown is the correlation coefficient between the measured pressure and the computer prediction obtained using the inverted Φ value (calculated over the 26-element array for consistency with previous results). The small but sudden drop in correlator output at about 22 kHz in Fig. 10 coincides directly with the cut-on frequency of the third mode. Moreover, results below about 9 kHz are unreliable due to low correlation between measurement and optimal prediction (upper figure of Fig. 10) caused by the low signal output of the B&K 8103 source hydrophone at these low frequencies.

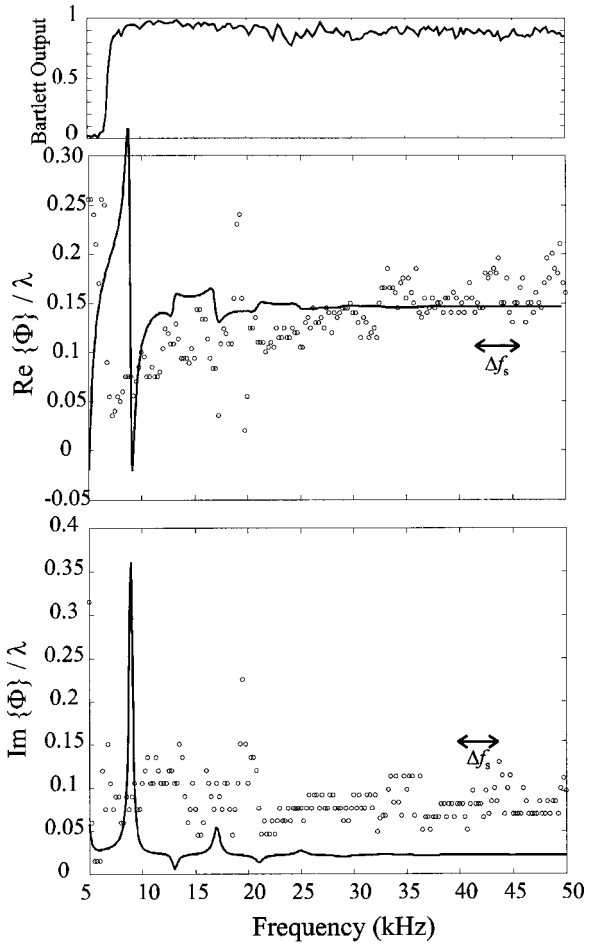


FIG. 10. Variation in the real and imaginary parts of the optimal complex reflection phase gradient Φ versus frequency, "σ." Predicted variation computed from the gradient of the low grazing angle reflection phase variation shown as a solid line. Bartlett output between fast field acoustic predictions obtained using Φ value and measurement, versus frequency shown in the upper plot.

Agreement between values of Re{Φ(f)} obtained by direct inversion, and that predicted from inverted wave speeds and densities, is reasonable overall, particularly as the measurements were made in just 10 cm of water at ultrasonic frequencies. The theoretical value of Im{Φ}, however, underpredicts the measured value by a factor of approximately 3. This finding implies that the actual bottom loss is significantly greater than can be accounted for by shear wave excitation at angles less than the critical angle. The main reason for this discrepancy is believed to be due to the use of pressure data measured close to the source. At this range there is a significant contribution to the acoustic field from modes that interact with the seabed at angles greater than the critical angle. The theoretical prediction, on the other hand, is calculated from the derivative of the theoretical reflection at small grazing angles much less than the critical angle. Attenuation of the compressional and shear wave speed in the bottom will also serve to increase the inverted values of Im{Φ}.

C. Deduction of sediment shear wave speed from the frequency dependence of Φ(f)

Both experimental and theoretical results in Fig. 10 exhibit weak oscillations in the behavior of Φ with roughly constant frequency equal to $\Delta f_s = 3.8$ kHz. The cause of this behavior in an elastic sediment layer overlaying a homogenous half-space is well documented. Unusually large reflec-

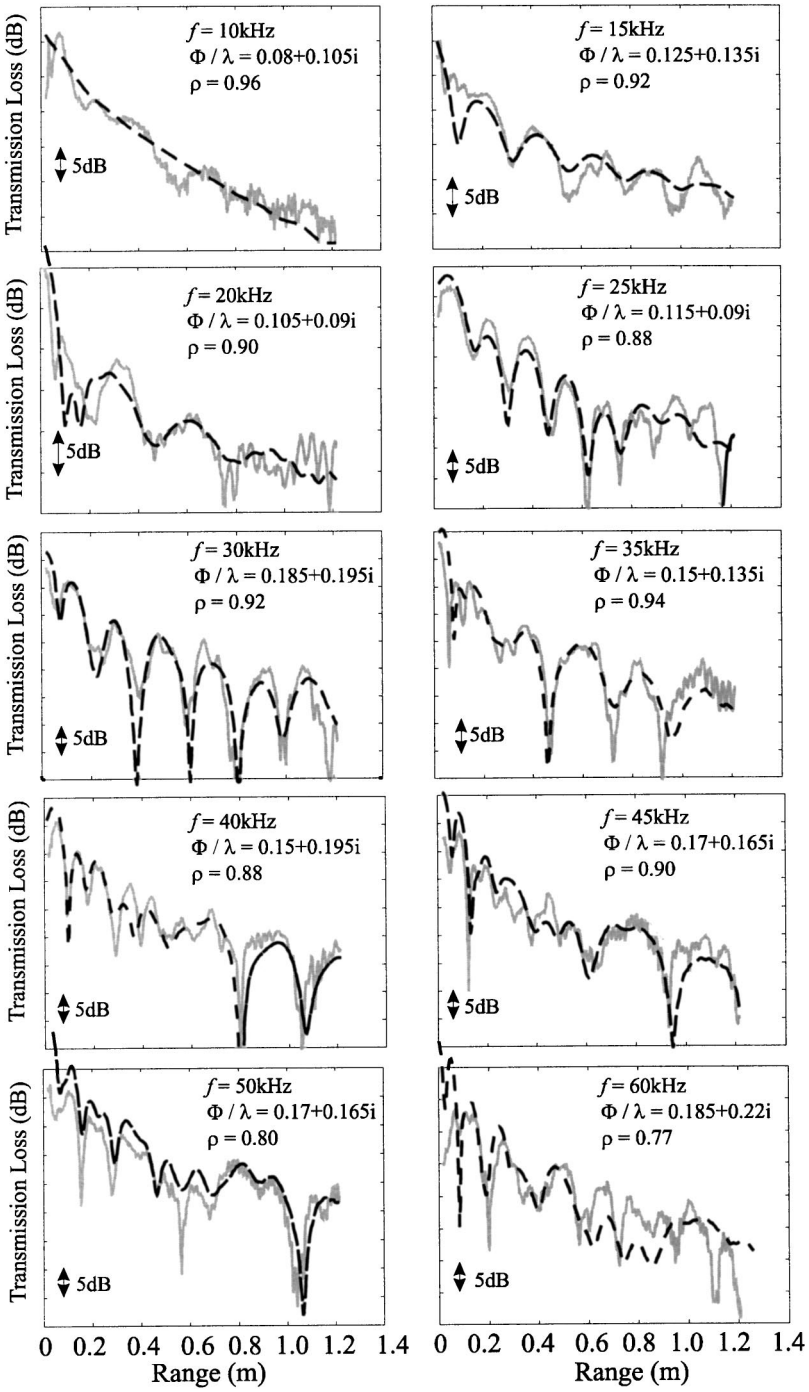


FIG. 11. Comparison of measured transmission loss (solid curve) with fast field model prediction obtained using optimal Φ values at frequencies: 10, 15, 20, 25, 30, 35, 40, 45, 50, and 60 kHz (dashed curve).

tion loss has been observed at frequencies corresponding to odd multiples of quarter wavelengths of the shear wave in the sediment layer.[9,10]

Assuming that this mechanism is also present in the laboratory tank bottom, and is the cause of oscillations in $\mathrm{Im}\{\Phi(f)\}$, the difference Δf_s between frequencies of high $\mathrm{Im}\{\Phi\}$ values (and hence high bottom loss since $BL \approx 8.7\gamma\,\mathrm{Im}\,\Phi$) is given by $\Delta f_s = c_{sII}/2h \sin\theta_{sII}$, where θ_{sII} is the grazing angle of the transmitted shear wave in the sediment layer and h is the thickness of the sediment layer. Relating θ_{sII} to the grazing angle of an incident plane in water θ_p using Snell's law and letting $\theta_p \to 0$ yields $\sin\theta_{sII} \underset{\theta\to 0}{=} \sqrt{1-(c_{sII}/c_{pI})^2}$. Eliminating $\sin\theta_{sII}$ from the two expressions above and solving for c_{sII} gives

$$c_{sII} = \frac{2h\Delta f_s}{\sqrt{1+(2h\Delta f_s/c_{pI})^2}}, \tag{22}$$

where Δf_s is the frequency separation between two peaks (and troughs) in the frequency variability in the bottom loss at small grazing angles. In sufficiently thin sediment layers $2h\Delta f_s/c_p < 1$, Δf_s, is a property predominantly of the shear wave speed in the sediment layer. Equation (22) in this case closely approximates to

$$c_{sII} \approx 2h\Delta f_s, \quad 2h\Delta f_s/c_{pI} < 1. \tag{23}$$

Measurement of Δf_s though the frequency periodicity of Φ may therefore be used to deduce the shear wave speed in the sediment layer. Evaluation of Eq. (23) for $\Delta f_s = (3800 \pm 500)$ Hz and $h = 0.07$ m gives a shear wave speed estimate equal to $c_s \approx (532 \pm 70)\ \mathrm{ms}^{-1}$. This is consistent with the value of $c_s \approx 590\ \mathrm{ms}^{-1}$ obtained by the direct inversion described in Sec. VI and plotted in Fig. 8(a).

As the frequency is increased, the peaks in the theoretical prediction of $\mathrm{Im}\{\Phi(f)\}$ in Fig. 10 become progressively smaller until oscillations are no longer observed. The theoretical predictions also indicate that the real and imaginary parts of $\Phi(f)$ tend asymptotically, in the high frequency limit, to the predictions of Eqs. (16a), (16b) based on a homogeneous, semi-infinite, elastic half space. Using the wavespeeds and density obtained in Sec. V for the sediment layer in these equations gives $\Phi/\lambda = 0.15 + 0.02i$. The real part of this prediction under-predicts the average high-frequency inverted values by about 15%, while the imaginary part is systematically underestimated by a factor of 3. This discrepancy is consistent with the error associated with the inversion principle generally, as indicated by the error in the simulation results in Sec. IV. Another reason for the disagreement is the use of pressure data made at close range from the source, which yields Φ values that are biased towards the heavily damped modes that interact with the bottom at grazing angles exceeding the critical angles. The theoretical prediction, however, is obtained from the gradient of the reflection coefficient at grazing angles below the critical angle. The use of pressure data measured at typical sonar detection ranges is therefore anticipated to yield Φ-values in closer agreement to the theoretical prediction.

D. Transmission loss prediction from Φ

Examples of the predicted transmission loss with range predicted from the fast field model with plane wave reflection coefficients calculated from optimal Φ values is presented in Fig. 11 at a number of discrete frequencies. Agreement is extremely good, even at the highest frequency presented of 60 kHz.

VII. CONCLUSION

A quantity is defined in this paper, referred to as the complex reflection phase gradient, whose two defining parameters may be easily inverted from pressure measurements made in a shallow water channel. It is based on the observation that at low grazing angles the reflection phase and bottom loss commonly exhibits an approximate linear relationship to the vertical component of the acoustic wave number at the seabed. The real part of the complex phase gradient specifies the rate at which the reflection phase varies with vertical acoustic wave number while the imaginary part quantifies the rate of change of bottom loss. Despite being defined with just two real parameters this, and other studies on the effective depth approximation, have shown that it provides accurate predictions of the sound field not too close to the source for a wide range of bottom types.

The motivation for this choice of model is that its two parameters are readily inverted from comparatively few acoustic pressure measurements. The paper demonstrates that the inversion results obtained from the pressure measurements at just 11 locations in a 50 m deep shallow water, iso-speed waveguide at 100 Hz overlaying an elastic half-space are in close agreement with the theoretical expression presented in the paper. These values were used to obtain estimates for the two wave speeds in a homogenous medium of known density. Furthermore, the complex phase gradient has been shown to be closely related to the input impedance of the seabed, becoming exact in the limit of zero grazing angle.

Application of the technique to a laboratory-scale 10 cm deep shallow water waveguide has been presented at frequencies between 10 and 75 kHz. The two inversion parameters were in satisfactory agreement with those predicted from the wave speeds and densities of the sediment layer and substrate obtained by conventional matched field inversion. Most importantly, the transmission loss predicted using the two-parameter representation of the sea-bottom is shown to be very close agreement with measurement over a wide range of frequencies.

ACKNOWLEDGMENTS

The author would like to thank DERA Winfrith for the sponsorship of this work. Assistance of Dr. Ben Cox in obtaining much of the experimental data is also gratefully acknowledged.

[1] M. J. Buckingham, "Array gain of a broadside vertical line array in shallow water," J. Acoust. Soc. Am. **65**, 148–161 (1979).
[2] D. M. F. Chapman, P. D. Ward, and D. D. Ellis, "The effective depth of the Pekeris Ocean waveguide, including shear wave effects," J. Acoust. Soc. Am. **85**, 648–653 (1989).
[3] S. A. L. Glegg, "The effective depth approximation for sound propagation in shallow water over a sediment layer and a hard rock basement," J. Acoust. Soc. Am. **94**, 3302–3310 (1993).
[4] M. V. Hall, "Acoustic reflectivity of a sandy seabed: A semianalytic model of the effect of coupling due to the shear modulus profile," J. Acoust. Soc. Am. **98**, 1075–1089 (1995).
[5] Z. Y. Zhang and C. T. Tindle, "Complex effective depth of the ocean bottom," J. Acoust. Soc. Am. **93**, 205–213 (1993).
[6] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," J. Acoust. Soc. Am. **83**, 571–578 (1988).
[7] H. P. Bucker, "Sound propagation in channel with lossy boundaries," J. Acoust. Soc. Am. **48**, 1187–1194 (1970).
[8] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer-Verlag, New York, 1990).
[9] N. R. Chapman and D. M. F. Chapman, "A coherent ray model of plane-wave reflection from a thin sediment layer," J. Acoust. Soc. Am. **94**, 2731–2738 (1995).
[10] M. A. Ainslie, "Plane-wave reflection and transmission coefficients for a three-layered medium," J. Acoust. Soc. Am. **97**, 954–961 (1995).

The contribution of bubbles to high-frequency sea surface backscatter: A 24-h time series of field measurements

Peter H. Dahl[a)]
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

(Received 13 February 2002; revised 18 October 2002; accepted 29 October 2002)

Measurements of acoustic sea surface backscattering, wind speed, and surface wave spectra were made continually over a 24-h period in an experiment conducted in 26 m of water near the Dry Tortugus collection of islands off south Florida in February 1995. The backscattering measurements were made at a frequency of 30 kHz and a sea surface grazing angle of 20°; a time series of the decibel equivalent of this variable, called SS20, was studied in terms of its dependence on environmental variables. On occasion reliable estimates of scattering in the grazing range 15°–27° were also obtained during the 24 hours. The scattering data exhibited evidence, in terms of scattering level and grazing angle dependence, of scattering from near-surface bubbles rather than scattering from the rough air–sea interface. The scattering data were compared with a model for σ_b, the apparent backscattering cross section per unit area due to bubble scattering, that is driven by a parameter, β_I, equal to the depth-integrated extinction cross section per unit volume. Using an empirical model for β_I based on data from a 1977 experiment conducted in pelagic waters, model predictions agreed reasonably well with the 1995 measurements presented here. Additional model–data comparisons were made using four measurements from a 1992 experiment conducted in pelagic waters. Finally, the 24-h time series of acoustic scattering exhibited a hysteresis effect, wherein for a given wind speed, there was a tendency for the scattering level to be higher if prior winds had been falling. A better understanding of this effect is essential to reduce uncertainty in model predictions.
 [DOI: 10.1121/1.1532029]

PACS numbers: 43.30.Cq, 43.30.Ft, 43.30.Gv, 43.30.Hw [WMC]

I. INTRODUCTION

It is well known that scattering from bubbles located close to the air–sea interface can be a significant, if not the dominant, source of apparent sea surface reverberation, particularly for grazing angles less than about 30° and for frequencies of $O(10)$ kHz and above.[1–3] Controlled studies performed in wave tanks, where the bubble-producing effects of natural breaking waves can be suppressed using short fetches, have shown that the area-normalized backscattering cross section of a rough air–water interface is the same when measured with microwave and acoustic systems operating at the same wavelength and surface grazing angle.[4] This is expected on the basis of first-order perturbation, or Bragg theory, which predicts the scattering levels to be the same whether the scattered wave originates from above or below the rough air–water interface.

In field studies[3] it has been shown that scattering levels from microwave and acoustic systems with equal wavelengths and grazing angles quickly part ways when the wind speed exceeds a threshold value between 2 and 3 m/s. The same study showed that above this threshold the acoustic cross section varied as U^5 whereas the microwave cross section varied as U^2, where U is wind speed. This rather low wind speed threshold has also been observed in a field study of ambient sound, for which an abrupt increase in the noise level was associated with the onset of sporadic, small scale wave breaking that occurred once the wind speed exceeded a value between 2 and 3 m/s.[5]

Further evidence of apparent sea surface reverberation caused by bubble scattering is presented here in the form of a 24-h time series of measurements made at 30 kHz in an experiment conducted off the coast of southern Florida in 1995. In addition to the scattering measurements, wind speed, directional surface wave spectra, and water column sound speed were continually measured. The time series illustrates features of the relation between acoustic scattering and wind speed not seen in individual or point measurements of scattering and wind speed. The scattering data are interpreted using a model for the apparent backscattering cross section per unit area due to bubble scattering σ_b. The model predicts the angular dependence of bubble scattering and is driven by a parameter describing the concentration of bubbles that is necessarily derived from field data. For a comparison, a model for microwave backscattering from the sea surface by Plant[6] is used to predict the backscattering cross section per unit area due to scattering from the rough air–sea interface σ_r. The inputs to this model include the same grazing angle, and a microwave frequency (close to C band), such that the microwave wavelength matches the acoustic wavelength at 30 kHz. In addition, the H-polarization is taken, the relative dielectric constant is set to infinity, and the acoustic convention for the definition of the cross section per unit area[3] is applied to the results.

The scattering measurements for the 24-h experiment were taken in a limited range of grazing angles, approximately 15°–25°, and scattering due to bubbles is expected to

a)Electronic mail: dahl@apl.washington.edu

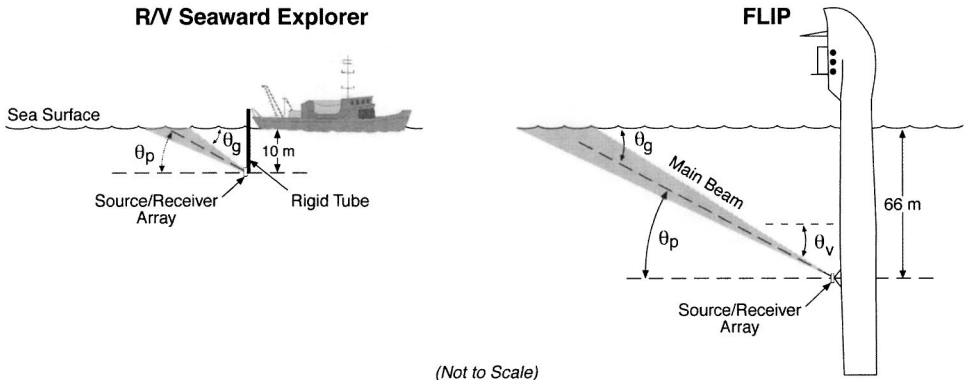


FIG. 1. Experimental apparatus and layout for measuring backscattering from the sea surface. The 1995 Florida measurements from the R/V SEAWARD EXPLORER (left), and 1992 measurements from the research platform FLIP (right). The three angles θ_p, θ_v, and θ_g are discussed in the text.

be angle independent according to the model for σ_b. To further illustrate the dependence on the sea surface grazing angle, three measurements from an experiment conducted in 1992 from the research platform FLIP are discussed. Each was made during a single wind-speed condition, but low ($\sim$1°–6°) and high (35°–45°) grazing angles were sampled. It is also of interest to compare the FLIP measurements made in pelagic conditions with the Florida measurements made in littoral conditions.

In Sec. II the Florida experiment and data reduction methods are described, and the 24-h time series of acoustic and environmental measurements is presented. In Sec. III the model for σ_b is given and compared with four experimental runs selected from the 24-h experiment. In Sec. IV the FLIP measurements are presented along with additional data reduction procedures required to handle the case of refraction. Results from both experiments are discussed in Sec. V, and a summary is given in Sec. VI.

II. FLORIDA EXPERIMENT: A 24-h TIME SERIES OF SEA SURFACE SCATTERING

The 24-h sea surface backscattering experiment was conducted 14–15 February 1995 near the Dry Tortugus collection of islands (24°36.7′ N, 82°50.7′ W) off southern Florida in waters 26-m deep from the research vessel SEAWARD EXPLORER (length 33 m). A 24-h measurement period was chosen to capture environmental effects in acoustic scattering associated primarily with diurnal variation in wind speed. During this period water column salinity and temperature were frequently measured with conductivity–temperature–depth (CTD) casts. Surface-to-bottom salinity remained near 36 ppt$\pm$0.05 ppt, temperature near 20.8 °C $\pm$0.15 °C, putting the sound speed at 1524.9 m/s $\pm$0.3 m/s. Wind speed measurements were made with a Coastal Climate anemometer (from which the air temperature was also recorded) mounted 6.9 m above the sea surface at the top of the U frame on the stern of the research vessel. Although this measurement position was not shadowed by a bridge superstructure, other structures located on top of the bridge (three large antennas, radar, etc.) may have influenced the wind measurements to a small degree that is not estimated. Wind speed was recorded continuously as a 5-min average every 10 min. Surface wave directional spectra were measured using a Datawell wave buoy (from which the sea temperature was also recorded) moored approximately 100 m from the SEAWARD EXPLORER and data were sent back to the vessel every hour via a radio modem. Over the 24-h period the air temperature varied slowly between 20.5 °C and 22.5 °C; thus stable air–sea boundary layer conditions are indicated by the air–sea temperature difference having remained $\leqslant$2°. An earlier work[7] discusses measurements of bistatic forward scattering made in the days following the 24-h experiment, including additional details on the environment.

Figure 1 (left side) shows the field apparatus and experimental layout for the backscattering measurements. An approximately rectangular transmitting array (14.9 cm by 21.4 cm) was suspended at a depth of 10 m off the stern of the SEAWARD EXPLORER, which was held in a four-point moor. The array was held in position, i.e., prevented from uncontrolled rotation and heave motion, by a rigid tube attached to an active heave compensation device, and the array's pitch angle was remotely monitored and set. The transmitting array's (measured) horizontal beamwidth was 15° and the vertical beamwidth was 20° (each defined as a full angle between points 3 dB down from the maximum response axis). The transmitting array was bisected horizontally and vertically into four, equal-sized subaperatures, from which a four-channel data stream was recorded. For each subaperature, the horizontal beamwidth was 36° and the vertical beamwidth was 23°. Henceforth the receiving subaperatures will be referred to as quad beams, and the transmitting main array as the sum beam. An experimental run consisted of a set of 20 pulses transmitted at 5-s intervals, thus taking 100 s to complete. The pulses were of length 1 ms with center frequency 30 kHz. Angular resolution for the scattering strength estimates in the grazing angle range 15°–27° is approximately 0.5°, based on the 1-ms pulse. Upon completion of an experimental run, the data were evaluated to confirm the proper setting of gains and geometry, given the ever-changing conditions and the possibility of spurious targets entering the scattering zone. After an initial evaluation of the data, another experimental run was started. The typical time interval between runs was 30 min, but on occasion this interval was reduced to about 5 min, with a total of 58 experimental runs completed in 24 h.

Scattering strength is the decibel equivalent to the backscattering cross section per unit area per unit solid angle (e.g., see Ref. 3). The reverberation level is first computed by averaging the squared envelopes of the 20 received echoes recorded on one channel, and thus our estimates are based on

transmitting on the sum beam and receiving on a wider, quad beam. The sonar equation is then used to estimate scattering strength from estimates of the calibrated reverberation level, from which an estimate of the noise has been subtracted. (The noise was estimated by making an experimental run otherwise identical to an actual run, but with transmitter turned off.) Here we remark only on the more subtle calculation of the effective scattering area, which in this case is pulse-length limited. A parameter $B(\theta_v;\theta_p)$ that incorporates the transmit and receive beam patterns, and the pitch angle θ_p of the maximum response axis of the transmit array, is computed by

$$B(\theta_v;\theta_p)=\int_{-\pi}^{\pi} b_T(\theta_v-\theta_p,\phi)b_R(\theta_v-\theta_p,\phi)d\phi, \quad (1)$$

where b_T and b_R are the transmit (sum beam) and receive (quad beam) beam pattern functions, θ_v is the vertical arrival angle, and ϕ is azimuthal angle. (Note that this parameter varies with the range or time delay by way of θ_v.) For the (effectively) nonrefracting conditions of the Florida experiment, θ_v equals the sea surface grazing angle θ_g. For computing $B(\theta_v;\theta_p)$, b_T and b_R are each modeled as the product of two, squared sinc functions, whose arguments are set to match the measured horizontal and vertical beam widths given previously. The effective scattering area for a given grazing angle $A(\theta_v)$ is defined as $A(\theta_v)=R(c\tau/2)B(\theta_v;\theta_p)$, where R is slant range, τ is pulse length, c is sound speed, and where the nonuniform ensonification of the scattering area is accounted for by $B(\theta_v;\theta_p)$, e.g., as done in Ref. 8. For grazing angles close to θ_p, $B(\theta_v;\theta_p)$ reduces to approximately the horizontal width of the transmit array, or 15°. To avoid dependence on the less reliable modeling of the transmit and receive beam patterns away from their respective main lobes, we reject scattering strength estimates for which the accompanying $10\log_{10}B(\theta_v;\theta_p)$ is more than 5 dB lower than its maximum value.

The uncertainty in the measurements is estimated as follows. First, the combined uncertainty associated with the calibration of all system parameters is ± 1 dB. Second, we assume that the 20 echo amplitudes gathered over a 100-s measurement period will be subject to the large time scale (i.e., >10 s) fluctuations in surface reverberation caused by bubbles, of the kind discussed in Ref. 9. Using the probability density function from Ref. 9, we estimate that statistical uncertainty for a 20-echo sample is ± 1.5 dB. These two uncertainties combine to give approximately ± 2 dB for the estimates of scattering strength expressed in decibels.

An estimate of the vertical arrival angle was computed for each experimental run using the quad beam data as a check to ensure that only reverberation originating from the sea surface entered into the final estimates of scattering strength. Complex data streams (each proportional to pressure) for the upper and two lower quad beams were coherently combined giving $s_U(t)$ for the upper, and $s_L(t)$ for the lower half of the main array. The phase of $s_U s_L^*$ averaged over 20 pings, or ψ, provides an estimate of the vertical arrival angle $\hat{\theta}_v$ via

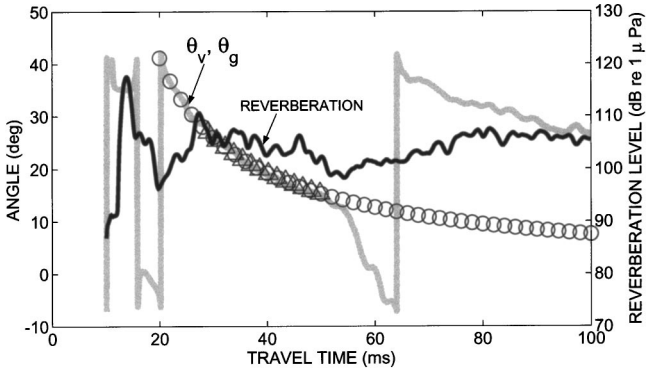


FIG. 2. Averaged reverberation level based on 20 echoes (black line, right-hand scale) and estimated vertical arrival angle (gray line, left-hand scale) for one experimental run from the 24-h Florida experiment. The line formed by circles is the expected vertical arrival angle (θ_v) and sea surface grazing angle (θ_g) based on a depth-independent sound speed of 1525 m/s. The line of triangles identifies the segment of data used to compute sea surface backscattering.

$$\hat{\theta}_v-\theta_p=\arcsin\left(\frac{1}{2\pi}\frac{\psi\lambda}{L}\right), \quad (2)$$

where θ_p is the transducer's pitch angle (17°), λ is acoustic wavelength (5.1 cm), and L is the center-to-center distance between the upper and lower halves of the main array (6 cm). The estimate is a function of delay time, and is smoothed with a boxcar filter of length equal to one pulse length. Figure 2 shows the reverberation level from one of the four quad beams (the four quad beams were well matched, and produced essentially equal calibrated output) plotted with a corresponding estimate of $\hat{\theta}_v$. The line of circles is the expected vertical arrival angle based on isovelocity propagation conditions with sound speed equal to 1525 m/s. The line of triangles corresponds to the time segment of reverberation data used in the sea surface scattering strength calculations. This line begins at the point where the $10\log_{10}B(\theta_v;\theta_p)$ criterion is satisfied and ends at the point where it is presumed that high-angle bottom scattering enters a down-looking side lobe. The usable time segment of reverberation data for this experimental run maps to sea surface grazing angles between 15° and 27°, and the estimated sea surface scattering strength is a constant—43 dB within this angular range. Not every experimental run successfully captured reliable estimates within this same angular range. On occasion, buoys associated with the ship's four-point mooring and other ongoing experiments drifted into a particular scattering zone and caused interference with the measurements.

The scattering zone associated with grazing angle $\theta_g=20°$ (or range 27 m) was always kept free of interference, and the most complete time series consists of 58 scattering strength estimates made at this angle; this variable is called SS20 and is based on the average of 20 echoes. Estimates of SS20 along with other environmental measurements taken during the 24-h period are displayed in Fig. 3. The origin of the time axis is 1510 UTC on 14 February 1995, or 1010 local time. The upper plot shows SS20 (red) plotted with variable U_{10} (blue), defined as the average wind speed in 5 min converted to a 10-m measurement height using the mea-

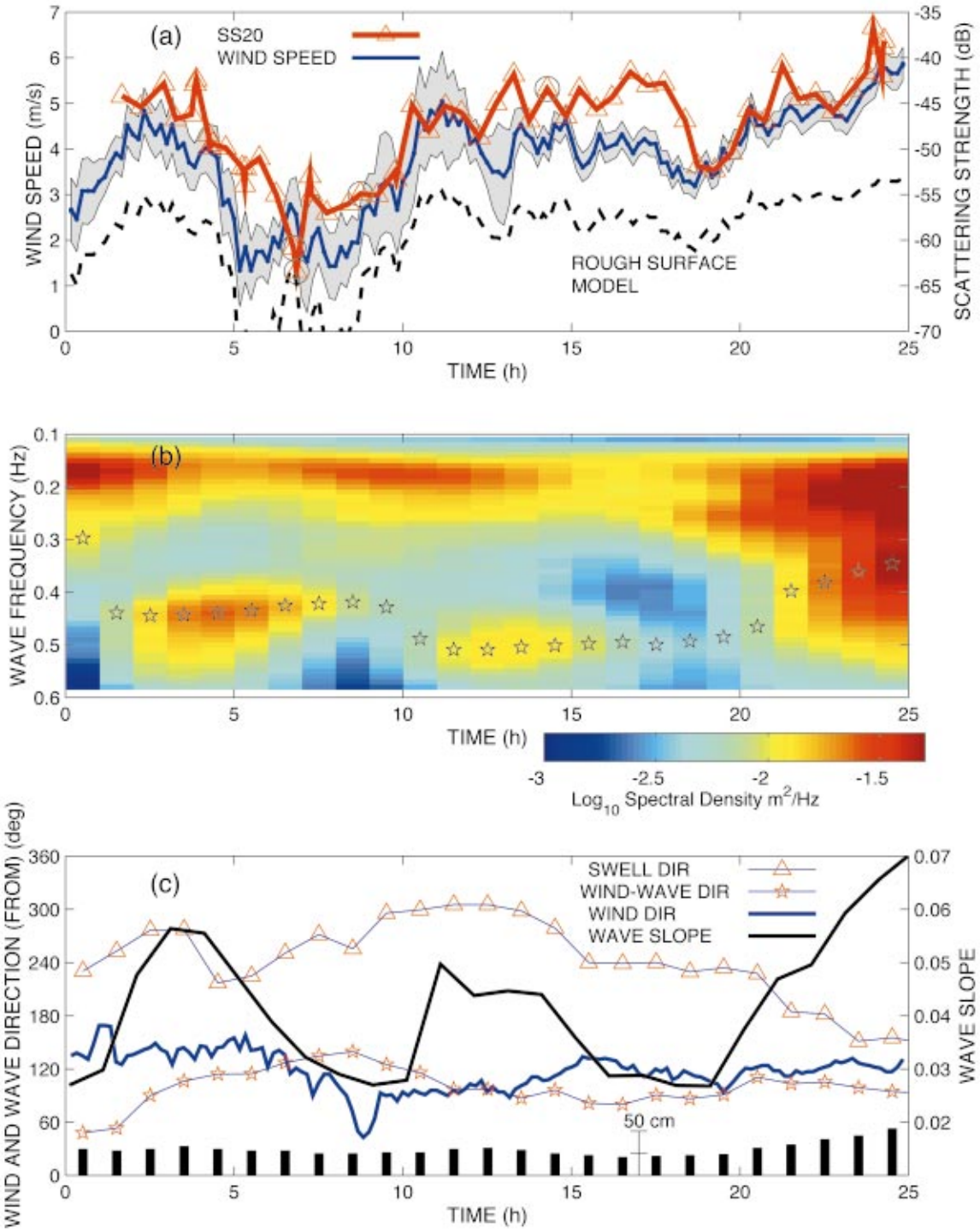


FIG. 3. (a) Sea surface backscattering strength at grazing angle 20°, or SS20 (red line and symbol, right-hand scale), plotted against U_{10} (blue line, left-hand scale), and $\pm$ the standard deviation of U_{10} (gray shade). The black, dashed line is a model estimate of SS20 based only on rough surface scattering (Ref. 6). Estimates of SS20 within the circles are discussed further in the text. (b) The time history of the sea surface wave variance spectrum; the star symbols mark the spectral peak of wind-generated waves. (c) Wave slope (black line, right-hand scale), and the direction of wind and wave systems (left-hand scale), these being the wave swell (triangle), wind-generated waves (star symbol), and wind (blue line). The significant wave height is given by the length of vertical bars with a 50-cm scale shown.

sured air and sea temperature difference and formulation given by Ref. 10 (resulting in a U_{10} that was nominally 2.5% greater than the wind speed measured at 6.9-m height). The gray area represents U_{10} $\pm$ the wind speed standard deviation for each 5-min averaging period. During the first 12 h, this value varies between 0.5–1 m/s, and during the second 12 h it remains at approximately 0.3 m/s. The black dashed line is a simulation for σ_r based on Plant's model,[6] assuming U_{10} as given by the blue line, but with a fixed standard deviation of 1 m/s. The standard deviation is important at low wind speeds,[11] but has little effect on the simulation for winds greater than about 4 m/s.

There is a manifest correlation between the two processes, SS20 and U_{10}, with each process showing four major

up/down swings occurring with a time scale of about 5 h. But there also appears to be a hysteresis effect, insofar as the U_{10}, SS20 relation depends somewhat on whether the prior winds had been rising or falling. In addition, at very low wind speeds (say <3 m/s) the relation between wind speed and scattering is weakened further.

The middle plot displays the set of sea surface variance spectra. These spectra are bimodal with a roughly constant, low-frequency peak at about 0.18 Hz that is associated with swell, plus a higher-frequency peak associated with wind waves, the location of which is identified for each spectrum by the star symbol. The lower plot displays the propagation direction for the swell (triangle) and wind-wave (star) systems, and wind (solid blue line). Note that the sonar look direction was maintained at $300° \pm 15°$, or very nearly downwind for the entire 24-h period. The swell system largely opposes the wind–wave system until about hour 20, at which point both systems become more closely aligned with the wind. Also plotted for reference are two integral measures of the surface variance spectra: the rms wave slope (black line), and the significant wave height (vertical bars) computed here by taking four times rms surface roughness, or H.

III. MODEL FOR APPARENT SEA SURFACE BACKSCATTER ORIGINATING FROM NEAR-SURFACE BUBBLES

Our approach to modeling apparent sea surface backscattering associated with scattering from near-surface bubbles follows Dahl *et al.*[3] Here, we give only the model for σ_b, which is

$$\sigma_b = \frac{\sin\theta_g}{8\pi}\frac{\delta_r}{\delta}(1+8\beta e^{-2\beta}-e^{-4\beta}), \tag{3}$$

where $\beta = \beta_I/\sin\theta_g$, δ_r is the radiation damping constant at resonance (taken to be 0.0136), and δ is the total damping coefficient at resonance (taken to be 0.0792 at 30-kHz frequency[3]). The basis of Eq. (3) is from Crowther's analysis,[12] and further discussions on variations of this model and its origins are available in Refs. 2, 13, 3. Equation (3) assumes that the parameter $\chi = 2kH\sin\theta_g$ is $\gg 1$, where k is the acoustic wave number, which at 30 kHz is satisfied by the sea surface conditions in effect during the 24-h experiment. Recent laboratory work[14,15] of the cases $\chi = 0$ and $0 < \chi < 1$, verifies the basic premise of Eq. (3) regarding multiple ensonification by four paths, two of which add coherently. The primary variable relating to the density of bubbles is β_I, defined as the depth-integrated extinction cross section per unit volume due to bubble scattering and absorption. This is a dimensionless quantity that succinctly characterizes the concentration of near-surface bubbles, and that must necessarily be estimated from field data. An empirical model for β_I is given in Sec. V. The complete model for scattering strength, or $10\log_{10}\sigma$, including both bubble and rough surface components, is

$$\sigma = \sigma_b + \sigma_r e^{-2\beta_I/\sin\theta_g}, \tag{4}$$

where the factor multiplying σ_r represents a reduction in intensity owing to two-way travel through the bubble layer.

Figure 4 shows scattering strength versus grazing angle

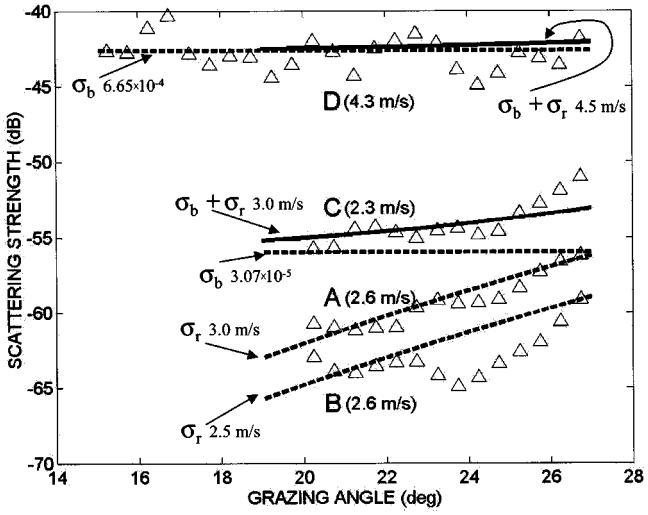


FIG. 4. Scattering strength versus grazing angle for four experimental runs A–D, in temporal order) identified by the four circles in Fig. 3(a). Runs A and B ($U_{10}=2.6$ m/s) are suggestive of Bragg scattering. The dashed lines near these data are the result of Plant's model for σ_r driven by $U_{10}=3$ m/s (upper line) and 2.5 m/s (lower line). Run C ($U_{10}=2.3$ m/s) exhibits a transition between Bragg and bubble scattering. Here the dashed line is σ_b from Eq. (3) with $\beta_I = 3.07\times10^{-5}$ and the solid line is the sum of σ_b and σ_r [via Eq. (4)] with σ_r based on $U_{10}=3$ m/s. Run D exhibits bubble scattering and the dashed line is σ_b with $\beta_I = 6.65\times10^{-4}$ and the solid line is Eq. (4) with σ_r based on $U_{10}=4.5$ m/s.

for four experimental runs (A–D), each identified by a circle in the upper plot of Fig. 3. Runs labeled A and B are suggestive of scattering from the air–sea interface, and at these grazing angles Plant's model[6] shows that this is Bragg scattering. These runs were taken 5 min apart (the data with the higher level taken first), and represent the weakest scattering levels measured in the 24-h period. The equivalent noise scattering strength level is approximately -70 dB, but interference (as previously mentioned) limits the lowest observable grazing angle to approximately 20° for these runs, plus the one with the next highest level. The dashed lines near A and B are the result of Plant's model for σ_r assuming U_{10} is 3 m/s (upper line) and 2.5 m/s (lower line), with a standard deviation of 1 m/s incorporated into the model for each case. Although the lower-level data drops out for grazing angles in the vicinity of 24°, quite possibly due to patchiness in sea surface roughness, both experimental runs display a nominal trend of increasing scattering strength with increasing grazing angle that is close to $\sigma \sim \tan^4\theta_g$, as expected for Bragg scattering from waves whose wave numbers are close to $2k\cos 20°$.

Clearly, however, at these very low wind speeds the relation between wind speed and scattering strength is encumbered with high variability. For example, the 5-min averaged wind speed taken just prior to both measurements A and B was the same, 2.6 m/s, yet their scattering levels differ by about 3 dB. Furthermore, the experimental run C giving the next higher level of scattering has associated with it an even lesser U_{10} of 2.3 m/s. It is possible that the two experimental runs exhibiting lower levels indicative of Bragg scattering are anomalous, i.e., taken at an instant during which there was a momentary gap in the bubble field with a high degree of patchiness in the spatial distribution. In any case, run C

appears to exhibit a kind of transition between Bragg and bubble scattering, with bubbles scattering having the dominant effect. The dashed line near the C data is Eq. (3) with $\beta_I = 3.07 \times 10^{-5}$, and this line remains constant with grazing angle (until θ_g falls well below 0.2°, at which point $\sigma_b \sim \sin \theta_g$). The solid line is the sum of σ_b and σ_r [via Eq. (4)], with σ_r based on $U_{10} = 3$ m/s, and it seems to slightly better capture the angular trend in the data, with the exception of the data near 26°. Note that were σ_r to be based on the $U_{10} = 2.5$ m/s, its addition to σ_b would have essentially no effect.

Finally, bubble scattering is apparent in the experimental run D. Here, the corresponding U_{10} was 4.3 m/s and the dashed line is Eq. (3) with $\beta_I = 6.65 \times 10^{-4}$, which remains flat until θ_g falls well below 0.5°. The solid line is the sum of σ_b and σ_r with σ_r based on $U_{10} = 4.5$ m/s; at this wind speed and range of grazing angles, σ_r becomes insignificant compared to scattering from bubbles.

IV. MEASUREMENTS FROM FLIP

The FLIP measurements were made as part of a series of experiments conducted in January 1992 from the research platform FLIP operating off the coast of California (32 °N, 125 °W, or 400 nm west of Los Angeles). Some results from these experiments (not involving sea surface backscattering *per se*) and additional information on the environment are described elsewhere.[16,17] Here, we present three sets of measurements of sea surface backscattering; each set is equivalent to one of the single measurements made during the course of the 24-h Florida experiment. In addition to covering both a lower and higher range of sea surface grazing angles, the FLIP sea surface backscattering measurements, collected in pelagic waters 4000-m deep, provide an interesting comparison to the Florida backscattering data collected in warmer, littoral waters 26-m deep.

Figure 1 (right side) shows the experimental geometry for the FLIP measurements. The transducer, mounted on FLIP's hull at a depth of 66 m, was the same one used in the Florida measurements, but rotated 90° such that the wider beam was horizontal and the narrower beam was vertical to reduce interference originating from FLIP's hull and other fixtures mounted on the hull. As in the Florida measurements, an experimental run consisted of 20 pulses transmitted at 5-s intervals. Presented first are two experimental runs designed to measure sea surface backscattering at very low grazing angles made on 24 January 1992 at 1653, for which $U_{10} = 8.3$ m/s, and at 1725, for which $U_{10} = 6.0$ m/s. (FLIP experimental times are UTC −8 h, and wind speed measurements from FLIP were made at a height of 10 m, and based on an averaging time of 10 min). For these measurements the transducer was pitched upward with $\theta_p = 2.5°$ (measured remotely with tilt meter) and the pulse length was 8 ms, which set angular resolution to approximately 0.1°. In contrast to the isospeed conditions that characterized the Florida site, the FLIP measurements were made in conditions for which the sound speed varied significantly with depth. We use ray theory to obtain the necessary mapping between the *observable* vertical arrival (launch) angle θ_v, and the corresponding local sea surface grazing angle θ_g that is not an observable in the data. This is an arguably subtle computation, the results of which impact interpretation of the surface scattering measurements made at low grazing angles.

First, we model the sound speed versus depth profile $c(z)$ using data from a CTD cast made on 24 January at 1908, combined with subsurface temperature measurements made simultaneously with the acoustic measurements. The temperature measurements were made continually at depths 10, 50, and 100 cm from a wave-following thermistor chain, and at depth 6 m from a sensor attached to FLIP's hull; all data were recorded in 10-min averages.[18] The thermistor chain data are combined with the salinity data from the 1908 CTD cast to produce a sound speed versus depth profile for the upper 1 m, with the remainder of $c(z)$ derived directly from the CTD data. This is done because there was a temperature drop in the upper 1 m of the water column of 0.04 °C between the first and second acoustic experimental run (approximately 0.5 h), followed by another temperature drop of 0.04 °C in the intervening 1.5 h between the second acoustic experimental run and the 1908 CTD cast. During this period there was essentially no change in the temperature data measured at 6 m. Because $c(z)$ varies both temporally and spatially, the one used to account for refraction in the acoustic measurements can, at best, be viewed as only one member of an ensemble of possible profiles.

Second, there will be some small change in sound speed near the surface Δc associated with the bubbly medium and not reflected in the CTD or thermistor chain data. Using the approach outlined in Ref. 19, we estimate Δc at the surface to be −0.1 m/s based on the 30-kHz frequency, an assumed bubble size spectrum $N(a)$ that goes as a^{-4} for bubble radii $a \geqslant 20\mu$ and is flat for $a < 20\mu$, and a nominal void fraction of 10^{-8}. Changes in $N(a)$, other than radical ones, have little effect on Δc, while changing the void fraction by an order of magnitude produces a magnitude change in Δc. A nominal void fraction for two experimental runs made on 24 January is estimated as follows. First, the e-folding vertical depth scale of the near-surface bubble layer L is taken to be ~0.25 m. This scale is estimated from Fig. 2 of Ref. 16, which is based on 30-kHz upward-looking sonar data from the same experiment; data from this figure, along with Eq. (9) of Ref. 16 produce a void fraction between $10^{-8.5}$ and $10^{-7.5}$. Next, we fit the measurements from 24 January with the model of Eq. (3) using $\beta_I = 0.004$; upon converting this value to a void fraction based on $L = 0.25$, the result is $10^{-7.7}$. We thus set the nominal void fraction to be 10^{-8}, with upper bound ($10^{-7.5}$) putting $\Delta c = -0.3$ m/s and lower bound ($10^{-8.5}$) putting $\Delta c = -0.03$ m/s. Finally, to incorporate Δc into the bubble-free sound speed profile $c(z)$, define $c_b(z) = c(z) + \Delta c e^{-z/L}$, where $c_b(z)$ includes the effects of a bubbly medium. The $c_b(z)$ profile is shown in Fig. 5(a) although it is indistinguishable from $c(z)$ on the scale of the figure.

The ray diagram [Fig. 5(b)] shows a subset of rays emanating from the source at depth 66 m based on $c_b(z)$. The solid lines are rays whose launch angles, i.e., θ_v, are between 5° and 7.6°; this angular range maps to θ_g between ~1° and 6°. The dashed lines are rays whose θ_v are less than 5°. Here, the sound field close to the sea surface begins to

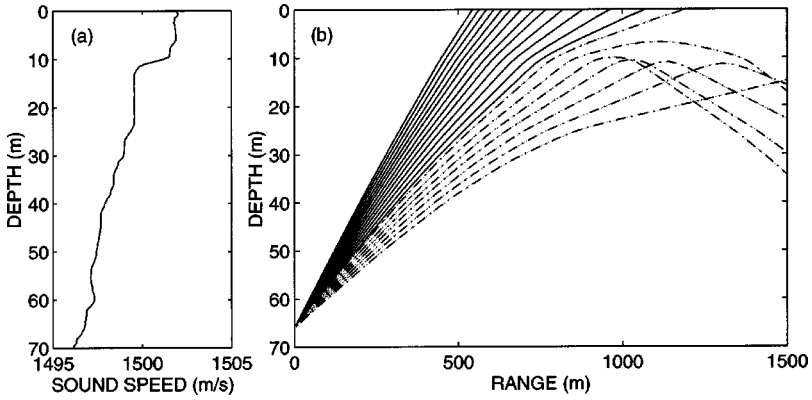


FIG. 5. (a) Sound speed versus depth profile $c(z)_b$ for the FLIP measurements on 24 January. (b) Ray diagram showing a subset of rays emanating from the source at depth 66 m, based on $c(z)_b$. The solid lines are rays with θ_v between 5° and 7.6° and the dashed lines are rays with θ_v less than 5°.

diverge with decreasing θ_v owing to refraction. With still smaller θ_v, rays undergo downward refraction, eventually producing a shadow zone and caustic region.

Figure 6 displays acoustic data for the measurements made at 1653 (a) and 1725 (b) similar to that shown in Fig. 2. These show calibrated reverberation level for one of the quad beams (black lines, right-hand scale), and corresponding split-beam estimates of vertical arrival angle (gray lines, left-hand scale), which are based on a transducer pitch angle of 2.5°. The estimated vertical arrival angle $\hat{\theta}_v$ and its relation to two-way travel delay time t is a key observable that

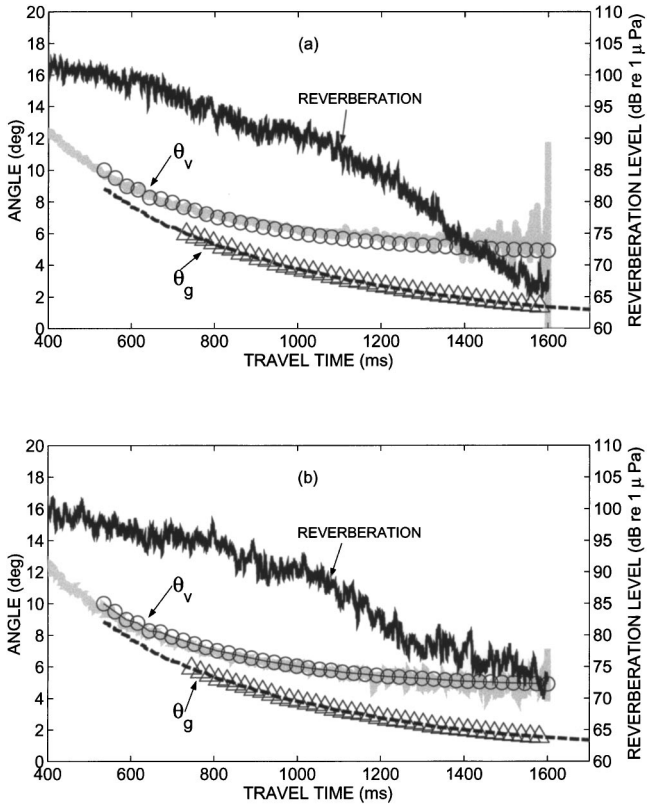


FIG. 6. (a) Averaged reverberation level based on 20 echoes (black line, right-hand scale) and estimated vertical arrival angle (gray line, left-hand scale) for FLIP measurement made 24 January at 1653. The line formed by circles is the expected vertical arrival angle (θ_v) based on ray theory with sound speed profile $c_b(z)$. The dashed line below it is the corresponding sea surface grazing angle (θ_g). The line of triangles identifies the segment of data used to compute sea surface backscattering. (b) The same for the FLIP measurement made 24 January at 1725.

can be compared with an equivalent estimate based on ray theory. This comparison $\theta_v(t)$, based on the $c_b(z)$ from Fig. 5(a), is shown by the line of circles that overlay the gray line in each plot. Associated with each $\theta_v(t)$ is, also by way of ray theory, a corresponding $\theta_g(t)$, shown by the lower dashed line in each plot. Given the satisfactory agreement between θ_v derived from ray theory and $\hat{\theta}_v$ estimated from the split beam observations, we proceed with $\theta_g(t)$ as our estimate of surface grazing angle to be associated with scattering strength. The line of triangles overlaying each dashed line shows the angular range of acceptable data for estimating scattering strength. In this case the upper angle of 6° is set by the aforementioned criterion for $10\log_{10}B(\theta_v,\theta_p)$, and the lower range is set by the signal to noise. For $\theta_g < 1°$ the reverberation level approaches the noise level (equivalent to a scattering strength of -67 dB) and the split beam phase estimate rapidly degrades, all consistent with the sound field diverging in excess of spherical spreading near the sea surface at ranges beyond about 1050 m, owing to downward refraction.

The ray calculations shown in Fig. 5(b) were performed using a generic program by the author and indicate transmission loss 1–2 dB greater than spherical spreading for ranges between 800 and 1000 m. As a check, transmission loss was also computed with the CASS (GRAB) program,[20] the results of which were consistent and ultimately incorporated into the sonar equation to compute scattering strength. Figure 7 shows estimates of scattering strength versus θ_g for the measurements made at 1653 (gray line and symbol) and 1725 (black line and symbol), based on a 20-echo-averaged reverberation level. The dashed line is the model result using Eq. (3) with $\beta_I = 0.004$, which, for comparison with the Florida results, puts scattering strength equal to -34.85 dB for $\theta_g = 20°$. Note that measurements made at 1725 extend down to a grazing angle of ~1.5°, whereas those made at 1653 stop at a slightly higher grazing angle because of noise contamination. Only Eq. (3), associated entirely with bubble effects, is used to model the data, as the Bragg scattering contribution at comparable wind speed is more than 40 dB lower than results produced either by Eq. (3) or seen in the actual data.

The agreement between model and data is better for grazing angles greater than about 3°, where the predicted scattering begins to level off with increasing grazing angle. But, as we have noted, the experimental results at low graz-

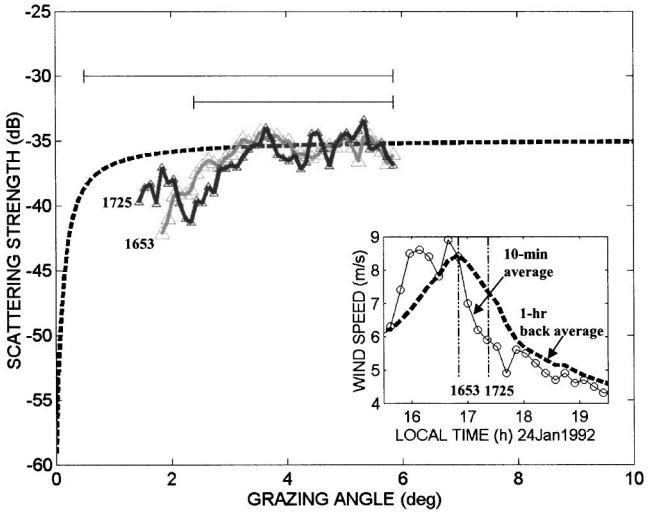


FIG. 7. Estimates of scattering strength versus θ_g for the measurements made 24 January at 1653 (gray line and symbol) and 1725 (black line and symbol). The dashed line is the model result using Eq. (3) with $\beta_I = 0.004$. The two horizontal bars show the overall uncertainty in the range of grazing angle. (Inset) Ten-minute-averaged wind speed (line and open circles), and its one-hour back-average (dashed line).

ing angle depend critically on the mapping between θ_g and θ_v. This mapping is, in turn, quite sensitive to small changes in sound speed near the surface, and uncertainties arise in establishing the smallest value of θ_g. One is the void fraction, which puts Δc in the range of -0.03 to -0.3 m/s, as discussed earlier. Another is the temperature change associated with spatial variation. For example, Plate 2 of Ferrari and Rudnick,[21] based on measurements made at the same time of year and within the same general region of the North Pacific, shows that within spatial scales of 1 km and at a fixed depth within the thermocline, the extent of horizontal variation in temperature can commonly be $\sim 0.2\,^{\circ}$C, which translates to a sound speed variation of ~ 1 m/s, or ± 0.5 m/s. The effect of changes in the near-surface sound speed on $\theta_g(t)$ is readily computed using Snell's law, given a reliable estimate of $\theta_v(t)$, i.e., $\cos\theta_v/c_L = \cos\theta_g/c_S$, where c_L is the sound speed at the source or launch depth, and c_S is the sound speed at the sea surface. This can be used to bound the uncertainty in sea surface grazing angle. Taking $c_L = 1496.72$ m/s, and $c_S = 1501.90$ m/s $+ 0.47$ m/s or -0.80 m/s, where the latter two values represent the nominal extremes in uncertainty based on Δc plus horizontal temperature variation (± 0.5 m/s), puts the smallest θ_g between about 0.5° and 2.4° for the smallest launch angle at the source of 5°. In this analysis we take the largest θ_g to be more reliably established at 6°, and further assume that any uncertainty in the transducer pitch angle is small compared to that introduced by the sound speed variation. The two horizontal bars above the data in Fig. 7 depict the two possible ranges in sea surface grazing angle. Were the data to be mapped to the angular range denoted by the longer bar, a model–data comparison for small grazing angles would improve markedly. The opposite happens, to some extent, for the shorter bar.

There was also a rather significant (~ 2 m/s) drop in wind speed during the period between the two measurements (see Fig. 7, inset) made in the early evening of 24 January. Yet, insofar as both experimental runs are modeled by the same β_I, we conclude that the depth-integrated bubble concentration was approximately the same during each measurement. Taking a 1-h back average of the wind speed time series substantially narrows the difference in the wind speed estimate associated with each acoustic measurement. But wind speed was also rising during the first measurement and falling during the second measurement based on the 1-h back average, which is suggestive of a hysteresis effect similar to that observed in the Florida data.

Presented next are results from an experimental run designed to measure sea surface backscattering at higher grazing angles. The run was made on 19 January at 1551, during which the wind speed was a steady 4.55 m/s (i.e., 10-min average and 1-h average are the same, but considered falling according to our definition), and the transducer was pointed downwind. The transducer pitch angle was set to 45°, and the pulse length was 1 ms, which put the angular resolution to approximately 1°. Note that in this case refraction effects are negligible and have no impact on the interpretation of the data because of the high grazing angles involved in the scattering.

Scattering at high grazing angles includes a significant contribution from the rough surface in addition to that from bubbles, and we hazard to identify these two components in the following manner. First, Plant's model[6] is used to create a starting vector of σ_r values (call this σ_{r0}) to correspond with grazing angles from the data, with the model input being a fully developed sea, and the same wind speed (4.55 m/s) and look direction. Test vectors of σ_r are then created by multiplying the σ_{r0} vector by factors uniformly distributed within the range of ± 5 dB (i.e., an arbitrary maximum offset of 5 dB is allowed in a test vector). Similarly, test vectors of σ_b are generated using Eq. (3) based on probable values of β_I. All combinations of the σ_r and σ_b test vectors are combined as in Eq. (4) to produce a model vector. We find that a minimum squared error between model and data vector is achieved when β_I equals 7.7×10^{-4} and when σ_r is offset from σ_{r0} by the relatively modest value of 2.5 dB. Three model curves are shown along with the data in Fig. 8. The horizontal dashed line is scattering strength based on bubble scattering alone, and the second dashed line is scattering strength based on rough surface scattering alone (i.e., σ_{r0} to which 2.5 dB has been added). The solid line is the model representing the incoherent sum of bubble and rough surface components as in Eq. (4). The sum model curve describes well the shape of the scattering strength versus the grazing angle relation suggested by the data. For example, the curve for rough surface scattering goes as $\sim \tan^4\theta_g$, as expected for Bragg scattering from ocean surface roughness at this Bragg scale, and the addition of the σ_b distorts this curve in a direction consistent with the data. Finally, for comparison with the Florida measurements, the β_I used in modeling gives a scattering strength at 20° equal to -42 dB.

V. DISCUSSION

More on the relation between our measurements of high-frequency sea surface backscattering and wind speed is shown in Fig. 9. Here, each of the 58 estimates of SS20 from

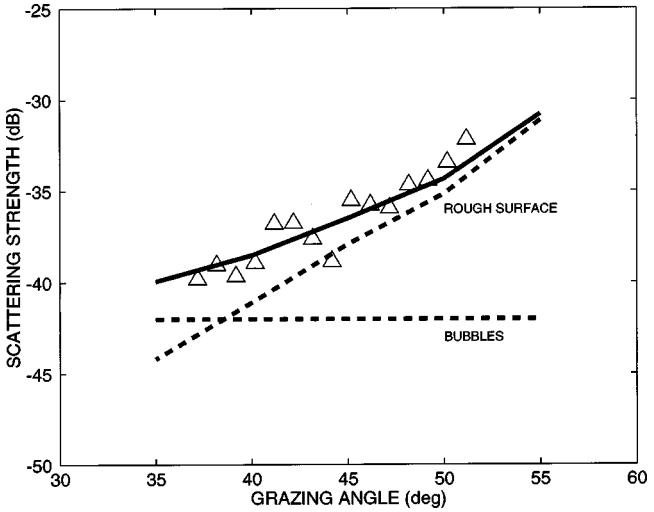


FIG. 8. Estimates of scattering strength versus θ_g for the measurements made 19 January at 1551 (10-min-averaged wind speed=4.55 m/s) compared with model curves. The horizontal dashed line (labeled BUBBLES) is the model result using Eq. (3) with $\beta_I = 7.7 \times 10^{-4}$, and the second dashed line (labeled ROUGH SURFACE) is scattering strength based on rough surface scattering alone to which 2.5 dB has been added. The solid line is a model based on the combination of the bubble and rough surface components using Eq. (4).

the 24-h Florida experiment are plotted against U_{10} according to whether the wind speed was rising (filled circle) or falling (open circle) prior to the acoustic measurement. The rising or falling property of the wind speed at a given time is determined by first smoothing the U_{10} time series over a 1-h time window extending backward from that time, then differentiating this result. The three estimates of SS20 inferred from the 1992 FLIP measurements made at grazing angles other than 20° are also plotted according to whether the wind speed was rising (filled triangle) or falling (open triangle) prior to the acoustic measurement. A fourth estimate associated with rising winds (filled square) is based on Eq. (3) with β_I equal to 0.06; this value originates from an estimate of the

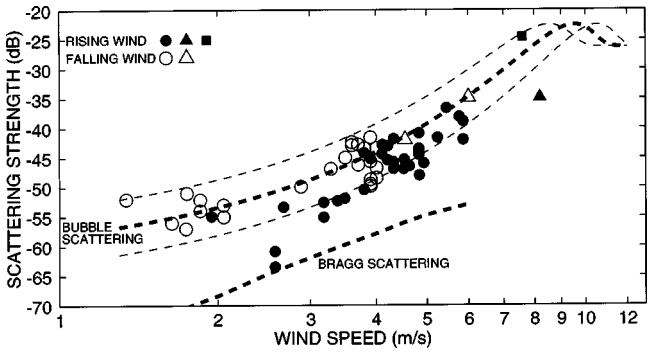


FIG. 9. Estimates of SS20 ($N=58$) from the 24-h Florida experiment plotted against U_{10} according to whether the wind speed was rising (filled circle) or falling (open circle) prior to the acoustic measurement. The three scattering-based measurements (triangle) and one attenaution-based measurement (square) from the 1992 FLIP experiment are also plotted using filled/open convention. Curve labeled BRAGG SCATTERING is $10 \log_{10} \sigma_r$, based on Plant's model (Ref. 18), and the curve labeled BUBBLE SCATTERING is $10 \log_{10} \sigma_b$, based on Eqs. (3) and (5); two model curves bracketing the central one result from plotting Eq. (3) versus U_{10}, but computing β_I with $U_{10}+1$ m/s (upper curve) and $U_{10}-1$ m/s (lower curve).

excess attenuation in the sea surface bounce path due to bubbles made during the same 1992 experiment.[22] Viewed this way, the data show a subtle hysteresis effect, i.e., the tendency of a slightly higher wind speed being associated with a given scattering level if the wind speed is rising. This is best seen in the large cluster of measurements from the 24-h experiment with levels between -50 and -40 dB, during which the wind speed was close to 4 m/s. The mean wind speed for the set of SS20 estimates that fall within -45 dB ± 5 dB and are associated with falling winds is 3.8 ± 0.09 m/s; the mean wind speed for the similar set of SS20 estimates but associated with rising winds is 4.6 ± 0.11 m/s.

The mechanism and further quantification of this hysteresis effect is beyond the scope of this paper, and a further analysis would require a much longer time series of measurements. It is quite plausible, however, that the hysteresis is, in part, linked to the persistence of bubbles created by the action of prior winds, e.g., as with the two FLIP measurements made on 24 January for which SS20 remained the same while U_{10} had fallen by about 2 m/s during the period between the measurements. Note that a recent study[23] demonstrated threshold and hysteresis effects in the growth and decay of wind waves in a wave tank setting, and determined a wind speed threshold for wind-wave growth and subsequent microwave scattering that is greater for rising winds than for falling winds. It is also plausible that a similar wind-wave growth effect influences the relation between wind speed and scattering for the acoustic case, in view of the sensitivity of high-frequency acoustic backscattering to dilute concentrations of bubbles created by sporadic, small scale wind-wave breaking, as shown by this study and Ref. 3.

The curve labeled BRAGG SCATTERING in Fig. 9 is $10 \log_{10} \sigma_r$ for 30 kHz frequency and $\theta_g = 20°$ based on Plant's model for microwave rough surface scattering. In view of the dominance of bubble scattering at high frequencies and low grazing angles as suggested in the data, we regard Plant's model as a lower bound for SS20 (though at higher grazing angles rough surface scattering does become significant, as demonstrated in Fig. 8). As discussed in the context of Fig. 4, scattering levels came reasonably close to this lower bound for only a brief period during the course of the 24 h, a period that did not, moreover, coincide precisely with the period of lowest wind speed. As noted earlier, the reason may have more to do with a momentary gap in the bubble field characterized by a high degree of patchiness. Likewise, the persistence of bubbles created by the action of prior winds was likely responsible for the maintenance of relatively high scattering levels at even lower wind speeds. We remark here that the possibility of bubbles introduced by ship heave and wave splash against the hull, and subsequently measured acoustically, is remote. However, the null effect from this cannot be proven. Still, during the frequent visual observations of the hull interaction zone near the stern throughout the 24-h period, significant bubble creation due to heave and wave-splash was not observed. Furthermore, the range to the scattering zone, approximately one ship length, would have provided some mitigation from this influence.

The curve labeled BUBBLE SCATTERING in Fig. 9 is

$10\log_{10}\sigma_b$, based on Eq. (3), with the following model for β_I:

$$\log_{10}\beta_I = -6.45 + 0.47U_{10} + 0.85\log_{10} f, \tag{5}$$

where f is frequency in kHz and U_{10} is 10-m height wind speed in m/s. This model was derived by the author prior to the availability of the Florida data discussed here, and was first published as part of Ref. 24. The model is based in large part on an empirical fit to scattering measurements made at frequencies between 15 and 60 kHz and grazing angles between 5° and 30° that originate from an experiment conducted in 1977 by Lilly and McConnell in pelagic waters 130 km west of San Diego.[25] The two model curves bracketing the central one result from plotting Eq. (3) versus U_{10}, but computing β_I with Eq. (5) while using $U_{10}+1$ m/s (upper curve) and $U_{10}-1$ m/s (lower curve). When Eq. (5) is applied to Eq. (3), the resultant curve for bubble scattering displays three wind speed regimes: (1) sensitivity to wind speed, (2) transition to saturation for which scattering is maximum, and (3) the saturation and subsequent insensitivity to a further increase in wind speed.[2,3]

The sensitivity of $10\log_{10}\sigma_b$ to wind speed is determined by the relation between β_I and wind speed. Among the candidate models for β_I are a power law in the form of $\beta_I \sim U^p$, and an exponential law in the form of $\beta_I \sim e^{pU}$, where U is some averaged value of wind speed. The 1977 field measurements, insofar as they are represented by Eq. (5), display an exponential sensitivity to wind speed, and we believe this property is repeated in the new measurements presented here. The differences are nonetheless subtle. For example, the relation between SS20 and U_{10} (exponential law), evaluated using orthogonal regression, gives a correlation coefficient of $r=0.82$, with β_I going very nearly as e^U. The resultant curve spanning the range of measurements would be approximately 1.5 dB below, but parallel to, the curve based on Eq. (5) as it, too, goes very nearly as e^U. The relation between SS20 and $\log U_{10}$ (power law) gives a correlation coefficient of $r=0.80$, with β_I going as $U^{3.3}$. In both evaluations we restricted the data to only that from the Florida experiment, and also excluded the two lowest estimates of SS20 that we argue are associated with Bragg rather than bubble scattering. Note that if we include only measurements for which U_{10} exceeded 3 m/s, the resultant power law changes to $\beta_I \sim U^{5.2}$. This is closer to the power law ($U^{5.5}$) derived from measurements over a similar range of wind speeds made in the North Sea at a frequency of 26.5 kHz.[3] The empirical model for β_I from that same work puts σ_b, on average, about 2.5 dB greater than that predicted by Eq. (5), although within the +1 m/s bound shown by the upper curve in Fig. 9. Proximity to land has been proposed as a systematic influence on bubble scattering, with coastal waters producing higher scattering levels than pelagic waters for a given wind speed.[2] The new data presented here do not support this hypothesis, given that the curve representing the Florida coastal measurements is 1.5 dB less than the curve representing the 1977 data taken in pelagic waters. Furthermore, there are other sources of systematic variability that can produce similar effects, such as that observed in this study pertaining to the effect of rising or falling winds.

In evaluating the relation between SS20 and wind speed using variable-length back-averages of U_{10}, we find that the correlation reaches a maximum $r=0.86$ for a 1-h back-average. This result mirrors that of Nicholas *et al.*,[26] who studied low-frequency ($\leqslant$1 kHz) backscattering data from the Critical Sea Test (CST) experiments. To study the influence of processes driven by the wind history, they varied the back-averaging time and found the error between the CST data and their empirical model, applicable to frequencies between 50 and 1000 Hz, was minimized for 1-h back-averaging of the instantaneous wind speed. At frequencies of $O(1)$ kHz or less, bubbles contribute to and can even dominate the scattering, yet Bragg scattering from the rough air–sea interface also plays a much more important role than it did in our case for which the frequency was of $O(10)$ kHz. Furthermore, theoretical models for low-frequency backscattering from bubbles involve coherent scatter from bubble plumes (e.g., Ref. 27) in contrast to the incoherent scatter that is the basis of Eq. (3). Still, common features are expected in an empirical relation between acoustic scattering and wind speed in regards to the influence of wind history on the concentration of near-surface bubbles.

In terms of surface wave field measures during the Florida experiment [Figs. 3(b)–(c)], the significant wave height remained between 20 and 30 cm up to about hour 20 and then slowly rose to about 60 cm during the next 5 hours; as would be expected, the wave height alone is a poor correlation to the scattering. The wave slope, however, does reflect the major up/down swings (of time scale $\approx$5 h) seen in both time series of wind speed and SS20, yet its simple correlation with SS20, in either logrithmic or linear forms, is poor ($r<0.4$). Note that the square of the wave slope is equal to the fourth moment of the wave spectrum m_4, and simultaneous field observations of wave breaking probability and m_4[28] agree with a predicted dependence of the breaking probability on m_4.[29] A drawback of the m_4 estimation, however, is its sensitivity to high-frequency cutoff, as set in our case by the highest frequency measured by the wave buoy (0.58 Hz). Another candidate predictor of SS20 that combines wind and wave measures is a wind forcing parameter (or reciprocal of wave age), defined as U_{10}/c_p, where c_p is a characteristic phase speed of the wind-driven waves. For this we set $c_p = 1.56/f_p$, where f_p is the peak frequency of the wind-wave system shown plotted by a star symbol for each wave spectral density displayed in Fig. 3(b). The wave forcing parameter varies from about 0.4 to 1.5 over the 24 h, and although it correlates reasonably with SS20 ($r=0.76$), it is by no means superior to wind speed alone as a correlation to SS20. Of potential significance in our case is the fact that the swell system largely opposed the wind-wave system until about hour 22, as shown in Fig. 3(c), at which point both systems become more closely aligned with the wind. There is some evidence that the presence of swell propagating against the wind leads to a larger drag coefficient than encountered in a pure wind–sea case.[30] Given this condition we would anticipate higher scattering levels for the counterpropagating case, owing to enhanced wave breaking, than if all waves

were aligned. We do not, however, have enough measurements taken under different wind-wave and swell alignment conditions to further evaluate this effect on scattering.

VI. SUMMARY

Measurements of sea surface backscattering, wind speed, and surface wave spectra, made continually over a 24-h period, are presented from an experiment conducted in 26 m of water near the Dry Tortugus collection of islands off south Florida in February 1995. Scattering strength measurements were made at frequency of 30 kHz and sea surface grazing angle of 20°, and a time series of this variable, called SS20, was studied in terms of its dependence on environmental variables. On occasion, reliable estimates of scattering in the grazing range 15°–27° were also obtained. The acoustic scattering data exhibited evidence, in terms of scattering level and grazing angle dependence, of scattering from near-surface bubbles rather than scattering from the rough air–sea interface.

The acoustic measurements were compared with the decibel equivalent of a model for σ_b, or the apparent backscattering cross section per unit area due to bubble scattering. This model is driven by β_I, a parameter characterizing the bubble concentration in terms of a depth-integrated extinction cross section per unit volume. Expressions for β_I must necessarily be derived empirically from field data, and the one for β_I used here is based largely on data from a 1977 experiment conducted in pelagic waters. It has β_I increasing exponentially with wind speed rather than as a power law of wind speed, a relation that is also supported by the more recent Florida measurements.

Four additional model-data comparisons were made using data from an experiment conducted in 1992, also in pelagic waters from the research platform FLIP. In this case, the estimates of SS20 were inferred from estimates of β_I that were, in turn, derived from measurements of backscattering at grazing angles other than 20° (three cases) and from a single measurement of attenuation in forward scattering. Problems inherent to obtaining reliable estimates of backscattering at very low grazing angles in waters for which the sound speed varies with depth were also discussed in the context of the 1992 backscattering measurements.

Unambiguous observations of the contribution of rough surface scattering were rare in the 30-kHz measurements, due to the dominant contribution of bubble scattering once the wind speed exceeded a threshold value between 2 and 3 m/s. This is likely to be true for the more general case of frequencies in the $O(10-100)$ kHz range and grazing angles $\lesssim$30°. Two cases were discussed for which evidence of rough surface scattering was exhibited in the data: one involving very low wind speed and one involving grazing angles greater than 30°. For these cases a model for microwave scattering from the ocean surface by Plant[6] was used to compute the backscattering cross section per unit area due to scattering from the rough air/sea interface.

The 1992 and 1995 data display reasonable consistency with the model for σ_b as driven by β_I, and the larger 1995 dataset (excluding the two lowest values argued to be associated with Bragg scattering) is, on average, only 1.5 dB less than that predicted by this model. The standard deviation of model-data error vector equals 3 dB when 5-min-averaged wind speed is used as model input, and this value reduces to 2.5 dB when a 60-min back-average of wind speed is used as model input. Model uncertainty, defined as the largest difference between model prediction and measured data, is 5 dB (again excluding the two lowest estimates of SS20). The $\pm$5-dB spread for a given wind speed is very close to the envelope that results from computing $10\log_{10}\sigma_b$ using a $\pm$1-m/s spread about a given wind speed. The 24-h time series also exhibits a hysteresis effect, wherein for a given wind speed there is a tendency for the scattering level to be higher if prior winds have been falling. A better understanding of this effect is essential in order to reduce uncertainty in model predictions.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research Code 321 Ocean Acoustics, via Contract No. N00039-91-C-0072. The author is grateful to William Plant for ongoing discussions on rough surface scattering and for providing computations from his rough surface scattering model, and to Andrew Jessup for providing thermistor chain data from the 1992 field experiment. The helpful suggestions from the three anonymous reviewers are also appreciated.

[1] C. S. Clay and H. Medwin, "High-frequency acoustical reverberation from a rough-sea surface," J. Acoust. Soc. Am. **36**, 2131–2134 (1964).
[2] S. T. McDaniel, "Sea surface reverberation: A review," J. Acoust. Soc. Am. **94**, 1905–1922 (1993).
[3] P. H. Dahl, W. J. Plant, B. Nützel, A. Schmidt, H. Herwig, and E. A. Terray, "Simultaneous acoustic and microwave backscattering from the sea surface," J. Acoust. Soc. Am. **101**, 2583–2595 (1997).
[4] W. J. Plant, P. H. Dahl, and W. Keller, "Microwave and acoustic scattering from parasitic capillary waves," J. Geophys. Res. **104**, 25853–25865 (1999). In this study, the experimentally measured microwave cross section per unit area for horizontal polarization, or $\sigma^{\circ}_{\mathrm{HH}}$, was found to be the same (within experimental uncertainty) as the acoustic cross section, or σ°_{A}, when Bragg wavelengths are equal, and when account is made for the fresh water dielectric constant, the microwave polarization ratio, and different conventions for cross section. The same levels of scattering were observed *only* when the orientation of the microwave and acoustic antenna were opposite (viz., upwind and downwind), owing to the effects of a mean tilt angle and scattering from bound, capillary waves that reside preferentially on the forward face of the wave and dominate the rough-surface scatter for short-fetch experimental conditions.
[5] P. C. Wille and D. Geyer, "Measurements on the origin of the wind-dependent ambient noise variability in shallow water," J. Acoust. Soc. Am. **75**, 173–185 (1984).
[6] W. J. Plant, "A stochastic, multiscale model of microwave backscatter from the ocean," J. Geophys. Res. **107**, 10.129/2001 JC000909, 2002.
[7] P. H. Dahl, "On bistatic sea surface scattering: Field measurements and modeling," J. Acoust. Soc. Am. **105**, 2155–2169 (1999).
[8] D. R. Jackson, A. M. Baird, J. J. Crisp, and P. A. G. Thomson, "High-frequency bottom backscatter measurements in shallow water," J. Acoust. Soc. Am. **80**, 1188–1199 (1986).
[9] P. H. Dahl and W. J. Plant, "The variability of high-frequency acoustic backscatter from the region near the sea surface," J. Acoust. Soc. Am. **101**, 2596–2602 (1997).
[10] S. D. Smith, "Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature," J. Geophys. Res. **93**, 15467–15472 (1988).
[11] W. J. Plant, "Effects of wind variability on scatterometry at low wind speeds," J. Geophys. Res. **105**, 16889–16910 (2000).
[12] P. A. Crowther, "Acoustical scattering from near-surface bubble layers," in *Cavitation and Inhomogeneities in Underwater Acoustics*, edited by W. Lauterborn (Springer-Verlag, New York, 1980), pp. 194–204.

[13] K. Sakar and A. Prosperetti, ''Coherent and incoherent scattering from oceanic bubbles,'' J. Acoust. Soc. Am. **96**, 332–341 (1994).
[14] G. Kapodistrias and P. H. Dahl, ''On scattering from a single bubble near a pressure release interface: Laboratory measurements and modeling,'' J. Acoust. Soc. Am. **110**, 1271–1281 (2001).
[15] P. H. Dahl and G. Kapodistrias, ''Scattering from a single bubble near a roughened air-water interface: Laboratory measurements and modeling,'' J. Acoust. Soc. Am. **113**, 94–101 (2003).
[16] P. H. Dahl and A. T. Jessup, ''On bubble clouds produced by breaking waves: An event analysis of ocean acoustic measurements,'' J. Geophys. Res. **100**, 5007–5020 (1995).
[17] P. H. Dahl, ''On the spatial coherence and angular spreading of sound forward scattered from the sea surface: Measurements and interpretive model,'' J. Acoust. Soc. Am. **100**, 748–758 (1996).
[18] A. T. Jessup and V. Hesany, ''Modulation of ocean skin temperature by swell waves,'' J. Geophys. Res. **101**, 6501–6511 (1996).
[19] K. W. Commander and A. Prosperetti, ''Linear pressure waves in bubbly liquids: Comparison between theory and experiments,'' J. Acoust. Soc. Am. **85**, 732–746 (1989).
[20] H. Weinberg and R. H. Keenan, ''Gaussian ray bundles for modeling high-frequency propagation loss under shallow-water conditions,'' J. Acoust. Soc. Am. **100**, 1421–1431 (1996).
[21] R. Ferrari and D. L. Rudnick, ''Thermohaline variability in the upper ocean,'' J. Geophys. Res. **105**, 16857–16883 (2000).
[22] P. H. Dahl, ''High-frequency forward scattering from the sea surface: The characteristic scales of time and angle spreading,'' IEEE J. Ocean. Eng. **26**, 141–151 (2001).
[23] M. A. Donelan and W. J. Plant, ''Threshold and hysteresis effects in wind wave growth and decay,'' submitted to J. Phys. Ocean., 2002.
[24] APL-UW author team, *APL-UW High-Frequency Ocean Environmental Acoustic Models Handbook*, Tech. Rep. APL-UW 9407, Applied Physics Laboratory, University of Washington, October, 1994.
[25] S. O. McConnell and J. C. Lilly, *Surface Reverberation and Ambient Noise Measured in the Open Ocean and Dabob Bay (U)*, Tech. Rep. APL-UW 7727, Applied Physics Laboratory, University of Washington, January, 1978.
[26] M. Nicholas, P. M. Ogden, and F. T. Erskine, ''Improved empirical descriptions for acoustic surface backscatter in the ocean,'' IEEE J. Ocean. Eng. **23**, 81–95 (1998).
[27] F. S. Henyey, ''Acoustic scattering from ocean microbubble plumes in the 100 Hz to 2 kHz region,'' J. Acoust. Soc. Am. **90**, 399–405 (1991).
[28] L. Ding and D. M. Farmer, ''Observations of breaking surface wave statistics,'' J. Phys. Oceanogr. **24**, 1368–1387 (1994).
[29] R. L. Synder and R. M. Kennedy, ''On the formation of whitecaps by a threshold mechanism. Part I: Basic formulation,'' J. Phys. Oceanogr. **13**, 1482–1492 (1983).
[30] M. A. Donelan, W. M. Drennan, and K. B. Katsaros, ''The air–sea momentum flux in conditions of wind sea and swell,'' J. Phys. Oceanogr. **27**, 2087–2099 (1997).

Low frequency coupled mode sound propagation over a continental shelf

D. P. Knobles,[a] S. A. Stotts, and R. A. Koch
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713

(Received 1 January 2002; revised 12 August 2002; accepted 23 September 2002)

A two-way integral equation coupled mode method is applied to a continental shelf ocean waveguide proposed for a special session devoted to range-dependent acoustic modeling at the 141st meeting of the Acoustical Society of America. The coupled mode solution includes both sediment trapped and continuum modes. The continuum is approximated by a finite number of leaky modes but neglects the branch cut contribution. Mode coupling matrix elements and the range evolution of the modal amplitudes show the nature of the mode coupling. Transmission loss versus range at 100 Hz predicted by the integral equation approach is compared to the transmission loss predicted by a wide angle parabolic equation method. While there is very good agreement, one observes small differences that can be interpreted as backscattering predicted by the integral equation solution. [DOI: 10.1121/1.1534847]

PACS numbers: 43.30.Bp, 43.30.Gv [DLB]

I. INTRODUCTION

Modeling sound propagation in continental shelf environments represented by idealized waveguides has received considerable attention in the ocean acoustics community.[1–4] Most acoustic computations with the effect of a penetrable seabed and a sloped bathymetry assume a horizontally stratified water column and a simple halfspace representing the seabed. However, realistic continental shelf environments often have a range-dependent sound speed profile, and the seabed has sediment layers over the bottom halfspace. Significant physical insight into the continental shelf problem is gained by expressing the acoustic field in terms of the local normal modes of the waveguide and examining the range evolution of individual modal wave number components and their couplings. Although the basic coupled-mode equations for the modal amplitudes have been established in differental[5–7] and integral forms,[8–10] implementing a consistent approach for realistic waveguide representations remains difficult.

The focus of this paper is the application of a two-way energy conserving coupled mode integral equation approach[4–6] to the continental shelf ocean waveguide proposed for a special session devoted to range-dependent acoustic modeling at the 141st meeting of the Acoustical Society of America (ASA). Attention is devoted to the nature of the mode coupling. The numerical computation is an attempt to establish a benchmark quality result for a case where mode coupling occurs over a large number of wavelengths and a small amount of backscattering is present. Section II defines the continental shelf problem. Section III briefly discusses the integral equation method and several issues associated with its application to the shelf problem. Section IV presents the numerical results.

[a]Electronic mail: knobles@arlut.utexas.edu

II. DESCRIPTION OF CONTINENTAL SHELF PROPAGATION PROBLEM

Figure 1 illustrates the bathymetry and sound speed structure of the continental shelf waveguide. Starting from the source, the waveguide has a constant depth of 400 m out to 2 km in range. Beyond 2 km the water depth decreases to 100 m over a 15 km range interval, making the bottom slope angle, $\theta_B = \tan^{-1}\ 0.02 = 1.146°$. For comparison the ASA benchmark wedge has a slope of 2.862°.[2] The water depth then remains constant at 100 m for an additional 3 km. The water column has a density of 1 g/cm^3 and no attenuation. The analytic expression for the sound speed profile (SSP) is

$$c(r,z) = 1450 + 4.6T(r,z) - 0.055T^2(r,z) + 0.016z, \tag{2.1}$$

$$T(r,z) = 5 + T_0(r)[\sinh(\pi(400-z)/200)/265]^2, \tag{2.2}$$

$$T_0(r) = 15 - 2.5[1 + \tanh(r-10)], \tag{2.3}$$

where z is in meters, and r is in kilometers. The SSP is essentially range-independent out to about 8 km. The largest range-dependence occurs between 9.5 and 10.5 km. After about 12 km the SSP again becomes approximately range-independent. Figure 2 shows on an expanded scale the sound speed range variation versus depth over the 0–20 km interval which again emphasizes the region (8–12 km) where the largest variation occurs. The maximum variation occurs at the surface of the water column with a range gradient of about 0.003 s^{-1}. The sound speed variation with range decreases with depth. The minimum in the sound speed is near the 150 m depth point, and for depths greater than this little variation in the profile occurs with range.

The seabed is comprised of a 50-m-thick sediment layer with a top sound speed of 2000 m/s and a linear increase with depth to 2500 m/s at the bottom of the layer. The density and the attenuation in the sediment are constant at 1.5 g/cm^3 and 0.1 dB/λ, respectively. The homogeneous halfspace below the sediment has a sound speed, density, and attenuation of 2500 m/s, 1.5 g/cm^3, and 0.1 dB/λ, respec-

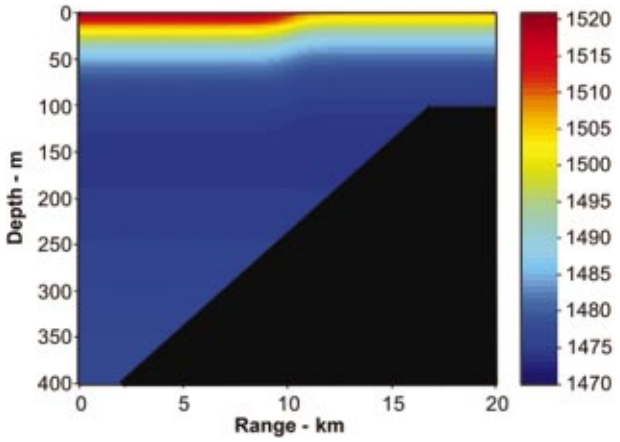


FIG. 1. Sound speed structure and bathymetry for continental shelf environment.

tively. For comparison the ASA benchmark wedge with a penetrable lossy bottom has a seabed consisting of a single halfspace with a sound speed, density, and attenuation of 1700 m/s, 1.5 g/cm^3, and 0.5 dB/λ, respectively.[2] Two aspects of the proposed continental shelf bottom do not represent typical ocean seabeds. First, the density and sound speed are not physically correlated. Second, a sound speed gradient of 10 s^{-1} is about an order of magnitude too large for rock. Nevertheless, this seabed is of interest because it allows for both trapped modes in the sediment, continuum modes, and the potential of backscattered energy because of the large sound speeds in the sediment.

III. APPLICATION OF INTEGRAL EQUATION FORMALISM TO CONTINENTAL SHELF PROBLEM

Since the integral equation formalism has been discussed in Refs. 8–10, we first briefly review the integral equation form of the coupled mode equations and then discuss several aspects of the formalism for the continental shelf problem.

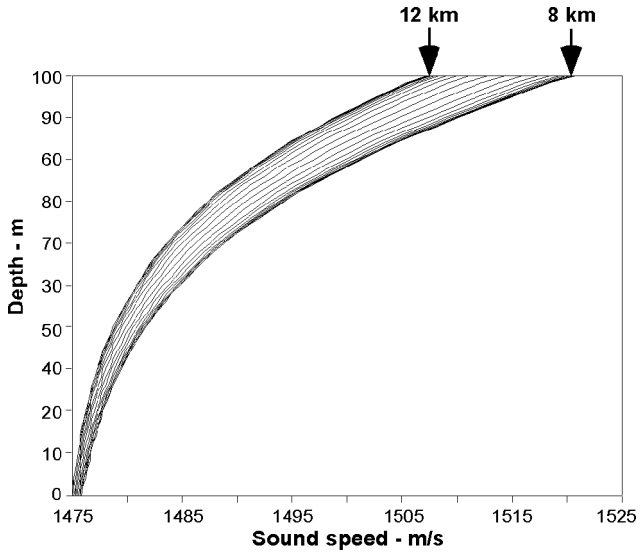


FIG. 2. Range variation of sound speed profile.

The equation solved in this analysis is the nonseparable Helmholtz equation for a cylindrically symmetric waveguide with a point source at $r=0$. The solution for the pressure, P, has the form

$$P(r,z)=\sum_{i=1}^{N}\Phi_i(z,r)R_i(r), \tag{3.1}$$

where the $\Phi_i(z,r)$ are local depth-dependent mode functions at range r, $R_i(r)$ is the solution to the coupled integral equation for the ith modal amplitude, and N is the total number of modes. The coupled integro-differential equation for the modal amplitudes is

$$R_n(r,z_0)=\int_0^\infty g_n(r,r')\Theta_n(r',z_0)r'\,dr' + \int_0^\infty g_n(r,r')\sum_{n'} C_{nn'}(r')R_{n'}(r')r'\,dr', \tag{3.2}$$

where $g_n(r,r')$ and $C_{nn'}(r')$ are the coordinate representations of the adiabatic Green's function and mode-coupling matrix operators, respectively. $\Theta_n(r,z_0)$ is the source function and z_0 is the depth of the source. The adiabatic Green's function for the mth modal component satisfies

$$\left\{\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}+k_m^2(r)\right\}g_m(r,r')=-\frac{\delta(r-r')}{r}. \tag{3.3}$$

If $y_m^{(0)}$ and $y_m^{(\infty)}$ are independent solutions to the homogeneous equation for g_m satisfying the boundary conditions at $r=0$ and $r=\infty$, respectively, then

$$g_m(r,r')=-\frac{y_m^{(0)}(k_m r_<)y_m^{(\infty)}(k_m r_>)}{r'W[y_m^{(0)},y_m^{(\infty)}]_{r=r'}}, \tag{3.4}$$

where $r_<=\min(r,r')$ and $r_>=\max(r,r')$. The Wronskian is $W[a,b]=a[(d/dr)b]-b[(d/dr)a]$. The boundary conditions are $g_m(r,r')$ is finite at $r=0$, and $g_m(r,r')$ has the form of an outgoing wave as $r\to\infty$. The components of the $N\times N$ coupling matrix are

$$C_{mn}(r)=A_{mn}+\alpha_{mn}+\frac{B_{mn}}{r}+(2B_{mn}+\beta_{mn})\frac{d}{dr}, \tag{3.5}$$

where

$$B_{mn}(r)=\int_0^\infty \frac{1}{\rho(z)}\phi_m(z,r)\frac{d}{dr}\phi_n(z,r)dz,$$
$$A_{mn}(r)=\int_0^\infty \frac{1}{\rho(z)}\phi_m(z,r)\frac{d^2}{dr^2}\phi_n(z,r)dz, \tag{3.6}$$

and α and β are the interface coupling matrices that result from sloping interfaces.[6]

Transmission loss is defined in the standard manner,

$$\mathrm{TL}(r,z,z_0) = -10\log_{10}\frac{P(r,z,z_0)P^*(r,z,z_0)}{\text{Intensity one meter from the source}}, \tag{3.7}$$

where the asterisk denotes complex conjugate.

As previously discussed in Refs. 8 and 9, the basic coupled integral equations are first transformed to an auxiliary set of coupled equations. The auxiliary equations are solved rather easily using the Lanczos method. The solution to the original equations are then recovered from the transformation. The Lanczos method is well suited for problems where the coupling occurs over many wavelengths. Unlike the Born series, each term in the Lanczos series includes an infinite series of forward and backscattered contributions which allows for rapid convergence.

In a realistic continental shelf environment the difficulties in a successful application of the integral equation approach to produce benchmark quality results are associated with the treatment of the modal continuum. A normal mode approach to the penetrable wedge problem introduces difficulties in dealing with the modal continuum associated with the lower homogeneous halfspace. Jensen and Ferla,[2] in the application of the Evans two-way stair step algorithm,[7] discretized the continuum by introducing a rigid bottom boundary condition about 50 wavelengths into the top of the bottom halfspace. In the limit that the depth of the rigid boundary condition approaches infinity, the continuum representation becomes exact. Such a representation is equivalent to an approach that includes both a leaky and branch line contribution to the continuum.[11] For the ASA benchmark penetrable wedge problem at 25 Hz there are 3 trapped modes at the base of the wedge. With the false bottom an additional 83 continuum modes were employed to represent the continuum.[2,10,15] Thus, the dimension of the coupled mode problem at 25 Hz was 86. If one scales the ASA benchmark to the present problem where the frequency is 100 Hz and the depth of the water column at the base of the slope is 400 m, the utilization of a false bottom would result in approximately 1250 modes. The numerical solution to a two-way coupled mode problem of dimension 1250 where the propagation interval is on the order of 1000 wavelengths, requiring about 15 000 range mesh points, is not feasible for typical high speed computers. In an actual application it may be possible to make a benchmark quality calculation using a fewer number of discrete modes to represent the continuum; however, such a computation still faces the difficulty of the coupling of hundreds of modes over a large range interval.

An alternative to the false bottom approach is to employ the Pekeris cut[12] to represent the continuum by a discrete set of leaky modes and to neglect the Pekeris branch line integral. As the water depth varies on the continental slope, the phenomena of mode cutoff requires attention because significant errors in the mode wave function can occur. An analogous problem occurs in broadband problems when a mode can pass through mode cutoff as a result of varying the propagation frequency. Westwood and Koch[13] addressed the mode cutoff difficulty for broadband acoustic problems in horizontally stratified waveguides by introducing a small complex sound speed gradient in the lower homogeneous halfspace. This had the beneficial effect of eliminating the branch point and making the transition of a leaky mode past the former position of the branch point continuous. The ability of computing a correct mode function as a mode passes through cutoff is generally made more robust by this approach. The introduction of the gradient also generates a set of discrete modes that represents the branch cut contribution. Finally, an additional benefit of the complex halfspace gradient is the leaky modes become physical in the sense they become bounded. This allows one to include mode coupling with the continuum in a coupled mode computation. It was observed for the ASA penetrable wedge problem that the branch line modes made almost no contribution to either the adiabatic calculation[14] or the coupled-mode calculation.[15] In this analysis the branch line modes are neglected.

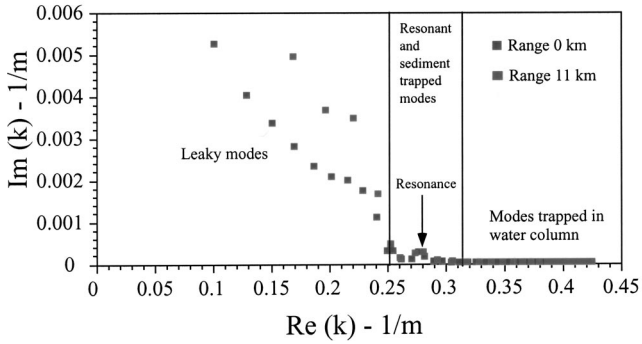


FIG. 3. Complex horizontal wave number eigenvalue spectra at ranges of 0 and 11 km.

IV. NUMERICAL RESULTS

In Sec. IV the results of the numerical computations at a frequency of 100 Hz are presented. First, the eigenvalue structure of the waveguide is presented. Next, parts of the mode coupling matrix elements are shown for several range points. The modal amplitudes as a function of range are also shown and are discussed in terms of the mode coupling matrix elements. Finally, the predicted transmission losses from the adiabatic approximation, the integral equation method, and a parabolic equation approach are compared over selected range intervals.

A. Eigenvalue structure

Figure 3 shows the modal spectrum in the complex wave number plane at ranges of 0 and 11 km. Not shown in the wave number diagram are the branch line modes. The number of modes, N in Eq. (3.1), is 56 at 0 km and is 33 at 11 km. The real part of the wave number, k_r, is 0.314 m^{-1} at the top of the sediment and 0.251 m^{-1} at the bottom of the sediment. These values are indicated by the two lines parallel to the imaginary k axis. For $k_r > 0.251\ \mathrm{m}^{-1}$, the modes lie along a single trajectory in the complex wave number plane. As the range increases the density of modes decreases and the modes move continuously along this trajectory. For $k_r < 0.251\ \mathrm{m}^{-1}$ the modes lie along a trajectory; however, the shape of this trajectory changes with range. On the basis of these values and the modal trajectories, one can define several groups of modes. First is the group of modes with $k_r > 0.314\ \mathrm{m}^{-1}$. These horizontal wave number eigenvalues have small imaginary components and correspond to modes trapped in the water column. A second group of modes with $k_r < 0.251\ \mathrm{m}^{-1}$ contains the leaky modes, and represents the

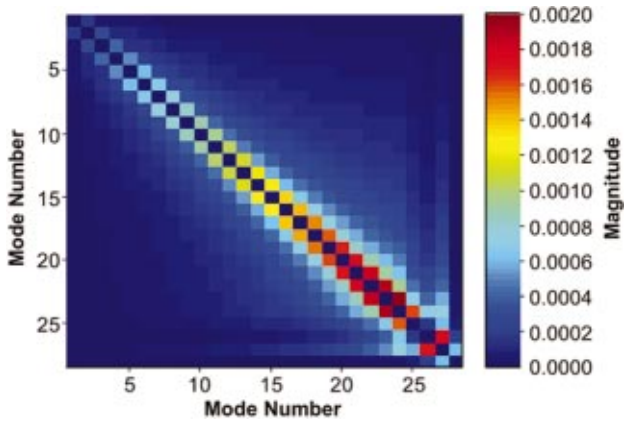


FIG. 4. First-order mode coupling matrix elements for trapped spectrum.

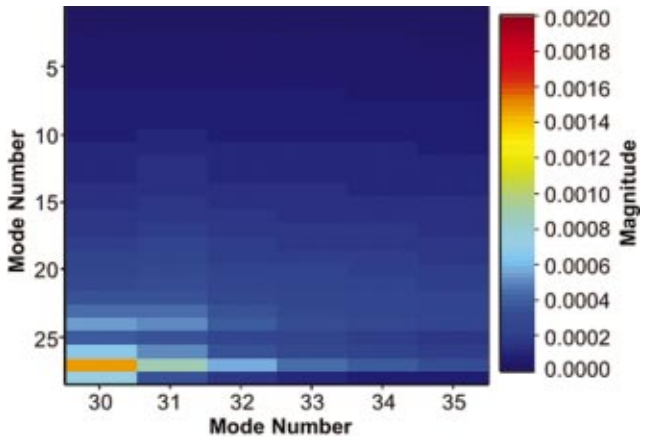


FIG. 5. First-order mode coupling matrix elements for trapped-continuum spectrum.

modal continuum associated with the lower halfspace. These modes are characterized by large imaginary components. In the region $0.251\ \mathrm{m}^{-1}<k_r<0.314\ \mathrm{m}^{-1}$ one observes several features. The most prominent feature is a resonance whose peak is located at approximately $k_r=0.278\ \mathrm{m}^{-1}$. The width of the resonance is approximately $0.01\ \mathrm{m}^{-1}$. On the ''right side'' of the resonance in the region $0.260\ \mathrm{m}^{-1}<k_r<0.273\ \mathrm{m}^{-1}$ and on the ''left side'' of the resonance in the region $0.285\ \mathrm{m}^{-1}<k_r<0.314\ \mathrm{m}^{-1}$ the modes penetrate into the sediment with an oscillatory wave function and have the same character as the modes in the first group with the exception that their turning points are in the sediment. In the region of the resonance $0.273\ \mathrm{m}^{-1}<k_r<0.283\ \mathrm{m}^{-1}$ an examination of these modes shows they have large amplitudes within the sediment as compared to the water column. For this reason one may refer to these modes as ''resonant modes.'' Finally, modes that have $k_r=0.251\ \mathrm{m}^{-1}+\delta k$, where δk is less than $0.01\ \mathrm{m}^{-1}$, exhibit a behavior similar to the resonant modes. These are modes that are near mode cutoff.

B. Mode coupling matrix elements and modal amplitudes

Figure 4 shows the magnitude of the first-order coupling matrix, B_{mn}, within the trapped spectrum at a range of 10 km. The term ''trapped'' refers to the portion of the modal spectrum such that $k_r>0.251\ \mathrm{m}^{-1}$. The largest coupling, and therefore the greatest energy exchange, occurs between the nearest-neighbor modes. This energy exchange increases continuously from modes 1 and 2 to the highest trapped pairs which have the real part of their eigenvalues less than $0.314\ \mathrm{m}^{-1}$. For instance, at 10 km the strongest coupling is for modes 23 and 24. Note also the diagonal coupling elements of B_{mn} are much weaker than the off-diagonal terms. The coupling of the trapped-continuum modes at this range is shown in Fig. 5. Their contribution is much weaker except for the coupling from the last resonance mode to the first continuum mode, which is considerable. Thus, it is demonstrated that considerable energy exchange is made from the trapped modes to the continuum via these resonance modes.[16]

Second-order coupling effects (the A_{mn} matrix) were observed to be at least two orders of magnitude weaker than the B_{mn} first-order terms. The strongest coupling contribution from these elements corresponds to the diagonal terms. Similar behavior is seen throughout the waveguide except that the mode numbers which give the strongest contribution decrease with decreasing water depth. Finally it was found that the boundary contributions to the mode coupling terms were as small or smaller than the second-order terms.

Figure 6 shows the magnitude of the modal amplitudes multiplied by $\sqrt{r}$ to compensate for cylindrical spreading. Out to 2 km, one observes what one would expect for an adiabatic case. Namely, there is no energy exchange between modes. However, starting at 2 km one observes nearest-neighbor mode coupling. For example, several neighboring mode pairs show energy being exchanged between modes. One observes that the range interval for a single cycle of

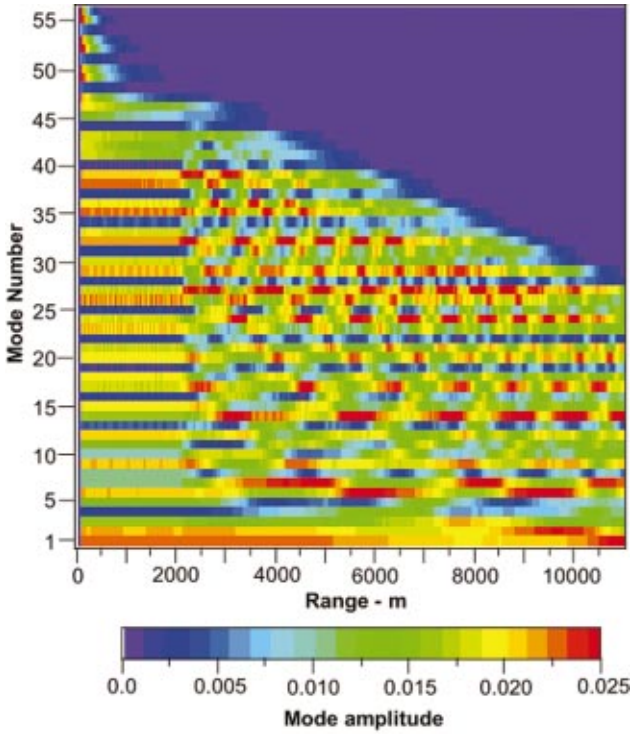


FIG. 6. Range variation of modal amplitudes times $\sqrt{r}$.

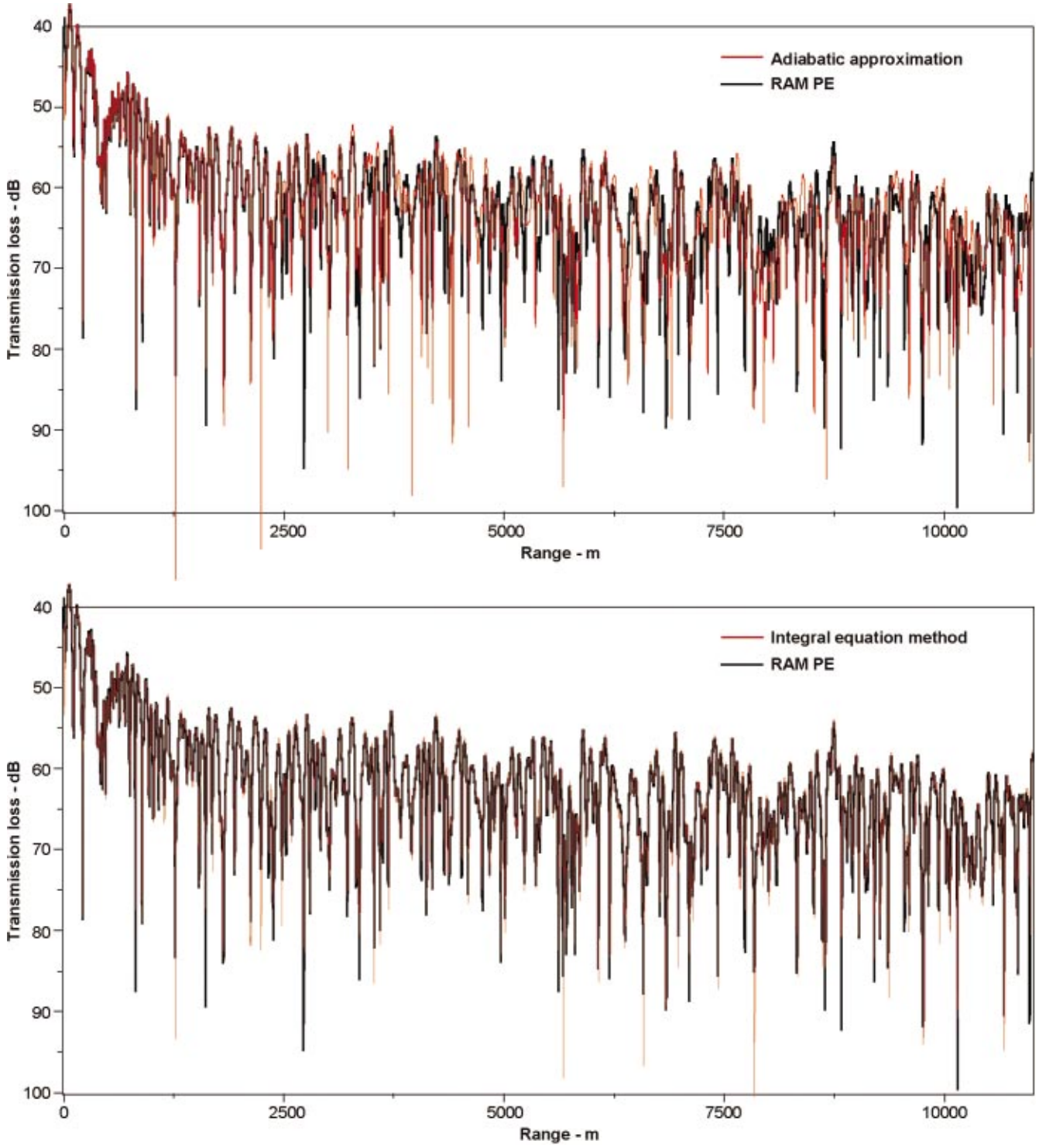


FIG. 7. (a) Transmission loss versus range produced using adiabatic approximation compared to that produced by the PE RAM model. Frequency is 100 Hz. Source depth and receiver depth are 133 and 70 m, respectively. (b) Transmission loss versus range produced using the integral equation method compared to that produced by the PE RAM model. Frequency is 100 Hz. Source depth and receiver depth are 133 and 70 m, respectively.

energy exchange between modes decreases with increasing mode number.

C. Transmission loss comparisons

The parabolic equation (PE) model RAM[17] version 1.5 was employed to generate transmission loss solutions to the continental shelf problem for the purpose of comparison to the integral equation method. The range, the depth, and the environment were sampled in the PE computations every 1, 0.25, and 1 m, respectively, the same as employed in the normal mode computation. Thus, any differences seen between the integral equation and PE solutions should not be a result of differences in environmental interpolation or spatial discretization. A false boundary was placed at 600 m beneath the sediment. This thick absorptive layer has an attenuation

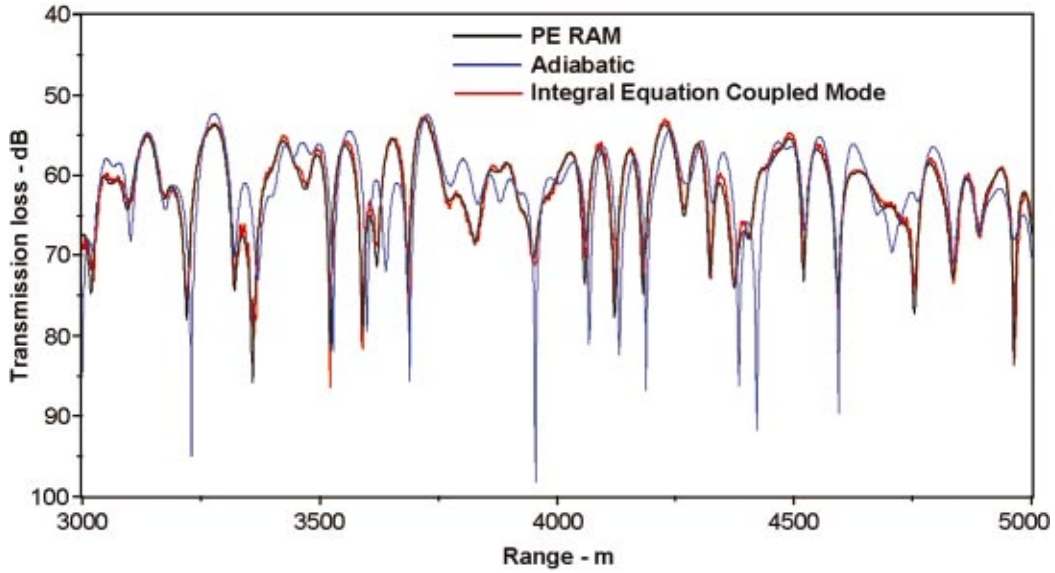


FIG. 8. Transmission loss versus range produced using the integral equation method compared to that produced by the PE RAM model and the adiabatic approximation in 3–5 km range interval. Frequency is 100 Hz. Source depth and receiver depth are 133 and 70 m, respectively.

that varies from 0.1 to 10.0 dB/λ. Convergence of the transmission loss results was obtained by using five Padé coefficients.

Figure 7(a) compares the 100 Hz transmission loss predicted by the adiabatic approximation to the solution predicted by the PE algorithm. The range interval of 0–11 km is large enough to show the overall agreement between the various computations but is small enough to discern many of the fine details of the differences. One observes the adiabatic approximation and RAM solution are in overall good qualitative agreement. As expected there is excellent agreement in the 0–2 km range interval. Beyond the range where the bathymetry changes (2 km), however, one observes disagreement in the fine details of the interference patterns. The level and nature of disagreement does not appear to change in the proximity of the rapid change in the SSP around 10 km. Figure 7(b) compares the integral equation solution to the PE solution. Only six Lanczos vectors were required to solve the coupled equations such that $\beta_6/\beta_1 \leqslant 10^{-2}$, where β_n is the magnitude of the nth Lanczos expansion coefficient. One observes very good agreement, both qualitatively and quantitatively. Since the PE and integral equation approach are based on two very different methods of solving the nonseparable wave equation, one can conclude that the transmission loss versus range prediction of the integral equation approach is at least very close to the exact answer. On the basis of the results in Figs. 7(a) and (b) one may also conclude that since both the adiabatic approximation and the PE algorithm neglect backscattering, the differences one observes between PE and the adiabatic approximation may be ascribed to mode coupling.

While the agreement between the RAM algorithm and the integral equation method is very good, there are differences. Figure 8 shows a selected range interval that demonstrates the general level of disagreement. The adiabatic approximation has also been included for reference. The differences observed between PE and the integral equation method are very small compared to their differences with the adiabatic approximation. Most of the differences between PE and the integral equation solution are on the order of a frac-

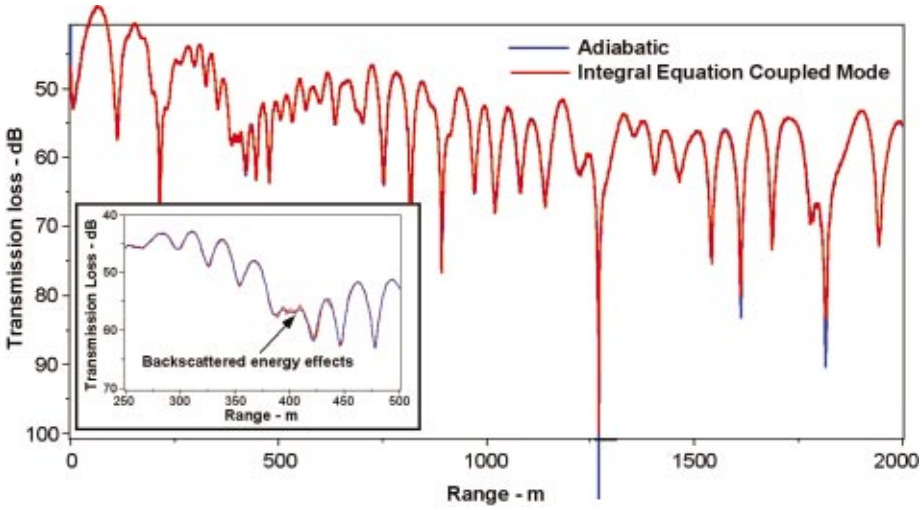


FIG. 9. Backscattering effects in 250–500 m range interval. Frequency is 100 Hz. Source depth and receiver depth are 133 and 70 m, respectively.

tion of a decibels; however, there are a few range points where the differences are larger. We may attribute some of these differences to a small backscattered field predicted by the integral equation method. One recalls that for the perfect ASA wedge[2] and the ''Fawcett'' hill,[6,8] both of which had a rigid bottom boundary condition, the acoustic fields had large backscattering components. For the ASA penetrable wedge, however, the amount of backscatter was found to be negligible.[2] This continental shelf case has a higher surface sediment speed with a very high sediment sound speed gradient. Because of the significantly higher impedance difference between the water and sediment, the high gradient in the sediment, and the higher frequency, one might expect a larger amount of backscatter for the continental shelf problem as compared to the ASA penetrable wedge. To illustrate that the integral equation method is indeed predicting backscatter, Fig. 9 compares the adiabatic and the full two-way coupled mode prediction of the transmission loss in the 250–500 m range segment. If there were no backscattered component to the acoustic field, the two solutions in this range interval would be identical. However, one does indeed observe small fluctuations around 400 m in the integral equation solution which are consistent with a small backscattered field. The current numerical implementation of the two-way integral formalism does not uniquely separate forward and back-scattered components. Clearly, it would be of interest to quantify the amount of back-scattered energy in a unique manner. One also observes a small shift in the interference pattern of the integral equation solution relative to the adiabatic solution between 415 and 435 m.

V. SUMMARY

A two-way integral equation method was applied to a continental shelf acoustic propagation problem. The nature of the mode coupling is dominated by nearest-neighbor coupling. The continuum associated with the lower halfspace was treated using a finite number of leaky modes whose mode functions were corrected at mode cutoff through the use of a small complex sound speed gradient in the lower halfspace. The RAM PE model was used for the purpose of comparisons. It was found that the integral equation and PE methods were in very good agreement. Small differences, however, are present between the PE and integral equation solutions. Using the adiabatic solution as a reference, it was concluded that these small differences could be attributed to a small backscattered component in the acoustic field.

ACKNOWLEDGMENTS

D.P.K. expresses his appreciation for helpful conversations with Dr. Richard Evans, Dr. Ellen Livingston, Dr. Kevin Smith, Dr. Alex Tolstoy, and Dr. Tom Yudichak.

[1] F. B. Jensen and W. A. Kuperman, ''Sound propagation in a wedge-shaped ocean with a penetrable bottom,'' J. Acoust. Soc. Am. **67**, 1564–1566 (1980).
[2] F. B. Jensen and C. M. Ferla, ''Numerical solutions of range-dependent benchmark problems in ocean acoustics,'' J. Acoust. Soc. Am. **87**, 1499–1510 (1990).
[3] C. T. Tindle, H. Hobaek, and T. G. Muir, ''Normal mode filtering for down-slope propagation in a shallow water wedge,'' J. Acoust. Soc. Am. **81**, 287–294 (1987).
[4] S. A. Chin-Bing, D. B. King, J. A. Davis, and R. B. Evans, *PE Workshop II: Proceedings of the Second Parabolic Equation Workshop* (Naval Research Laboratory, Washington, DC, 1993).
[5] V. Maupin, ''Surface waves across 2-D structures: A method based on coupled local modes,'' J. Geophys. **93**, 173–185 (1988).
[6] J. Fawcett, ''A derivation of the differential equations of coupled-mode propagation,'' J. Acoust. Soc. Am. **92**, 290 (1992).
[7] R. B. Evans, ''A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom,'' J. Acoust. Soc. Am. **74**, 188–195 (1983).
[8] D. P. Knobles, ''Solutions of coupled-mode equations with a large dimension in underwater acoustics,'' J. Acoust. Soc. Am. **96**, 1741–1747 (1994).
[9] D. P. Knobles, S. A. Stotts, R. A. Koch, and T. Udagawa, ''Integral equation coupled mode approach applied to internal wave problems,'' J. Comput. Acoust. **9**, 149–167 (2001).
[10] S. A. Stotts, ''Coupled-mode solutions in generalized ocean environments,'' J. Acoust. Soc. Am. **111**, 1623–1643 (2002).
[11] D. C. Strickler, ''Normal-mode program with both the discrete and branch line contributions,'' J. Acoust. Soc. Am. **57**, 856–861 (1970).
[12] C. L. Pekeris, ''Theory of propagation of explosive sound in shallow water,'' Geol. Soc. Amer. Mem. **27** (1948).
[13] E. K. Westwood and R. A. Koch, ''Elimination of branch cuts from the normal mode solution using gradient half spaces,'' J. Acoust. Soc. Am. **106**, 2513–2523 (1999).
[14] R. A. Koch, E. K. Westwood, J. Lemond, and D. P. Knobles, ''Improving a practical broadband adiabatic normal mode model by including untrapped modes,'' J. Acoust. Soc. Am. **103**, 2857 (1998).
[15] D. P. Knobles, S. A. Stotts, R. A. Koch, and E. K. Westwood, ''Inclusion of continuum effects in coupled mode theory using leaky modes,'' J. Acoust. Soc. Am. **103**, 2857 (1998).
[16] C. T. Tindle and Z. Y. Zhang, ''Continuous modes and shallow water sound propagation,'' Proceedings of the IEEE Oceans '93 Conference, 1993, pp. I-81–I-86.
[17] M. Collins, ''A split-step Padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742 (1993).

Spectral integral representations of monostatic backscattering from three-dimensional distributions of sediment volume inhomogeneities

Kevin D. LePage
SACLANT Undersea Research Centre, I-19138 La Spezia, Italy

Henrik Schmidt
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 1 May 2002; revised 17 August 2002; accepted 15 September 2002)

A theory is developed for generating short time, monostatic reverberation realizations caused by three-dimensionally distributed volume inhomogeneities in stratified media. A wave number integral approach to treating the propagation to and from the scatterers, combined with a two-dimensional spectral representation of the azimuthally averaged scatterer realizations and a novel numerical implementation, combine to yield an efficient, high fidelity reverberation simulator for predicting monostatic backscatter from horizontally stratified sediments. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1528625]

PACS numbers: 43.30.Bp, 43.30.Gv, 43.30.Hw [WMC]

I. INTRODUCTION

The prediction of scattering from sediment volume inhomogeneities has been the subject of extensive study over the years.[1–13] Typically the sediment is treated as a homogeneous medium in which small perturbations in sound speed and density are superimposed. Due to the nature of the problem, predictions of scattered intensity are usually obtained under one or more subsequent approximations. When propagation is treated with a full wave approach, the perturbations are usually assumed to be uncorrelated vertically in order to obtain closed from expressions for the expected value of scattered intensity.[3,5] If the propagation is simplified under the assumption that the scatterers are excited only by propagating waves, scattered intensity may also be obtained in closed form.[6,7] If the interaction of the field with the scatterers is asymptotically evaluated, the full scattering of propagating and inhomogeneous waves may be evaluated for three-dimensional distributions of scatterers, away from the part of the sediment ensonified near the critical angle.[9,10] However, if the propagation is to be treated ''exactly'' and a three-dimensional distribution of scatterers is to be retained, then Monte Carlo techniques for estimating the scattered intensity caused by sediment volume inhomogeneities are required. In a previous paper the authors laid out the theoretical basis for a unified treatment of sediment volume inhomogeneity scattering using spectral techniques.[11] In that paper the treatment of scattering from sound speed and density inhomogeneity scattering in general horizontally stratified acoustic media, including stratifications with background sound speed gradients, was presented for the two-dimensional wave equation. While this development was useful for understanding some of the behaviors of sediment scattering in refractive layered media, the results themselves cannot be used directly to interpret data because of the 2-D geometry used. We have therefore been motivated to extend the wave number integral approach to treat laterally monostatic scattering in three dimensions, which we view as a crucial step in extending the utility of the model.

Three-dimensional scattering under the perturbation approximation is theoretically very similar to the two-dimensional problem, especially in the fundamental equations. However the implementation is more computationally demanding. The wave number integrals which must be evaluated for three-dimensional problems are Hankel transforms rather than Fourier transforms. For bistatic problems there is the requirement to evaluate $k_o r$ Fourier orders n, where r is the bistatic separation between the source and the receiver, and k_o is the background sound speed.[14,15] Finally, there is potentially a much larger number of contributing scatterers to be summed over to give the backscattered pressure, due to the addition of a dimension. For these reasons the generation of wave theoretic scattered field predictions for general bistatic geometries and scatterer distributions is a computationally demanding area which is just beginning to be explored.[12,14]

In this paper we take the first step toward a general three-dimensional wave theoretic scatter prediction capability. We show that simplified expressions for backscatter may be obtained for laterally monostatic scenarios with horizontally isotropically distributed sediment inhomogeneities. In these cases the wave number integral representation of scattering from three-dimensional sound speed fluctuations may still be obtained with one-dimensional integral transforms. The result is that a high fidelity computational capability for the generation of backscattering time series is obtained for relatively small cost over the previous two-dimensional implementation. Since the results are wave theoretic, the propagation is treated properly for general stratified acoustic media, including the presence of background sound speed gradients. Propagation near the critical angle is accurately calculated for fast sediments, and propagation near turning points is accurately calculated for refractive sediments. And since the distribution properties of the inhomogeneities of

the scatterers are intrinsically three dimensional, scattering from realistic scatterer distributions may be evaluated. The combination of these two factors and a unique numerical implementation introduces a new prediction capability into the field of sediment acoustics.

II. THEORY

The first order scattered pressure observed at the field point $\{x, y, z\}$ due to small perturbation volume inhomogeneity scattering from sound speed perturbations $\Delta c/c$ in the depth interval $[z_1, z_2]$ is given by the convolution integral[11]

$$p_s(x,y,z) = 2k_b^2 \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \int_{z_1}^{z_2} dz' G(\mathbf{r},\mathbf{r}') \times \frac{\Delta c}{c}(x',y',z') p(x',y',z'), \tag{1}$$

where under the Born Approximation, p is the pressure incident on the scatterers in the absence of scattering effects, $G(\mathbf{r},\mathbf{r}')$ is the Green function from the scatterer at $\mathbf{r}' = (x',y',z')$ to the observer at $\mathbf{r} = (x,y,z)$, and the three-dimensional integral integrates over the space of contributing scatterers $\Delta c/c$. In a wave number integral modeling framework for an axisymmetric horizontally stratified medium, the Green function has the wave number integral representation[16]

$$G(R,z,z') = \int_0^{\infty} g(q_r;z,z') J_0(Rq_r) q_r \, dq_r, \tag{2}$$

with $R = |\mathbf{r} - \mathbf{r}'| = \sqrt{(x-x')^2 + (y-y')^2}$ representing the horizontal separation of scatterer and observer. $g(q_r;z,z')$ is the *depth-separated Green function* which, for example, in the case of a homogeneous background medium with acoustic wave number k_0, is[17]

$$g(q_r;z,z') = \frac{i}{4\pi} \frac{e^{i|z-z'|\sqrt{k_o^2 - q_r^2}}}{\sqrt{k_o^2 - q_r^2}}. \tag{3}$$

Similarly, for scatterers in a horizontally stratified ocean model, the incident acoustic pressure may be represented by a harmonic sum over Hankel transforms

$$p(x',y',z') = \sum_{n=0}^{\infty} \begin{Bmatrix} \cos n\theta' \\ \sin n\theta' \end{Bmatrix} \int_0^{\infty} \tilde{p}_n(k_r,z') J_n(k_r r') k_r \, dk_r, \tag{4}$$

where $r' \equiv \sqrt{x'^2 + y'^2}$ and $\theta' \equiv \arccos(x'/r')$.

A. Monostatic backscatter

In the case of observations of scattering in a laterally monostatic geometry $x, y \equiv 0$, Eq. (1) considerably simplifies when the source is horizontally omnidirectional. Then the only nonzero Fourier harmonic in the source expansion is $n = 0$, and upon substitution of Eqs. (2) and (4), Eq. (1) may be written

$$p_s(z) = 2k_b^2 \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \int_{z_1}^{z_2} dz' \frac{\Delta c}{c}(x',y',z') \times \int_0^{\infty} g(q_r;z,z') J_0(q_r r') q_r \, dq_r \times \int_0^{\infty} \tilde{p}_0(k_r,z') J_0(k_r r') k_r \, dk_r, \tag{5}$$

or equivalently, converting to cylindrical coordinates and setting $\delta c \equiv \Delta c/c$,

$$p_s(z) = 2k_b^2 \int_0^{\infty} r' dr' \int_{z_1}^{z_2} dz' \int_0^{2\pi} d\theta \, \delta c(r',\theta,z') \times \int_0^{\infty} g(q_r;z,z') J_0(q_r r') q_r \, dq_r \times \int_0^{\infty} \tilde{p}_0(k_r,z') J_0(k_r r') k_r \, dk_r. \tag{6}$$

This is not in the most efficient form for calculating scattering in a layered waveguide. This is because Eq. (6) requires an azimuthally integrated realization of the sound speed defect, which implies that a full three-dimensional realization of the defects has to be computed on a cylindrical grid, which must then be integrated over azimuth by numerical quadrature. In general, we would like to be able to generate a realization of an azimuthally integrated realization directly without having to implement the quadrature. As we show below, if we know the two-dimensional Cartesian coordinate power spectrum of the defects as a function of depth, and this spectrum is for a homogeneous, isotropic process, then it is possible to calculate the azimuthally integrated realizations directly.

We first represent the scatterer realization in cylindrical coordinates as a two dimensional Fourier transform of the Cartesian scatterer spectrum at depth z', with a change of variables to cylindrical wavenumber coordinates

$$\delta c(r',\theta,z') = \int_0^{\infty} p_r' \, dp_r' \int_0^{2\pi} d\theta' \, \tilde{\delta c}(p_r' \cos\theta', p_r' \sin\theta', z') \times e^{ip_r' r' \cos\theta' \cos\theta} e^{ip_r' r' \sin\theta' \sin\theta}. \tag{7}$$

We may then integrate δc with respect to θ by utilizing the trigonometric identity $\cos(\theta - \theta') = \cos\theta\cos\theta' + \sin\theta\sin\theta'$ and recognizing the integral representation of the Bessel function J_o. The result is

$$\begin{aligned} \int_0^{2\pi} d\theta \, \delta c(r',\theta,z') &= \int_0^{\infty} p_r' \, dp_r' \int_0^{2\pi} d\theta' \, \tilde{\delta c}(p_r' \cos\theta', p_r' \sin\theta', z') \times \int_0^{2\pi} d\theta \, e^{ip_r' r' \cos(\theta - \theta')} \\ &= 2\pi \int_0^{\infty} p_r' \, dp_r' \int_0^{2\pi} d\theta' \, \tilde{\delta c}(p_r' \cos\theta', p_r' \sin\theta', z') \times J_o(p' r'). \end{aligned} \tag{8}$$

$\tilde{\delta c}$ is the Fourier transform of a random variable conforming to a known two-dimensional power spectrum $P_{\delta c}$. If we want to know the equivalent power spectrum of the azimuthal integral of δc at wave number p_r' then we must evaluate the expected value of the square of the Hankel transform of Eq. (8)

$$\int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \langle \delta c(p_{r1},\theta_1,z')\,\delta c(p_{r2},\theta_2,z')\rangle$$

$$= 4\pi^2 \int_0^{2\pi} d\theta_1' \int_0^{2\pi} d\theta_2' \langle \tilde{\delta c}(p_{r1}\cos\theta_1', p_{r1}\sin\theta_1', z')$$

$$\times \tilde{\delta c}(p_{r2}\cos\theta_2', p_{r2}\sin\theta_2', z')\rangle. \tag{9}$$

If δc is a homogeneous random process, such that its correlation function is a function only of $[r_1'-r_2', \theta_1-\theta_2]$, then each wave number of the power spectrum is an independent random process. In this case the ensemble average inside the integrals over θ_1' and θ_2' on the second line of Eq. (9) is

$$\langle \tilde{\delta c}(p_{r1}\cos\theta_1', p_{r1}\sin\theta_1', z')\,\tilde{\delta c}(p_{r2}\cos\theta_2', p_{r2}\sin\theta_2', z')\rangle$$

$$\equiv P_{\delta c}(p_{r1}\cos\theta_1', p_{r1}\sin\theta_1')\,\delta(p_{r1}-p_{r2})\,\frac{\delta(\theta_1'-\theta_2')}{p_{r1}}. \tag{10}$$

Furthermore, if the power spectrum is not only for a homogeneous process but also for an isotropic one, then $P_{\delta c}(p_{r1}\cos\theta_1', p_{r1}\sin\theta_1') \equiv P_{\delta c}(p_{r1})$ and the left hand size of Eq. (9) may be integrated yielding

$$\int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \langle \delta c(p_{r1},\theta_1,z')\,\delta c(p_{r2},\theta_2,z')\rangle$$

$$= 8\pi^3 P_{\delta c}(p_{r1})\,\frac{\delta(p_{r1}-p_{r2})}{p_{r1}}. \tag{11}$$

Equation (11) is the variance of the azimuthally integrated sound speed defect δc for wave numbers p_{r1} and p_{r2} at a depth of z'. The delta function indicates that the wave number bins are uncorrelated. Thus it follows that if we want to generate realizations of $\int_0^{2\pi} d\theta\ \delta c(r,\theta,z')$ consistent with the spectral variance in Eq. (10), we need only evaluate the Hankel transform of a kernel which is an uncorrelated Normal random variable with variance $8\pi^3[P_{\delta c}(p_r)/p_r]$

$$\int_0^{2\pi} d\theta\, \delta c(r',\theta,z')$$

$$= 2\pi\sqrt{2\pi} \int_0^{\infty} dp_r \sqrt{p_r} J_o(p_r r') N(0, P_{\delta c}(p_r)). \tag{12}$$

Insertion of Eq. (12) into Eq. (6) then yields

$$p_s(z) = 4\pi\sqrt{2\pi}k_b^2 \int_0^{\infty} r'\,dr' \int_{z_1}^{z_2} dz' \int_0^{\infty} \sqrt{p_r} N(0, P_{\delta c}(p_r, z'))$$

$$\times J_0(p_r r')\,dp_r \int_0^{\infty} g(q_r; z, z') J_0(q_r r')\, q_r\, dq_r$$

$$\times \int_0^{\infty} \tilde{p}_0(k_r, z') J_0(k_r r')\, k_r\, dk_r. \tag{13}$$

Changing the orders of integration, Eq. (13) can be cast into a form which is directly suited for implementation in a wave number integration modeling framework for ocean waveguides,

$$p_s(z) = 4\pi\sqrt{2\pi}k_b^2 \int_0^{\infty} q_r\, dq_r \int_{z_1}^{z_2} dz' g(q_r; z, z')$$

$$\times \int_0^{\infty} J_0(q_r r')\, r'\, dr' \int_0^{\infty} \sqrt{p_r} N(0, P_{\delta c}(p_r, z'))$$

$$\times J_0(p_r r')\, dp_r \int_0^{\infty} \tilde{p}_0(k_r, z') J_0(k_r r')\, k_r\, dk_r. \tag{14}$$

In this form the two inner integrals are uncoupled and may be evaluated directly *a priori* for a selected grid of scatterers in the spatial (r', z') domain. The integral over k_r represents the incident pressure field, while the integral over p_r represents the cylindrically averaged scattering strength at (r', z'). The product of these integrals then represents the horizontal distribution of virtual sources at depth z' contributing to the scattered field. The Hankel transform integral over r' converts this sheet of virtual sources into an equivalent wave number spectrum, which when multiplied by the depth-dependent Green function represents the generated scattered field component at horizontal wave number q_r. The depth integral then superimposes the contributions from all scatterer depths to produce the total kernel of the Hankel transform integral for the monostatic reverberent field on the coordinate axis $r=0$, which is evaluated using any of the standard wave number integration techniques.[10]

B. Fast Hankel transform evaluation

In underwater acoustic modeling the Hankel transform wave number integrals have traditionally been evaluated using the so called Fast Field Program (FFP) approach,[18] which is based on the large argument asymptotic of the Hankel functions. However, for the bottom scattering problem this approach is inadequate because of the significance of the steep angle contributions from scatterers at short range. Even on todays computers the generation of Bessel functions is time consuming, and a direct numerical quadrature technique for evaluating the Hankel transforms in Eq. (14) is infeasible. A number of *Fast Hankel Transforms* (FHT) have been developed,[19–21] but they generally have nonuniform sampling requirements and therefore not best suited to the ocean acoustics problem where the field is desired at a predetermined grid of horizontal ranges. Being based on the Fast Fourier Transform (FFT), the advantage of the FFP is that it efficiently approximates the Hankel transform on a regular grid of wave numbers and ranges. However, it is possible to design a numerically efficient correction to the FFP which allows for accurate accounting for the small argument contributions.

The Hankel transform integrals are of the form

$$f(r) = \int_0^{\infty} f(k) k J_m(kr)\, dk$$

$$= \frac{1}{2} \int_0^{\infty} f(k) k [H_m^{(1)}(kr) + H_m^{(2)}(kr)]\, dk, \tag{15}$$

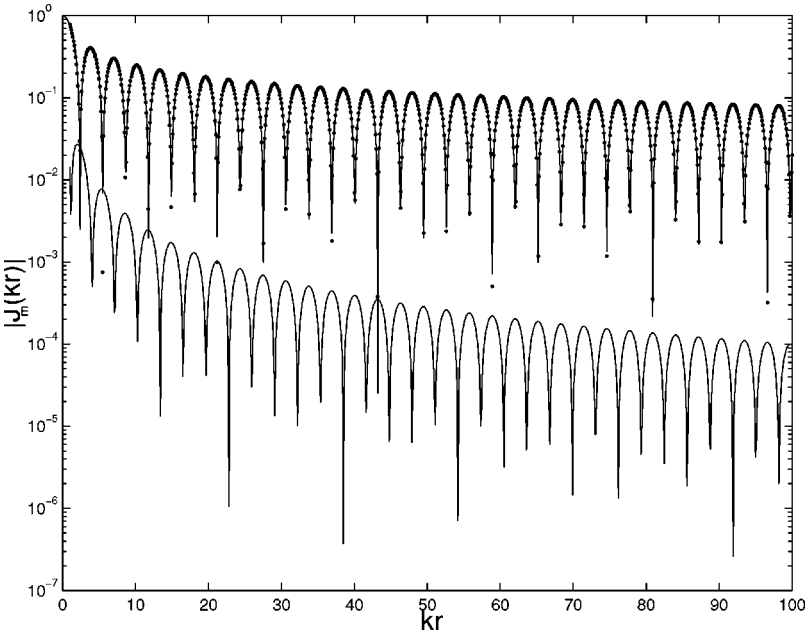


FIG. 1. Error of far-field approximation of $J_0(kr)$. Solid: Exact. Dots: Large-argument approximation. Lower curve: Error.

where $H_m^{(1)}$ for $\exp -i\omega t$ time-dependence corresponds to outgoing waves and $H_m^{(2)}$ to incoming waves, both of azimuthal Fourier order m dependence. For classical acoustic forward propagation modeling the incoming waves are ignored, eliminating the second term. However, for the backscattering problem they must be retained. We can represent the Hankel functions $H_m^{(1,2)}(k_r r)$ by their asymptotic form[22]

$$\lim_{kr\to\infty} H_m^{(1,2)}(k_r r) = \sqrt{\frac{2}{\pi k_r r}}\, e^{\pm i[k_r r-(m+1/2)\pi/2]}, \tag{16}$$

to arrive at the following approximation to the Hankel transform,

$$f^*(r) = \sqrt{\frac{1}{2\pi r}} \int_0^\infty f(k)\sqrt{k}[e^{i[kr-(m+1/2)\pi/2]} + e^{-i[kr-(m+1/2)\pi/2]}]dk. \tag{17}$$

This integral can be rewritten as a two-sided Fourier transform,

$$f^*(r) = \int_{-\infty}^{\infty} g(k) e^{ikr}\, dk, \tag{18}$$

with the kernel

$$g(k) = \begin{cases} f(-k)\sqrt{\dfrac{-k}{2\pi r}}\, e^{i(m+1/2)\pi/2} & k < 0 \\ f(k)\sqrt{\dfrac{k}{2\pi r}}\, e^{-i(m+1/2)\pi/2} & k \geqslant 0 \end{cases}. \tag{19}$$

The Fourier transform is efficiently evaluated using a standard FFT if the range r and wave number k are discretized equidistantly, with the sampling intervals being constrained by

$$\Delta r \Delta k_r = \frac{2\pi}{M}, \tag{20}$$

where M is the length of the discrete Fourier transform sequence M, which is generally a power of 2.

The error associated with using Eq. (18) is clearly associated with the approximation of the Bessel function in terms of the asymptotic expressions of the Hankel functions. Figure 1 shows the exact value of the absolute value of the Bessel function $J_0(kr)$ for $kr \leqslant 100$ as a solid curve and the asymptotic values indicated by the dots. The absolute error is indicated by the lower curve, and it is clear that the relative error of the approximation is greater than 10^{-5} for $kr \leqslant 20\pi$. Consequently, to achieve a more accurate evaluation of the Hankel transforms of order $n=0$ in Eq. (14) it is only necessary to correct the contributions corresponding values of $kr \leqslant \mathrm{KR} = 20\pi$. This is performed in a numerically stable manner by a weighted average of the contributions of the exact Bessel function and the approximate FFP kernel,

$$f(r) = f^*(r) + \int_0^\infty f(k) w(kr) k J_0(kr) dk - \int_0^\infty f(k) w(kr) \sqrt{\frac{k}{2\pi r}}[e^{i(kr-(m+1/2)\pi/2)} + e^{-i(kr-(m+1/2)\pi/2)}]dk, \tag{21}$$

where we have chosen the tapered weight function $w(kr)$ as

$$w(kr) = \begin{cases} 1 & kr \leqslant \mathrm{KR}/2 \\ [1+\cos(\pi(kr-\mathrm{KR}/2)/(\mathrm{KR}/2))]/2 & \mathrm{KR}/2 < kr < \mathrm{KR} \\ 0 & kr \geqslant \mathrm{KR} \end{cases}. \tag{22}$$

Equation (21) can be evaluated very efficiently. First of all, with the wave number and range sampling constrained by Eq. (20), all values of the exponentials are computed as part of the FFT evaluation of Eq. (17). Second, the Bessel func-

tions will only be needed for a limited number of discrete values of the argument,

$$kr = n\Delta k \Delta r, \quad n = 0,...,\mathrm{KR}/(\Delta k \Delta r), \tag{23}$$

which can be precomputed into a look-up table.

The performance of this FHT is illustrated by Fig. 2, which shows the evaluation of the Hankel transform

$$p(r,z) = \int_0^\infty \frac{e^{ik_z|z|}}{ik_z} kJ_0(kr)dk \tag{24}$$

representing the free-field point source field. $k_z = \sqrt{(\omega/c)^2 - k^2}$ is the vertical wave number at angular frequency ω. Figure 2(a) shows the FFP approximation which clearly breaks down at steep angles and short ranges, while Fig. 2(b) shows the correct spherical spreading behavior at all propagation angles produced using Eq. (21).

III. RESULTS

Using the FHT outlined in Eq. (21), Eq. (14) and its wave theoretic equivalent for scattering from density inhomogeneities[11] has been implemented inside the SAFARI/OASES[23] code in a numerically efficient way. In the algorithm, the code is first run to generate the incident field through the scattering volume, as indicated by the third line of Eq. (14). Under the Born Approximation this field is assumed to be unaffected by the scattering process. This incident field is then multiplied in the spatial domain with the azimuthally averaged scatterer distributions at each range and depth, as shown by the second line in Eq. (14). The resulting virtual source distribution at each scattering depth is then transformed to the scattered wave number domain. This gives the spectral distribution of the right-hand side of the inhomogeneous depth-separated wave equation for the scattered field. The source spectrum for the scattered field at each scatterer depth is essentially a more complicated version of a point source expansion, since there is now an array of point sources at each depth whose strengths depend both on the incident field and the distribution properties of the scatterers.

The resulting inhomogeneous depth separated wave equation for the scattered field is solved using the *Global Matrix* approach.[23] The result is the spatial distribution of the scattered field for the desired frequency. To generate scattered field time series, Eq. (14) must be solved over an array of frequencies for the same scatterer distribution. Thus the second line of Eq. (14) is evaluated once for the scatterer distribution and is then multiplied by the incident field at each frequency of interest at all wave numbers and depths in order to determine the virtual source representation. The resulting frequency dependent source term is then used to drive the depth separated wave equation for the scattered field, which is also frequency dependent, over the desired frequency band. The result is a complex transfer function between the incident field which was used to determine the third line of Eq. (14) and the scattered field at the desired horizontally monostatic receiver depths z.

The complex scattered field transfer functions may be used to synthesize scattered field time series using Fourier synthesis. The sampling requirements of the scattered field time series must be in harmony with those of the spatial sampling of the scatterer realization and the incident field. In practice the spatial integrals on the second and third lines of Eq. (14) are truncated at a range sufficient to give scattered returns from all times of interest in the temporal Fourier synthesis. For instance, for a 1 Hz sampling interval, scatterers up to a range of roughly 750 m are expected to contribute to the scattered field time series at the 1 s upper limit of the corresponding temporal window of the time series. Given a sample frequency of f_s the number of time samples is then $f_s/\Delta f$. In addition, the spatial sampling interval must be sufficiently fine to properly excite backscattered waves, which means that the maximum spatial sampling interval which is possible is one half a wavelength. Since the scatterer realization is generated once, this criteria is required to be satisfied at the upper end of the frequency band of interest. In practice the spatial sampling is also required to be smaller than the correlation length scale of the scatterers. The final requirement used was $\Delta x = 2\pi/k_{\max} \leqslant \min((\lambda/2),(\ell/10))$.

A. Scenario 1: Volume scattering from a slow sediment layer

Scattered field time series evaluated by Fourier synthesis of Eq. (14) have been generated for three scattering scenarios. Scenario 1 is illustrated in Fig. 3. A point source is situated 25 m above a slow sediment layer 100 m thick with a background sound speed of 1455 m/s, a background density of 1.024 g/cm^3, and an intrinsic attenuation of 0.01 dB/λ. In the sediment layer there is a distribution of sound speed and density fluctuations which are 100% correlated. The rms amplitude of the sound speed fluctuations is 1 m/s, while the rms density fluctuation is 7/1455 g/cm^3. The fluctuations are distributed according to a power law with a lateral correlation length scale $\ell_x = \ell_y = 10$ m and a vertical correlation length scale $\ell_z = 2$ m. These values are notional. Note that the relatively flat roll-off of the inhomogeneity spectrum ensures roughness in these realizations at a broad spectrum of spatial scales. The scatterer power spectrum required in Eq. (14) has the form

$$P_{\delta c}(p_r,p_z) = \langle \delta c^2 \rangle \frac{\ell_r^2 \ell_z \Gamma(1.75)}{\pi^{1.5}\Gamma(0.25)} \times (1 + p_r^2\ell_r^2 + p_z^2\ell_z^2)^{-1.75}, \tag{25}$$

which corresponds to a three-dimensional von Karman (as discussed by Turgut[13]) spectrum with fractal dimension 3.5.[24]

In all the scenarios, the properties of the basement are chosen to be identical to the background properties of the sediment layer. This is not a requirement of the implementation, which can indeed insert a volume scattering layer into an arbitrary horizontally stratified waveguide, but rather a choice which tends to make the benchmarking and explanation of the results more straightforward by eliminating multipath.

In Fig. 4 the absolute value of the complex envelope of the scattered field time series received on an array of receiv-

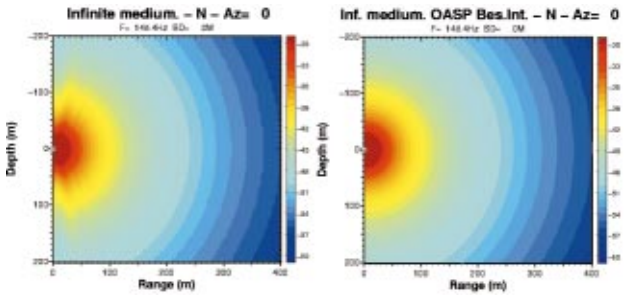


FIG. 2. Acoustic point source field. (a) shows the FFP approximation which clearly breaks down at steep angles and short ranges, while (b) shows the correct spherical spreading behavior at all propagation angles, produced using Eq. (21).

ers at the origin are shown in a stacked format for a particular realization of sediment inhomogeneities. The source pulse is Gaussian shaded

$$p_{\text{source}}(t,r=1\ \text{m}) = \exp\{-2i\pi f_o t - t^2 \Delta\omega^2/2\},$$

with a center frequency of $f_o = 500$ Hz and a bandwidth of $\Delta\omega/2\pi = 20$ Hz. The results show that the receivers closest to the sediment water interface receive the scattered field first, with the slope of the first arrival at higher receivers corresponding to the sound speed in the upper half space.

To further interpret the scattered field predictions it is useful to beamform the scattered field received on the vertical line array to transform the results to the time-angle domain. The resulting time-beam evolution of scattered field intensity, averaged over 16 independent scatterer realizations conforming to the distribution properties in Eq. (25), is shown in the lower panel of Fig. 5. In the upper panel an independent estimate of the beam-time evolution of backscatter intensity obtained with the ray trace code SCARAB[7,8] is

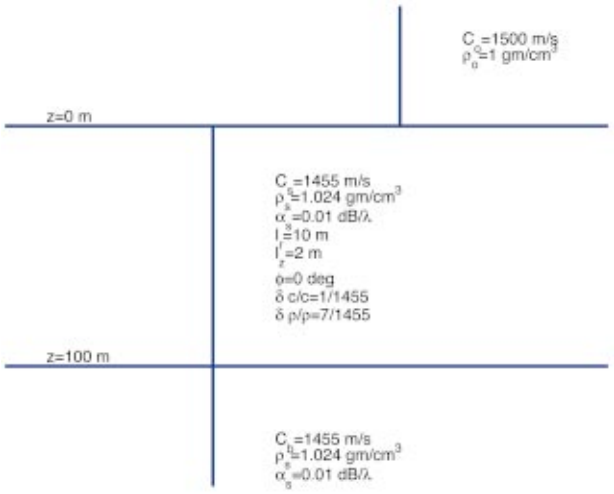


FIG. 3. Sound speed profile and scatterer properties of volume scattering Scenario 1. An isospeed halfspace overlies a slow sediment layer 100 m thick with a background sound speed of 1455 m/s and density of 1.024 g/cm^3. The basement halfspace has identical properties. In the scattering layer there are 100% correlated sound speed and density inhomogeneities conforming to a power law distribution with a fractal dimension of 3.5. The rms sound speed fluctuation is 1 m/s and the density fluctuation is 7/1455 g/cm^3. The horizontal correlation length scales of the scatterers is 10 m and the vertical correlation length scale is 2 m. The high frequency asymptote of the scatterer distribution power spectrum is $k^{-7/2}$.

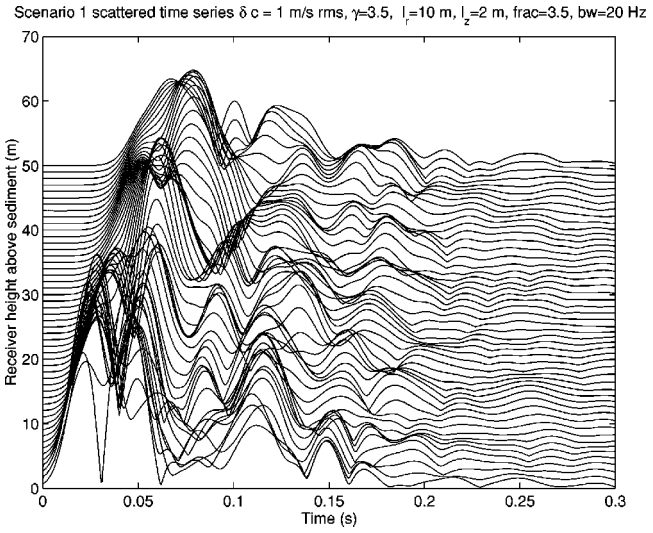


FIG. 4. Scenario 1 stacked scattered field time series. The absolute value of the complex envelope of the field received over a 50 m aperture array of receivers as a function of time after the source pulse. The source pulse is Gaussian shaded with a center frequency of 500 Hz and a bandwidth of 20 Hz.

shown for purpose of comparison. In the wave number integral representation the source pulse has a peak level of 1 μPa@1 m and is Gaussian shaded with a bandwidth of 20 Hz. In the SCARAB implementation the source pulse window is a rectangle of uniform amplitude of 1 μPa@1 m and duration $T = 5$ ms. The spectral results have therefore been adjusted to correct for the differences in incident energy between these two source pulses. The incident energy in the SCARAB source pulse is

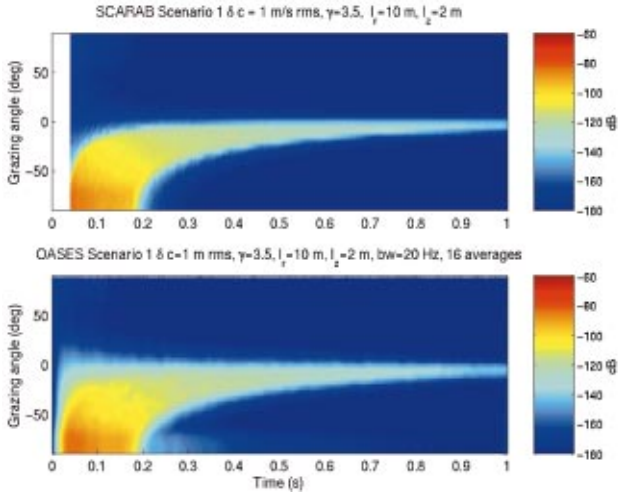


FIG. 5. The beam-time evolution of volume scattering intensity predicted by SCARAB (upper plot) and the wave number integral implementation discussed in this paper (lower plot). In the wave number integral approach, time series are collected and beamformed on an array with a 50 m aperture at the origin whose center is co-located at the source depth. The beamformed results are then squared and averaged over an ensemble of 16 independent realizations of the sediment inhomogeneities to give an estimate of the expected value of the scattered intensity. Good agreement is seen between the two approaches, with the scattered intensity falling between the time-angle trajectories of scatterers at the top and bottom of the sediment layer. Note that the beamforming of the wave number integral results introduces some sidelobes. The SCARAB results have no sidelobes because they are obtained directly in the time-angle domain.

$$E_{SCARAB} = \int_0^T \cos^2(i\omega_o t)\,dt = T/2 = 2.5E-3, \tag{26}$$

while for the wave number integral with the Gaussian shaded source pulse it is

$$E_{OASES} = \int_{-\infty}^{\infty} (\cos(i\omega_o t) e^{-(\omega-\omega_o)^2/2\Delta\omega^2})^2\,dt$$

$$= \sqrt{\pi}/2\Delta\omega = 1/80\sqrt{\pi}. \tag{27}$$

With the normalization of the scattered pressure computed by the wave number integral technique by the factor $\sqrt{0.2\sqrt{\pi}}$ it is possible to directly compare the scattered field intensity received by a single hydrophone with the SCARAB predictions. The results are shown in Fig. 6 for a receiver at the source depth of 25 m above the sediment-water interface. In general the agreement between the two results is quite good, within 1 or 2 dB over the range of times and angles.

The agreement between the wavenumber integral and SCARAB representations of the scattered field intensity is satisfying because the differences between Eq. (14) and the sonar type equation for RL in the SCARAB manual are quite significant (for convenience, a short description of the SCARAB model is given in the Appendix). While both models use the Born Approximation and first order perturbation theory, both the Green functions and the scattering are treated differently in practice. SCARAB has a ray trace approximation for the Green function, while the wave number integral implementation accurately models the total Green function, including evanescent waves, turning points, and propagation near the critical angle for fast bottoms. For scattering SCARAB integrates the differential plane wave scattering cross-section of the sediment inhomogeneities over the sediment volume to estimate the scattered intensity, while the wave number integral calculates the scattered pressure from each scatterer over a particular realization of scatterers. In the case of wave number integral results, the expected value of the scattered intensity must be estimated by Monte Carlo averaging the square of the scattered pressure over an ensemble of scatterer realizations.[25] Finally, the expression SCARAB uses for the differential cross-section, Eq. (42) from Yamamoto,[6] also does not require the gradient of the scatterer distribution or of the incident field, as do the density scattering contributions in the wave number integral representation.[11] Thus good agreement between the SCARAB and the wave number integral results lends confidence that Eq. (14) and the numerical implementation are consistent with the sonar equation approach to describing the sediment inhomogeneity scattering process for slow bottoms. As discussed below, we expect the two approaches to be equivalent only in cases where the sediment scattering process is well described by a plane wave differential scattering cross section and where the propagation is well approximated by ray theory.

B. Scenario 2: Volume scattering from a fast sediment layer

One significant advantage of using the wave number integral approach to modeling sediment volume inhomogeneity scattering is that the Green function is accurately modeled. Failings in the ray theory approximation to the Green functions are especially evident near turning points in refractive sediments and near the critical angle for fast sediments. Another significant advantage is that the wave number integral approach correctly includes the scattering response of scatterers which have been excited by evanescent, or nonpropagating waves. This type of scattering is typically found in fast sediments along the sediment water interface at ranges where the incident angle of the source is smaller that the critical angle. For scatterers ensonified by these types of waves, the differential plane wave scattering cross section is undefined. Yet it is known that scatterers ensonified by these types of waves can be important contributors to the total scattered field.[9] In fact sub-critical penetration and scattering is the basis for advanced mine hunting sonar concepts for buried object detection.[26,27] If this mechanism is going to be included for modeling target response, it is important that a similar mechanism be included for modeling the reverberation from sediment volume inhomogeneities, especially for smooth bottoms. For these reasons in Scenario 2 we compare the scattering predictions of the wave number integral approach to SCARAB for a fast sediment scattering scenario in order to evaluate the differences between the two approaches when there is a critical angle.

Scenario 2 is illustrated in Fig. 7. A point source is situated 25 m above a fast sediment layer 100 m thick with a background sound speed of 1600 m/s, density of 2 g/cm^3, and attenuation of 0.01 dB/λ. In the sediment layer there is the similar distribution of 100% correlated sound speed and density fluctuations as were described in Scenario 1. The rms sound speed fluctuation is 1 m/s, the rms density fluctuation is 7/1600 g/cm^3, and the correlation length scales are $\ell_x = \ell_y = 10$ m and $\ell_z = 2$ m. The source characteristics are the same as in Scenario 1.

In Fig. 8 the time-angle evolution of backscattered intensity is illustrated for the fast sandy bottom type of Scenario 2. The SCARAB result is shown in the top plot and the wavenumber integral result is shown in the bottom plot. The results are in good agreement for angles greater than the critical angle of 20°, but the wave number integral result also shows a shallower angle branch at late time which conforms to the time-angle trajectory of the sediment water interface. This upper trace corresponds to the inhomogeneous forcing of scatterers by nonpropagating evanescent waves at the sediment water interface. Since the waves are not propagating in the sediment, there is no differential scattering cross section for SCARAB to use to estimate this branch. Instead SCARAB estimates that the only propagation path of importance to scatterers near the sediment water interface is the headwave path, which enters the sediment at the critical angle and propagates at near-horizontal angles to these scatterers and back. This is a fundamental limitation of the SCARAB assumption that all of the scattering of the sediment volume inhomogeneities can be parameterized in terms of differential scattering cross section.

In Fig. 9 the scattered intensity received monostatically for Scenario 2 is illustrated in red. Here it is seen that although the SCARAB result neglects the inhomogeneous scat-

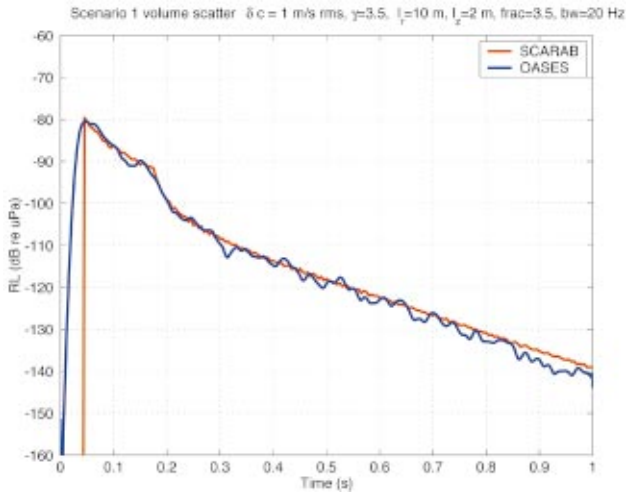


FIG. 6. Single phone scattered field intensity estimates from SCARAB (red) and the current implementation (blue). The wave number integral estimates of the scattered pressure have been squared and averaged over 16 independent realizations of the sediment layer inhomogeneities. Agreement with the SCARAB results is quite good, typically within 1 or 2 dB.

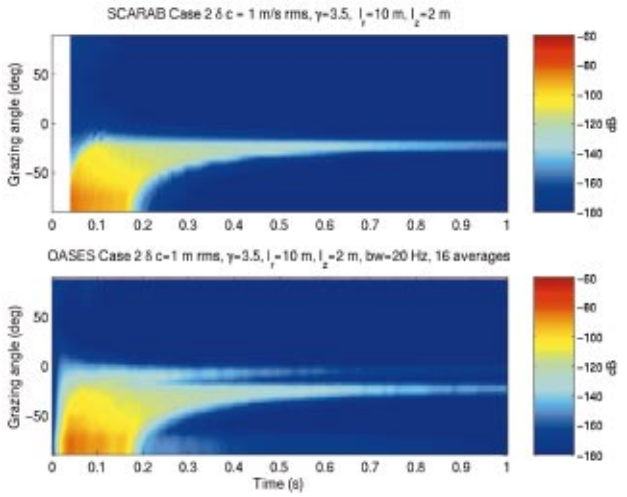


FIG. 8. The time-beam evolution of scattered intensity for Scenario 2. While the general agreement at angles greater than the critical angle is very good, the forced evanescent scattering branch following the time-angle trajectory of the sediment-water interface is missing in the SCARAB result (top). This branch is seen in the wave number integral results (bottom) as well as the headwave scattering branch arriving at and above the critical angle of 20°. The presence of the fast sediment causes the scattering associated with the propagating part of the Green function to be confined to angles greater than or equal to the critical angle of the bottom for all times greater that the time necessary to interrogate the sediment layer at the critical angle, about 90 ms for the current geometry.

tering mechanism, the overall agreement between the wave number integral result and SCARAB is still good. This is because in this case the attenuation in the sediment, 0.01 dB/λ, is sufficiently small that the headwave propagation path is the strongest contributor to the total scattered field for the times of interest in this simulation. However, in Scenario 2a when we increase the sediment attenuation to 0.5 dB/λ, the blue curves show that SCARAB and the wave number integral result no longer agree for times later than about 0.16 s. This is due to the lack of an evanescent scattering mechanism in SCARAB. Taken together, the results for Scenarios 2 and 2a indicate that the evanescent scattering mechanism is about 20 dB down from the headwave mechanism in the absence of loss. For the sake of completeness, the beam time trajectories for Scenarios 2 and 2a are illustrated in Fig. 10.

C. Scenario 3: Volume scattering from an upward refracting sediment layer

Gradients in the background properties of sediments are commonly observed in coring data and in bottom sound speed inversions.[28,29] Gradients are important to the physics of the scattering process because they control how the subbottom is illuminated.[4,8] These gradients are often composed of microstructure with significant sound speed and density variance. However, some of the characteristics of scattering from these types of sediments may be accounted for by assuming that the background properties are smoothly varying

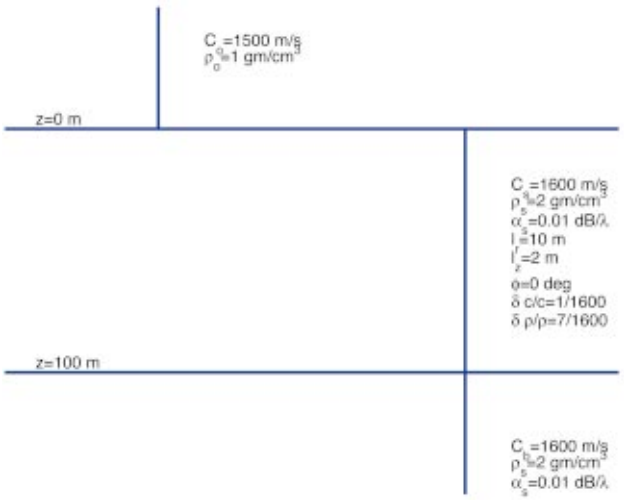


FIG. 7. Scenario 2 is typical of a sandy sediment layer, with a sediment background sound speed of 1600 m/s, density of 2 g/cm^3, and attenuation of 0.01 dB/λ. The basement properties are identical to the background properties of the sediment, and the scatterer distribution properties are identical to those in Scenario 1.

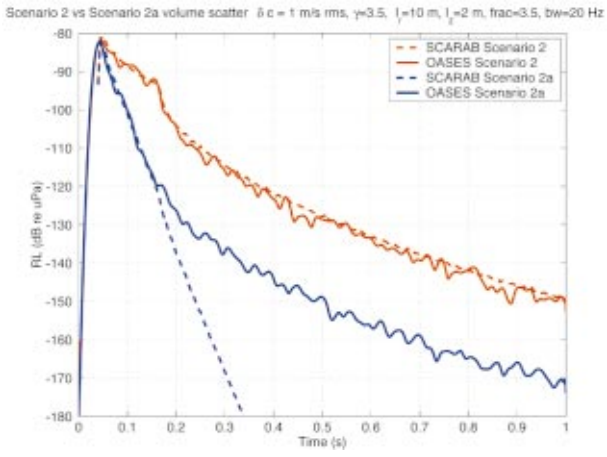


FIG. 9. Single phone average intensity estimates for Scenario 2 (red) and Scenario 2a (blue) which has a higher sediment attenuation of 0.5 dB/λ. The equivalent SCARAB predictions are indicated by the dashed lines of the appropriate color. The results show that in Scenario 2a the SCARAB and wave number integral results do not agree. This is due to the absence in SCARAB of an evanescent wave scattering mechanism.

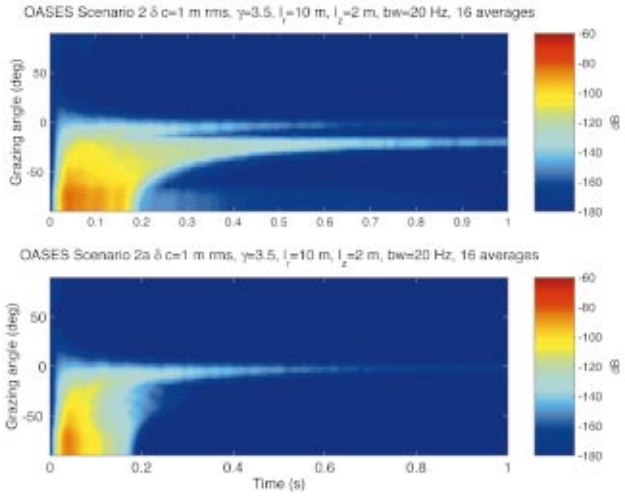


FIG. 10. The time-beam evolution of scattered intensity for Scenario 2 (upper panel) and Scenario 2a (lower panel) which has a lossier sediment. In the lossier sediment case, the headwave scattering branch is attenuated after about 0.3 s. At times greater than this, the evanescent scattering path associated with the sediment water interface is the only scattering mechanism of importance.

over the depth of the sediment layer. In wave number integral techniques, sound speed gradients which obey linearity in the square of the index of refraction may be handled exactly, as in these cases the depth separated wave equation becomes a differential equation whose solutions are Airy functions.[4,16] In Scenario 3 such a gradient is introduced into the background sound speed properties of the fast sandy sediment in Scenario 2 (Fig. 11). The sound speed at the bottom of the layer is increased to 1700 m/s, while all the other properties of the sediment layer are held fixed. To eliminate multipath, the basement sound speed is also set at 1700 m/s.

The effect of the sound speed gradient on the incident field is illustrated in Fig. 12, where the range-depth transmission loss is shown for the point source 25 m above the sediment water interface at 500 Hz for Scenarios 2 and 3. In both scenarios, the incident field passes unhindered into the sediment layer save for a small reflection loss for ranges up to 67 m, where the critical angle occurs. However in Scenario 2, the field then passes down into the lower halfspace, while in Scenario 3 a significant portion of the transmitted energy is refracted back up into the sediment layer.

In Fig. 13 the time angle distributions of backscattered intensity for Scenarios 2 and 3 are shown. The results show that these distributions are significantly different. While the high angle early time backscatter and the forced inhomogeneous backscatter associated with the sediment water interface are identical for the two scenarios, the deep refracted or headwave path contains significantly more energy at late time for Scenario 3 than it does for Scenario 2. The angular width of the refracted scattering branch is also wider for Scenario 3, because higher angle energy is able to ensonify the sediment layer at late times due to the refraction of this energy back into the sediment layer before it can escape into the basement.

In Fig. 14 the backscattered intensity for the two scenarios is shown. Here it can be seen that the decay rate of the

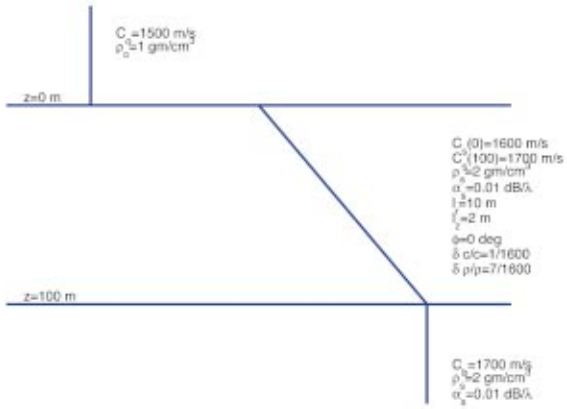


FIG. 11. Scenario 3 is an upward refracting sandy sediment layer, with a sediment background sound speed of 1600 m/s at the sediment water interface, 1700 m/s at the sediment-basement interface, density of 2 g/cm^3, and attenuation of 0.01 dB/λ. The basement has a sound speed of 1700 m/s and has properties which are otherwise identical to those of the sediment layer. The scatterer distribution in the sediment layer are identical to those in Scenario 2.

reverberation from the sediment volume inhomogeneities in the upward refracting sediment in Scenario 3 is significantly slower that for the iso-speed sediment in Scenario 2. The implication is that sediment gradients can have a very significant effect on both the levels and the decay characteristics of sediment scattering.

IV. CONCLUSIONS

The wave number integral representation for monostatic backscatter from three-dimensional sediment inhomogene-

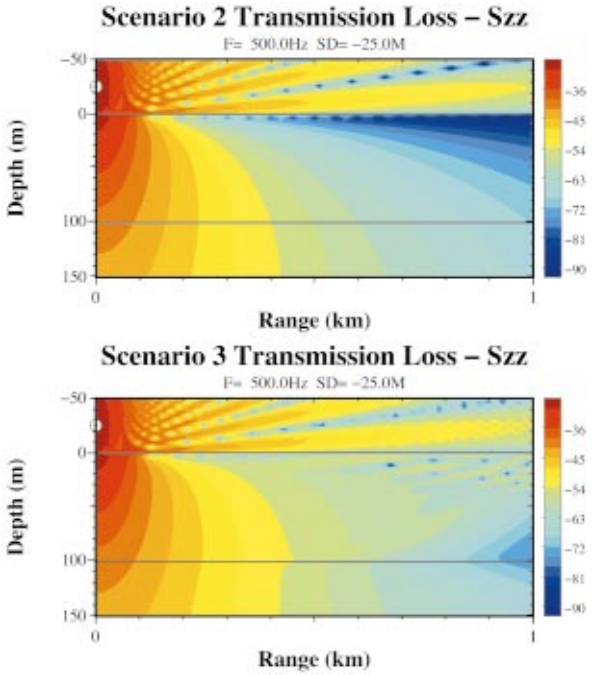


FIG. 12. The range-depth transmission loss for Scenarios 2 (top) and 3 (bottom) at 500 Hz. In Scenario 2 the energy passing into the bottom above the critical angle passes through the sediment basement interface and is lost to the sediment layer. In Scenario 3 the upward refracting sediment sound speed causes a significant amount of the incident energy to be retained at longer ranges from the source.

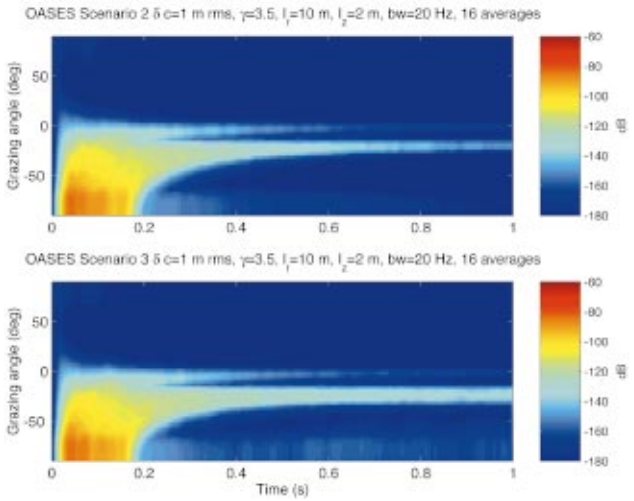


FIG. 13. Time-angle evolution of scattered intensity for the iso-speed fast bottom in Scenario 2 (upper plot) and the upward refracting sediment in Scenario 3 (lower plot). The results for Scenario 3 indicate that upward refracting sediments can have significantly more late time reverberation over a broader range of grazing angles than otherwise similar sediment layers without gradients.

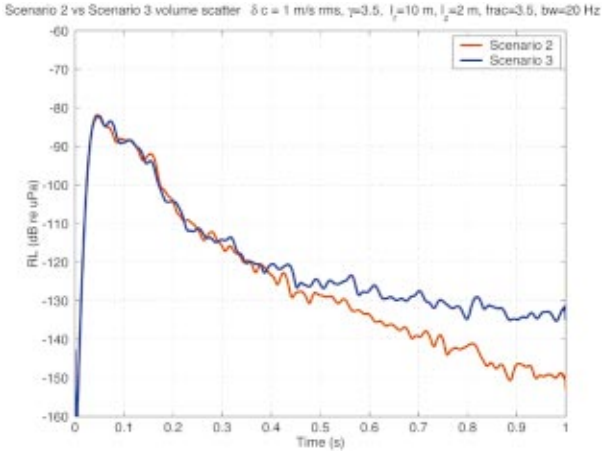


FIG. 14. The scattered intensity received in a monostatic configuration for a fast sediment without a sound speed gradient (Scenario 2, red curve), and a fast upward refracting sediment (Scenario 3, blue curve). The results indicate that sediment gradients can have a very significant effect on scattered levels from sediment volume inhomogeneities at late time.

ities has been derived using the method of small perturbations and the Born Approximation, and these equations have been implemented in a computationally efficient manner. The code offers the capability for evaluating Monte Carlo estimates of backscattered intensity and the time-angle evolution of backscattered intensity for isotropic, horizontally stratified sediment scenarios with critical angles and turning points where the scatterers are treated using the perturbation and single scattering approximations. The potential also exists for treating sediment volume inhomogeneity scattering in more complicated horizontally stratified bottoms by superimposing the single layer results derived and implemented here for all the contributing layers. For single layers, the results obtained here indicate that for slow sediments, the wave number integral estimates of backscattered intensity are in excellent agreement with those obtained by SCARAB using a simple sonar equation approach, as expected. However for fast and upward refracting sediments, the complete first order perturbation solution for the scattered field can only be obtained through the use of the wave number integral approach.

For fast sediments, it is found that a second scattering branch exists in the time-angle plane of scattered intensity. This second branch follows the time-angle evolution of scattering from the sediment-water interface, and is associated with the inhomogeneous excitation of the volume inhomogeneity scatterers closest to the sediment water interface. Since these scatterers are ensonified by nonpropagating waves, they have no defined farfield scattering cross section. Yet the results show that these scatterers can and do contribute to the backscattered signal in an important way. For inhomogeneities in fast smooth sediments with significant attenuation, this scattering mechanism is the only contributor to reverberation at late time.

For upward refracting sediments, it is shown that the intensification of the incident field on sediment inhomogeneities beyond the turning point of the basement grazing rays leads to increased backscattered levels at late time. This phenomenon is most likely to be observable in weakly lossy sediments, and is accurately modeled near ray caustics with the wave theoretic approach used here.

The numerical implementation of a benchmarked wave number integral formalism for three-dimensional sediment inhomogeneity scattering into the SAFARI/OASES framework makes it possible to evaluate the quality of the various approximations which are routinely taken for perturbation estimates of backscattered intensity from sediments. In this paper we have concentrated on evaluating the suitability of using approximate Green functions, but the code also can be used to evaluate the suitability of assuming that the vertical correlation length scale of the bottom inhomogeneities is infinitesimal.

ACKNOWLEDGMENTS

The authors wish to acknowledge Charles Holland, Peter Neumann, and Andrew Rogers for supplying the SCARAB code and Ivar Bratberg for his assistance modifying and running it. The MIT part of this work was funded by the Office of Naval Research, Ocean Acoustics Program.

APPENDIX: EXPRESSIONS FOR SCATTERED INTENSITY USED BY SCARAB

The SCARAB model was developed to address deficiencies of many of the then existing bottom scattering models. Holland and Neumann[8] departed from the commonly used local plane-wave approximation (which they showed was inadequate) by modeling the received level in the time domain with all of the experiment complexities, then calculated bottom scatter under exactly the same assumptions as the measured data were processed. This is the essence of the SCARAB model: to model as closely as possible the measurement conditions, including assumptions used in the data processing.

Invoking the Born Approximation, the received level from a single ping can be written as a sum of delayed, scaled, replicas of the transmit source intensity $S(t)$:

$$RL(T,\theta) = 10\log\left[\sum_{i=1}^{N} S(t-t_i)10^{TL_i^1/10} \times 10^{-TL_i^2/10}\frac{dP}{d\Omega}(\theta_i, z_i)\right], \quad \text{(A1)}$$

where TL_i^1 and TL_i^2 are the transmission losses (including the beam pattern) from the source to scattering element i and from element i back to the receiver respectively. For sub-bottom volume scattering $dP/d\Omega$ is the differential scattering cross-section of a scatterer i at depth z; N is the number of scatterers. In the initial version of the model used in Ref. 8, the propagation to and from the scatterers was computed using OASES and the scattering was calculated from randomly distributed 3-D ellipsoids. A later version of the model employs a number of simplifications including use of ray theory for the propagation and replacing the scattering from randomly distributed deterministic scatterers with a stochastic approach.[7] It is this simple version of the model that is used in the comparisons in this paper.

[1] A. N. Ivakin and Y. P. Lysanov, "Theory of underwater sound scattering by random inhomogeneities of the bottom," Sov. Phys. Acoust. **27**, 61–64 (1981).
[2] A. N. Ivakin, "A unified approach to volume and roughness scattering," J. Acoust. Soc. Am. **103**, 827–837 (1998).
[3] D. Tang, "Acoustic Wave Scattering from a Random Ocean Bottom," Doctoral Dissertation, MIT-WHOI Joint Program, WHOI-91-25, June 1991.
[4] P. D. Mourad and D. R. Jackson, "A model/data comparison for low frequency bottom backscatter," J. Acoust. Soc. Am. **94**, 344–358 (1993).
[5] B. H. Tracey and H. Schmidt, "A self-consistent theory for seabed volume scatter," J. Acoust. Soc. Am. **106**, 2524–2534 (1999).
[6] T. Yamamoto, "Acoustic scattering in the ocean from velocity and density fluctuations in the sediments," J. Acoust. Soc. Am. **99**, 866–879 (1996).
[7] A. Rogers and C. Holland, "SCARAB version 1.01, Users Manual," August 1997, Planning Systems Incorporated.
[8] C. W. Holland and P. Neumann, "Sub-bottom scattering: A modeling approach," J. Acoust. Soc. Am. **104**, 1363–1373 (1998).
[9] P. C. Hines, "Theoretical model of acoustic backscatter from a smooth seabed," J. Acoust. Soc. Am. **88**, 324–334 (1990).
[10] P. C. Hines, "Theoretical model of in-plane scatter from a smooth sediment seabed," J. Acoust. Soc. Am. **99**, 836–844 (1996).
[11] K. D. LePage and H. Schmidt, "Spectral integral representations of volume scattering in sediments in layered waveguides," J. Acoust. Soc. Am. **108**, 1557–1567 (2000).
[12] H. Schmidt and K. D. LePage, "Spectral integral representations of multistatic scattering from sediment volume inhomogeneities," J. Acoust. Soc. Am. **108**, 2564 (2000).
[13] A. Turgut, "Inversion of bottom/subbottom statistical parameters from acoustic backscatter data," J. Acoust. Soc. Am. **102**, 833–852 (1997).
[14] H. Schmidt and J. Lee, "Physics of 3-D scattering from rippled seabeds and buried targets in shallow water," J. Acoust. Soc. Am. **105**, 1605–1617 (1999).
[15] H. Schmidt and J. Glattetre, "A fast field model for three-dimensional wave propagation in stratified environments based on the global matrix method," J. Acoust. Soc. Am. **78**, 2105–2114 (1985).
[16] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, Eq. 4.93, *Computational Ocean Acoustics* (Springer-Verlag, New York, 2000).
[17] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, Eq. 2.103, *Computational Ocean Acoustics* (Springer-Verlag, New York, 2000).
[18] F. DiNapoli and R. Deavenport, "Theoretical and numerical Green's function solution in a plane multilayered medium," J. Acoust. Soc. Am. **67**, 92–105 (1980).
[19] J. A. Ferrari, D. Perciante, and A. Dubra, "Fast Hankel transform of nth order," J. Opt. Soc. Am. A **16**, 2581–2582 (1999).
[20] Q. H. Liu and W. C. Chew, "Applications of the conjugate gradient fast Fourier Hankel transfer method with an improved fast Hankel transform algorithm," Radio Sci. **29**, 1009–1022 (1994).
[21] Q. H. Liu and Z. Q. Zhang, "Nonuniform fast Hankel transform (NUFHT) algorithm," Appl. Opt. **38**, 6705–6708 (1999).
[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[23] H. Schmidt, "SAFARI: Seismo-acoustic fast field algorithm for range independent environments. User's Guide," SR-113, SACLANT Undersea Research Centre, La Spezia, Italy (1988) (AD A 200 581).
[24] J. A. Goff and T. H. Jordan, "Stochastic modeling of seafloor morphology: Inversion of sea beam data for second order statistics," J. Geophys. Res. **93**, 13589–13608 (1988).
[25] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1988).
[26] A. Maguer, W. L. J. Fox, H. Schmidt, E. Pouliquen, and E. Bovio, "Mechanisms for subcritical penetration into a sandy bottom: Experimental and modeling results," J. Acoust. Soc. Am. **107**, 1215–1225 (2000).
[27] A. Maguer, E. Bovio, W. L. J. Fox, and H. Schmidt, "*In situ* estimation of sediment sound speed and critical angle," J. Acoust. Soc. Am. **108**, 987–996 (2000).
[28] C. W. Holland and J. Osler, "High resolution geoacoustic inversion in shallow water: A joint time and frequency domain technique," J. Acoust. Soc. Am. **107**, 1263–1279 (2000).
[29] C. W. Holland, R. Hollett, and L. Troiano, "Measurement technique for bottom scattering in shallow water," J. Acoust. Soc. Am. **108**, 997–1011 (2000).

Modal analysis of broadband acoustic receptions at 3515-km range in the North Pacific using short-time Fourier techniques

Kathleen E. Wage[a)]
Department of Electrical and Computer Engineering, George Mason University, Fairfax, Virginia 22030

Arthur B. Baggeroer
Departments of Ocean and Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

James C. Preisig
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 3 May 2001; revised 3 September 2002; accepted 23 September 2002)

In 1995–1996 the Acoustic Thermometry of Ocean Climate (ATOC) experiment provided an opportunity to study long-range broadband transmissions over a series of months using mode-resolving vertical arrays. A 75-Hz source off the California coast transmitted broadband pulses to receiving arrays in the North Pacific, located at ranges of 3515 and 5171 km. This paper develops a short-time Fourier transform (STFT) processor for estimating the signals propagating in the lowest modes of the ocean waveguide and applies it to analyze data from the ATOC experiment. The STFT provides a convenient framework for examining processing issues associated with broadband signals. In particular, this paper discusses the required frequency resolution for mode estimation, analyzes the broadband performance of two standard modal beamforming algorithms, and explores the time/frequency tradeoffs inherent in broadband mode processing. Short-time Fourier analysis of the ATOC receptions at 3515 km reveals a complicated arrival structure in modes 1–10. This structure is characterized by frequency-selective fading and a high degree of temporal variability. At this range the first ten modes have equal average powers, and the magnitude-squared coherence between the modes is effectively zero. The coherence times of the peaks in the STFT mode estimates are on the order of 5.5 min. An analysis of mean arrival times yields modal dispersion curves and indicates that there are statistically significant shifts in travel time over 5 months of ATOC transmissions. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1530615]

PACS numbers: 43.30.Qd, 43.30.Re, 43.30.Bp, 43.60.Gk [DLB]

I. INTRODUCTION

In deep water, the lowest-order acoustic normal modes are associated with the most energetic late arrivals at long range. Numerous authors, notably Munk and Wunsch,[1] have suggested using these arrivals in applications such as tomography and matched field processing. Using mode signals as observables requires the ability to associate a mode arrival with a particular path or section of the water column. In range-invariant environments, this problem is trivial because the modes propagate independently without exchanging energy, i.e., an arrival in mode 1 is known to have traversed the entire path in mode 1. For ranges on the order of megameters, however, inhomogeneities such as internal waves cause significant coupling of energy among the modes, resulting in complicated arrival patterns that are difficult to interpret. To date, tomographers have primarily relied on the earlier-arriving wavefronts, rather than the low-mode signals, in long-range experiments. Internal-wave-induced fluctuations associated with the rays can be studied within the context of an existing theory,[2] whereas no comparable framework exists for the modes. Theoretical[3,4] and empirical[5,6] investigations of the long range propagation of modes through internal waves have yielded interesting results, but there have been few opportunities to relate them to experimental measurements of the mode arrival structure.

In 1995–1996 the Acoustic Thermometry of Ocean Climate (ATOC) experiment provided the first opportunity to observe broadband receptions over a period of months using mode-resolving vertical line arrays (VLAs). Two VLAs were part of a large network that also included U.S. Navy Sound Surveillance System (SOSUS) arrays.[7] Figure 1 shows paths from the bottom-mounted ATOC source on Pioneer Seamount to the two VLAs, located near Hawaii and Kiritimati (Christmas Island). Ranges to these arrays were 3515 and 5171 km, respectively. Each of the vertical arrays had 40 elements and spanned an aperture of 1400 m, providing adequate sampling of the first ten modes of the local environment. The source transmitted phase-encoded pseudo-random sequences with a center frequency of 75 Hz and bandwidth of 37.5 Hz. Transmissions occurred at 4-h intervals during periods set by the ATOC Marine Mammal Research Program.

[a)]Formerly at Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, and Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution.

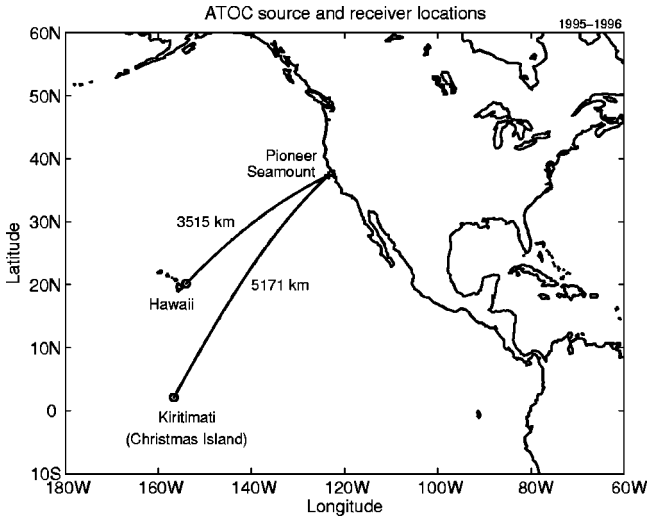


FIG. 1. ATOC source and vertical line array receivers.

This paper develops a short-time Fourier transform (STFT) method for estimating broadband signals propagating in the lowest modes of the deep water channel and analyzes receptions from the ATOC VLA at Hawaii using this approach. Since the modes are a frequency-dependent basis, broadband mode estimation requires separating the signal into frequency bins and using narrow-band mode filtering for each bin. Typically, researchers have implemented the required frequency decomposition using either a single long Fourier transform or a set of bandpass filters. These approaches are both special cases of the short-time Fourier transform. The STFT provides a convenient framework for exploring the time-frequency resolution tradeoffs that have not been addressed in previous work.

Short-time Fourier analysis of the Hawaii data set provides a detailed characterization of the low mode arrival structure at megameter range. In particular this paper quantifies the coherence of the first ten modes, estimates average dispersion curves, and examines trends in arrival time over the course of 5 months of ATOC transmissions.

The rest of the article is organized as follows. The next section reviews relevant aspects of long-range mode propagation, using simulations of the Pioneer–Hawaii path to illustrate the impact of internal waves on the mode arrivals. Section III reviews the broadband mode estimation problem and highlights the important issues associated with estimating broadband signals. Following that, Sec. IV presents a short-time Fourier framework for broadband mode estimation. Section V describes the results of the STFT analysis of the Hawaii data set. A summary concludes the paper.

II. LONG-RANGE MODE PROPAGATION

Normal modes are the eigenfunctions of the ocean waveguide, derived from the frequency domain acoustic wave (Helmholtz) equation.[8] For a given environment and frequency, the mth mode is characterized by a horizontal wavenumber k_m, which defines its phase and group velocity, and a modeshape ϕ_m, which defines its vertical structure. The modes are an orthonormal basis for narrowband signals, thus the acoustic pressure at frequency Ω, range r, and depth z can be represented as the weighted sum

$$p(r,z,\Omega)=\sum_m a_m(r,\Omega)\phi_m(r,z,\Omega), \tag{1}$$

where a_m is the frequency-dependent coefficient for mode m. From a simple input/output viewpoint, the underwater channel transforms the mode coefficients at the source into a set of coefficients at the receiver. In an adiabatic waveguide, the modes propagate independently without exchanging energy. For typical deep water channels, the low modes travel slowest since they represent energy trapped around the sound speed minimum. In addition, modal group velocities in these channels usually decrease with frequency.

When the medium is nonadiabatic, the modes exchange energy as they propagate. Range-dependent waveguides are often modeled as a cascade of range-independent segments, with the boundary conditions at the segment interfaces determining the mode coupling coefficients. Assuming that an environment is only weakly range-dependent, the adiabatic approximation simplifies the modeling problem by neglecting the coupling terms. Under this assumption, each propagating mode adapts with range (changes shape and wavenumber), but does not transfer energy into other modes. The validity of the adiabatic approximation is related to the nature of the inhomogeneities in the medium, and Desaubies concluded that its accuracy depends strongly on frequency, mode number, range and the acoustic quantity of interest, e.g., intensity, phase, travel time.[9,10]

Sound speed fluctuations due to internal waves are the dominant source of mode coupling in long-range propagation scenarios. To understand the impact of internal waves on the mode arrivals, consider two simulations for the 3515 km California–Hawaii path of the ATOC experiment. Figure 2 shows the results of a broadband parabolic equation (PE) simulation through a deterministic, range-varying model of the environment. This model was defined using sound speed profiles derived from the Levitus winter climatology[11,12] and bathymetry from the ETOPO-5[13] topography database.[14] The top plot in Fig. 2 is the synthesized pressure time series at the Hawaii array location, generated using the RAM PE code.[15] Spatial patterns associated with individual mode arrivals are evident in the pressure field, e.g., modes 2 and 3 (mode 1 is not strongly excited in this simulation). The figure beneath the pressure plot is the modal time series obtained by projecting the field (finely sampled in depth) onto the mode functions at the receiver. Note that the modes arrive in order from highest to lowest, as is consistent with deep water dispersion. Constructive interference of the higher modes results in the planewave (ray) arrivals in the early part of the reception.

Figure 3 illustrates how the results change when internal waves are present. For this simulation, the background sound speed profiles were perturbed by internal wave fluctuations at $\frac{1}{2}$ Garrett-Munk strength, computed using the method of Colosi and Brown.[16] Instead of a single, dispersive arrival in each mode, there are multiple arrivals. This "modal multi-

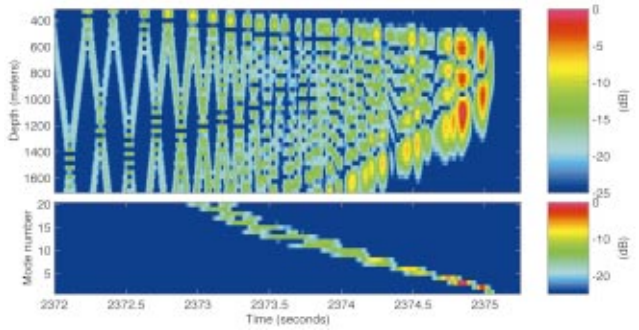


FIG. 2. Deterministic, range-varying simulation for California-ATOC path. Top panel shows the synthesized pressure time series, $20\log_{10}|p(t)|$; bottom panel shows the corresponding time series, $20\log_{10}|a(t)|$, for the first 20 modes.

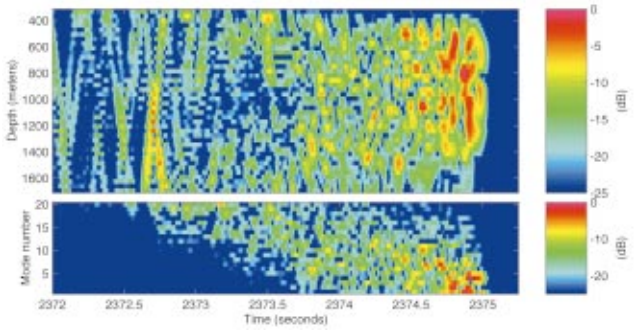


FIG. 3. Simulation for California–Hawaii environment perturbed by internal waves ($\frac{1}{2}$ GM level). Top plot: pressure time series, $20\log_{10}|p(t)|$; bottom plot: modal time series, $20\log_{10}|a(t)|$.

path'' creates the more complicated interference patterns seen in the pressure time series.

From a theoretical standpoint, the effects of internal waves on long-range sound propagation are not fully understood. As indicated in the introduction, most previous work focused on the ray arrivals because they are amenable to analysis via a geometrical optics approximation. The monograph by Flatte *et al.* summarizes the path integral theory that predicts the fluctuations and coherence of resolved rays.[2] While there is no comparable theory for predicting the behavior of the mode arrivals, several authors have addressed aspects of the mode propagation problem. Some of the key results are described below.

In two seminal papers Dozier and Tappert derived modal intensity statistics for narrow-band signals in the presence of internal waves.[3,6] Their theory and numerical simulations showed that internal-wave-induced scattering eventually results in an equipartition of energy among the modes. The derivation of this result depends on several simplifying assumptions, including one that says there is no loss of energy into the bottom.

In one of the few studies of long-range experimental data, Colosi *et al.* compared pressure measurements from the 1000-km SLICE89 experiment to broadband PE simulations.[17] Their results showed that the broadening of the transmission finale in the data is attributable to the exchange of energy among the modes, caused by internal waves. In a later paper Colosi and Flatte explored the subject of mode coupling via internal waves using PE simulations designed to model certain aspects of the ATOC experiment.[5] They demonstrated the strong nonadiabatic character of propagation through these random fields and quantified the travel-time bias/spread and intensity fluctuations for the modes. A recent review article by Colosi *et al.* indicates that internal-wave-induced mode coupling is a major factor at 75 Hz, but may be significantly reduced at lower frequencies, e.g., 28 Hz.[18] Several papers have examined the degradation of mode coherence by internal waves concentrating on the implications for various signal processing methods, e.g., matched filtering,[19] horizontal array beamforming,[20,21] and vertical array beamforming.[22] Sazontov developed an approximate analytic method for computing the modal cross-coherences,[23] and Gorodetskaya *et al.* applied this technique to the study of horizontal and vertical array gain limitations due to internal wave fluctuations.[4] These approximate expressions for coherence have not yet been validated by experimental data.

The ATOC experiment provided a significant opportunity to learn more about mode propagation out to megameter ranges. This paper examines the mode arrivals in the Hawaii data set, focusing in particular on the broadband characteristics and temporal stability of these signals. The following section reviews the problem of estimating mode arrivals using vertical arrays, specifically highlighting the issues associated with using broadband signals.

III. BROADBAND MODE ESTIMATION PROBLEM

In many experiments, including ATOC, the low-order modes are not temporally resolvable, meaning that a vertical array is required to separate the mode signals based on their spatial characteristics.[24] An array measures the sum of modes (associated with the signal) plus noise, i.e., in vector notation,

$$\mathbf{p}[r,\Omega]=\mathbf{\Phi}[r,\Omega]\mathbf{a}[r,\Omega]+\mathbf{n}[\Omega]. \tag{2}$$

$\mathbf{\Phi}$ is the matrix of sampled modeshapes, $\mathbf{a}$ is the vector of mode amplitudes, and $\mathbf{n}$ is the vector of observation noise. For a broadband source, the measurement is a vector time series

$$\mathbf{\Psi}(r,t)=\int_{\Omega}(\mathbf{\Phi}[r,\Omega]\mathbf{a}[r,\Omega]+\mathbf{n}[\Omega])e^{j\Omega t}\,d\Omega. \tag{3}$$

The objective of mode processing is to estimate the vector of mode coefficients ($\mathbf{a}[r,\Omega]$), or equivalently to estimate the corresponding mode time series. Section III A discusses two important design criteria for mode filters, and Sec. III B reviews previous work on this topic.

A. Design issues

The most important issue to consider in broadband mode estimation is the frequency dependence of the modeshapes. As an example, Fig. 4 illustrates how the first ten modes at the ATOC Hawaii array vary as a function of frequency. A comparison of the modeshapes at 60 and 90 Hz indicates that they change significantly over the 30-Hz interval (approximate bandwidth of the ATOC source). Obviously, any spatial

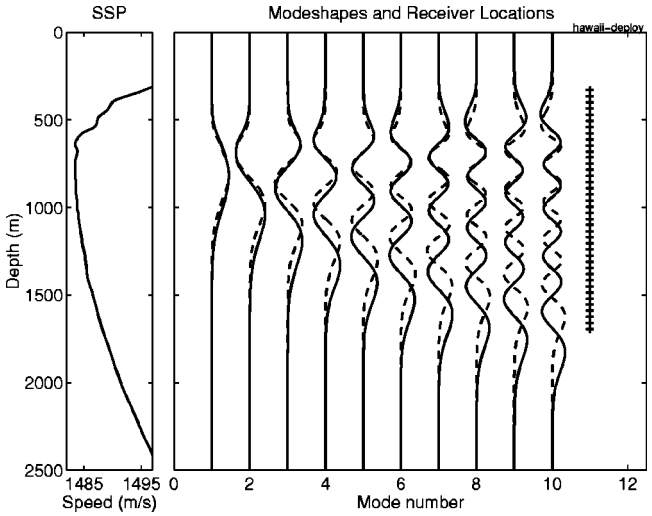


FIG. 4. Measured sound speed profile and first ten modeshapes (at 60 and 90 Hz) for the ATOC Hawaii environment. Water depth at array location is 5426 m; only the top 2500 m are shown in the plot. The + 's indicate nominal sensor positions for the 40-element VLA.

processing that requires a replica of the sampled mode shape must be done on a set of subbands to avoid mismatch problems. In general the maximum width of these bands is determined by the environment and center frequency. Since bandpass filtering smears signals in time, the broadband mode filter design process inherently involves time and frequency resolution tradeoffs.

Separating the signal into subbands reduces mode estimation to a classical linear inverse problem.[25,26] The key issue to consider in solving each of the narrowband problems is the spatial sampling of the modeshapes by the array since that determines how well the processor can resolve a mode and reject noise.[27] From the point of view of estimating a single mode, there are two types of noise to consider: structured interference from signals propagating in other modes and uncorrelated measurement noise (due to ships, etc.). Based on Fig. 3, it is reasonable to assume that time-windowing can be used to limit the structured interference from the earliest ray arrivals (corresponding to high order modes), since they do not overlap in time with the lowest modes.

B. Previous work

Most previous research focused on experimental settings where a narrowband assumption is valid, meaning that either the source is continuous wave (CW) or the variations in the mode functions across the source band are negligible. Two basic types of mode filters have been developed for narrowband signals: the matched filter (MF)[28,29] and the pseudo-inverse (PI) filter.[30] The performance of these filters is well understood. The matched filter (sometimes called the sampled modeshapes filter) has the advantage of a simple and stable implementation, but it does not guarantee that energy from one mode will not leak into an adjacent mode. On the other hand, the PI filter ensures good modal crosstalk rejection at the expense of increased noise sensitivity. Headrick *et al.* provide a useful discussion of the matched filter in the context of the shallow water SWARM 95 internal wave scattering experiment.[31] They assess the variability in cross-mode rejection due to array tilt and temporal fluctuations of the modeshapes. SWARM 95 used a 400-Hz broadband source (100-Hz bandwidth), but the authors conclude that a narrowband mode filter is sufficient because the modes do not change significantly across the source band.

For some broadband experiments, variations in the modeshapes and wavenumbers as a function of frequency are too large to ignore. The standard solution to the broadband problem consists of separating the signal into frequency bins using a single long Fourier transform, doing narrow-band mode processing for each bin, and obtaining a time series via an inverse transform.[32–34] Sutton *et al.*[35] and Heaney and Kuperman[36] suggest using a windowed Fourier transform to compute the mode estimates. Others propose using bandpass filters to process the data in bands where the modeshapes may be assumed constant.[37] None of these studies discuss how the frequency resolution (determined by the length of the window or the width of the bandpass filter) affects the mode estimates.

This paper generalizes the previous approaches by using a short-time Fourier transform (STFT) framework to analyze the time/frequency tradeoffs inherent in broadband mode estimation. An STFT-based processor separates the signal into a set of subbands and estimates the modal time series in each band. The results are time-varying mode spectra that can be used to examine the frequency-dependent structure in the signals, e.g., to quantify dispersion or frequency-selective fading. An important advantage of the STFT approach is that it provides a method of analyzing the characteristics of individual multipath arrivals within a mode, provided that short enough time windows can be used in the processing. This facilitates the search for frequency-coherent mode arrivals. The extent of the modeshape variations with frequency determines the maximum temporal resolution that is attainable with a mode processor for a particular environment.

The following section develops the short-time Fourier mode processing framework, explores important time/frequency resolution issues, and selects the processing parameters for the ATOC Hawaii experimental dataset.

IV. SHORT-TIME FOURIER MODE PROCESSING FRAMEWORK

The short-time Fourier transform is a standard signal processing technique for examining the characteristics of transient or time-varying signals.[38,39] STFT analysis consists of computing discrete Fourier transforms for a sequence of finite-length data segments. There are two equally valid interpretations of the resulting time-dependent spectrum: (1) as the output of a filterbank or (2) as the output of a windowed Fourier transform operation. This section relies on the first interpretation to describe the application of STFT techniques to broadband mode estimation. The discussion is organized as follows. Section IV A provides an overview of the short-time Fourier mode processor. Following that, Secs. IV B and IV C discuss narrowband mode filters and their broadband

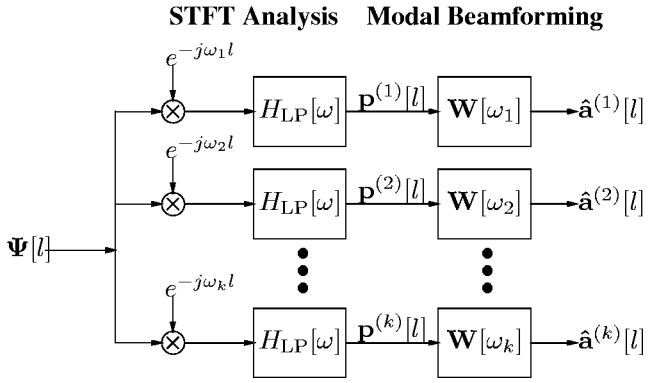


FIG. 5. Block diagram of STFT-based mode processor.

performance characteristics, respectively. Finally, Sec. IV D illustrates the properties of short-time mode estimates using an adiabatic propagation example.

A. Overview

In STFT-based mode analysis, the processor separates the received pressure into a set of subbands and computes mode estimates for each subband. Figure 5 illustrates these steps and introduces some notation. The input to the filterbank is $\mathbf{\Psi}[l]$, a sampled vector time series from an N-element receiving array. As shown, the filtering operation consists of complex demodulation followed by lowpass filtering; this is equivalent to bandpass filtering followed by demodulation. The result of the STFT analysis step is an N-point complex vector time series of pressures for each band: $\mathbf{p}^{(k)}[l]$, where k denotes the band (bin) number and l is the time index. After filtering, the processor computes a vector time series of mode coefficients for each band using a set of narrow-band modal beamformers. The result is an M-point vector $\hat{\mathbf{a}}^{(k)}[l]$ containing the estimated short-time Fourier transform of the first M modes in the bin centered around ω_k.[40]

Within the STFT framework, the length of the lowpass filter, $H_{\mathrm{LP}}[\omega]$, determines the time and frequency resolution of the estimates. Long filters (equivalent to using long data windows for the Fourier transform) have good frequency resolution, implying that the required operating bandwidth for the modal beamformer $\mathbf{W}[\omega_k]$ is small. The disadvantage of long filters is that they smear the arrivals in time. Short filters provide much better temporal resolution, but they have wider passbands, meaning that the processor may be more sensitive to the frequency-dependent variations of the modeshapes. In general H_{LP} can be any filter with a lowpass characteristic, but short-time Fourier analysis typically employs the same filters (windows) that are used in spectral estimation. Harris provides an extensive list of window functions in his classic paper.[41]

The STFT approach reduces the broadband estimation problem to a set of narrowband problems. Assuming that the modeshapes are constant over the band defined by the lowpass filter, the pressure measurement in the kth band becomes

$$\mathbf{p}^{(k)}[l]=\mathbf{\Phi}[\omega_k]\mathbf{a}^{(k)}[l]+\mathbf{n}^{(k)}[l], \tag{4}$$

where $\mathbf{\Phi}[\omega_k]$ is the matrix of sampled modeshapes at the band's center frequency, $\mathbf{a}^{(k)}$ is a vector of bandpass-filtered, time-varying mode coefficients, and $\mathbf{n}^{(k)}$ is a noise vector. As noted above, estimating the mode amplitudes from the measured pressure is a classic inverse problem. This paper focuses on solutions of the form

$$\hat{\mathbf{a}}^{(k)}[l]=\mathbf{W}_k^{\mathrm{H}}\mathbf{p}^{(k)}[l], \tag{5}$$

where $\mathbf{W}_k^{\mathrm{H}}$ is a matrix containing the deterministic, time-invariant spatial filter for the kth band (the superscript H denotes the conjugate transpose operator). Time- and data-adaptive solution methods could be incorporated into the STFT framework once more is known about the structure of mode signals at long ranges.

The next section briefly describes standard narrowband mode filters, and a subsequent section reviews their broadband performance.

B. Narrow-band mode filters

As indicated in Sec. III B, the two standard solutions to the narrow-band mode estimation problem are the matched filter and the pseudo-inverse mode filter. There are several ways to derive these filters. The approach described below is based on optimizing array gain and provides a complementary perspective to the estimation theory derivation that is more common in the mode filtering literature (see Ref. 27 for a summary). An advantage to viewing mode filtering as a constrained optimization problem is that it emphasizes the inherent tradeoff between interference rejection and processor sensitivity.

Array gain represents the improvement in the signal-to-noise ratio (SNR) due to processing. It is typically defined as the ratio of the SNR at the output of a beamformer to the SNR at a single sensor. Since the signal and noise characteristics often vary across an array, the input SNR is taken to be an average (arithmetic or geometric) of the single-phone SNRs. In the case of modal beamforming, the signal levels vary from one sensor to another because the modeshapes are functions of depth. White noise gain, G_w, represents the gain of the processor when the noise is assumed to be spatially white. For the mode processing problem, the gain for mode m is defined as

$$G_w=N\frac{|\mathbf{w}_m^{\mathrm{H}}\boldsymbol{\phi}_m|^2}{|\mathbf{w}_m|^2|\boldsymbol{\phi}_m|^2}, \tag{6}$$

where $\mathbf{w}_m$ is the weight vector (filter) and $\boldsymbol{\phi}_m$ is the sampled modeshape vector for the mth mode. Equation (6) assumes that the input SNR is equal to the arithmetic average SNR across the array. Application of the Schwartz inequality shows that the maximum value of the white noise gain is N, the number of sensors in the array. In addition to describing the noise response, G_w provides a useful measure of the sensitivity of the processor to mismatch, as discussed by Cox *et al.*[42]

The matched filter (MF) results from choosing the weight vector for mode m that maximizes white noise gain while maintaining a unit gain in the desired mode. Maximizing G_w subject to a unity gain constraint is mathematically

equivalent to minimizing the squared length of the weight vector subject to the same gain constraint, thus the optimization problem becomes

$$\min|\mathbf{w}_m|^2 \quad \text{subject to} \quad \mathbf{w}_m^{\mathrm{H}}\boldsymbol{\phi}_m = 1. \tag{7}$$

Standard optimization techniques yield the following solution,

$$\mathbf{w}_m{}^{\mathrm{H}} = \frac{1}{|\boldsymbol{\phi}_m|^2}\boldsymbol{\phi}_m^{\mathrm{H}}, \tag{8}$$

or in terms of the weight matrix for M modes

$$\mathbf{W}^{\mathrm{H}} = \begin{bmatrix} \frac{1}{|\boldsymbol{\phi}_1|^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{|\boldsymbol{\phi}_M|^2} \end{bmatrix} \mathbf{E}^{\mathrm{H}}, \tag{9}$$

where $\mathbf{E}$ is the sampled modeshape matrix containing the M desired modes, i.e., the first M columns of $\boldsymbol{\Phi}$. The matched filter is optimal in the sense that it achieves the maximum white noise gain ($G_{w-\mathrm{MF}} = N$), but it does not explicitly prevent the signal in one mode from leaking into another. Instead, it relies on the orthogonality of the modes to separate them. It is important to note that while the modeshapes are orthogonal functions of the continuous depth variable z, the sampled modeshape vectors are not guaranteed to have this property.

The pseudo-inverse (PI) filter results from constraining mode leakage by placing nulls in the modal beampattern at the locations of a set of interfering modes. In this case, the optimization problem consists of maximizing the white noise gain (minimizing the weight vector length) subject to multiple constraints, i.e.,

$$\min|\mathbf{w}_m|^2 \quad \text{subject to} \quad \begin{cases} \mathbf{w}_m^{\mathrm{H}}\boldsymbol{\phi}_m = 1, \\ \mathbf{w}_m^{\mathrm{H}}\boldsymbol{\phi}_{n\neq m} = 0, \quad 1 \leqslant n \leqslant M. \end{cases} \tag{10}$$

It is useful to rewrite the problem as

$$\min|\mathbf{w}_m|^2 \quad \text{subject to} \quad \mathbf{w}_m^{\mathrm{H}}\mathbf{E} = \mathbf{c}_m^{\mathrm{T}}, \tag{11}$$

where $\mathbf{E}$ contains the first M columns of the sampled modeshape matrix and $\mathbf{c}_m$ is an M-point column vector with a one in the mth position and zeros everywhere else. When the sampled modeshape matrix is full rank, standard optimization techniques yield the following solution for the weight vector

$$\mathbf{w}_m^{\mathrm{H}} = \mathbf{c}_m^{\mathrm{T}}(\mathbf{E}^{\mathrm{H}}\mathbf{E})^{-1}\mathbf{E}^{\mathrm{H}}. \tag{12}$$

Equation (12) corresponds to one row of the pseudo-inverse of the sampled modeshapes matrix containing the first M modes, thus $\mathbf{W}^{\mathrm{H}}$ is simply

$$\mathbf{W}^{\mathrm{H}} = (\mathbf{E}^{\mathrm{H}}\mathbf{E})^{-1}\mathbf{E}^{\mathrm{H}}. \tag{13}$$

Assuming that $\mathbf{E}$ has full rank, the null constraints in Eq. (11) are met exactly and the processor for mode m rejects the $M-1$ other modes included in the pseudo-inverse. In terms of interference rejection, the PI filter guarantees better performance than the matched filter, but this improvement may come at the expense of increased sensitivity to noise or other perturbations. Consider the white noise gain for the mth mode:

$$G_{w-\mathrm{PI}} = \frac{N}{|\boldsymbol{\phi}_m|^2} \cdot \frac{1}{\mathbf{c}_m^{\mathrm{T}}(\mathbf{E}^{\mathrm{H}}\mathbf{E})^{-1}\mathbf{c}_m}. \tag{14}$$

Recalling that $\mathbf{c}_m^{\mathrm{T}}\mathbf{c}_m$ is equal to unity, the second term in Eq. (14) can be written as the inverse of a Rayleigh quotient[43] and is therefore known to be bounded by the squared singular values of the sampled modeshape matrix, i.e.,

$$\sigma_{\min}^2 \leqslant \frac{\mathbf{c}_m^{\mathrm{T}}\mathbf{c}_m}{\mathbf{c}_m^{\mathrm{T}}(\mathbf{E}^{\mathrm{H}}\mathbf{E})^{-1}\mathbf{c}_m} \leqslant \sigma_{\max}^2, \tag{15}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are the minimum and maximum singular values of $\mathbf{E}$, respectively. If the sampled modeshapes are orthogonal, $\mathbf{E}^{\mathrm{H}}\mathbf{E}$ is a diagonal matrix where the mth term on the diagonal is equal to $|\boldsymbol{\phi}_m|^2$. In this case the quotient in Eq. (15) reduces to $|\boldsymbol{\phi}_m|^2$, and the white noise gain is equal to the optimal value of N. On the other hand, if the array does not adequately sample the modes, the modeshape vectors are

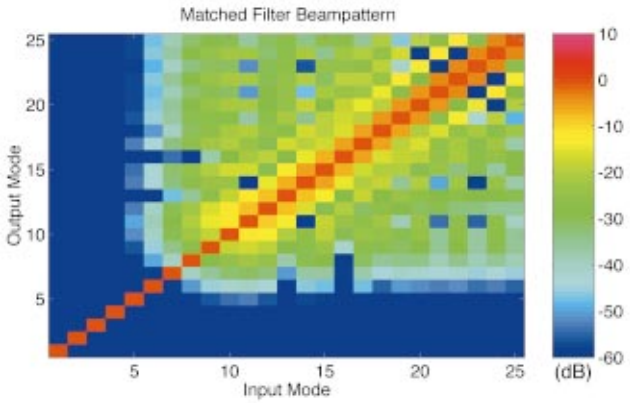


FIG. 6. Matched filter beampattern (75 Hz) for the 40-element ATOC VLA at Hawaii.

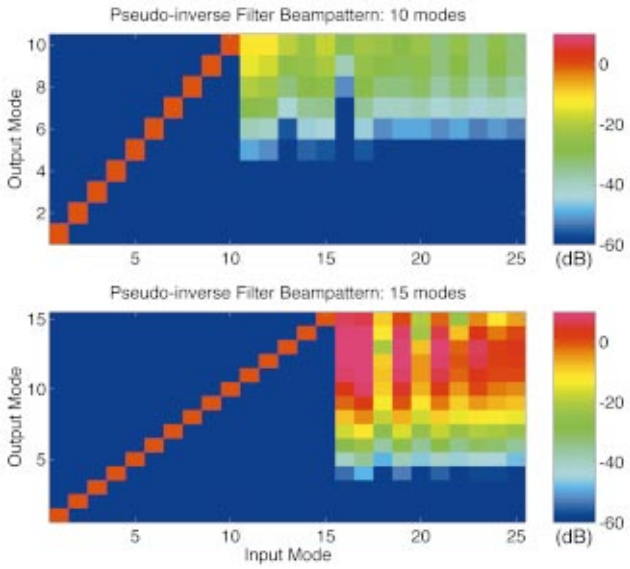


FIG. 7. Comparison of 10- and 15-mode pseudo-inverse filter beampatterns (75 Hz) for the ATOC VLA at Hawaii.

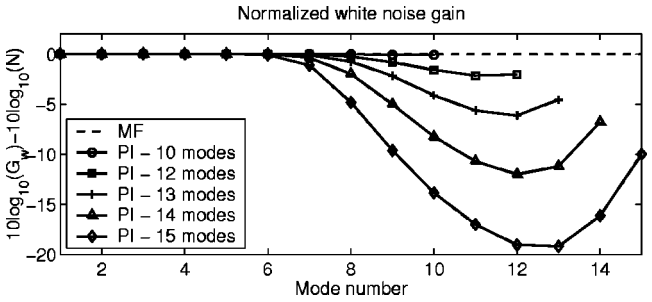


FIG. 8. Comparison of white noise gain for the matched filter and 10-, 12-, 13-, 14-, and 15-mode pseudo-inverse filters at 75 Hz for the ATOC Hawaii VLA. G_w is normalized by N, the maximum gain for an array with N hydrophones. Results are shown in dB.

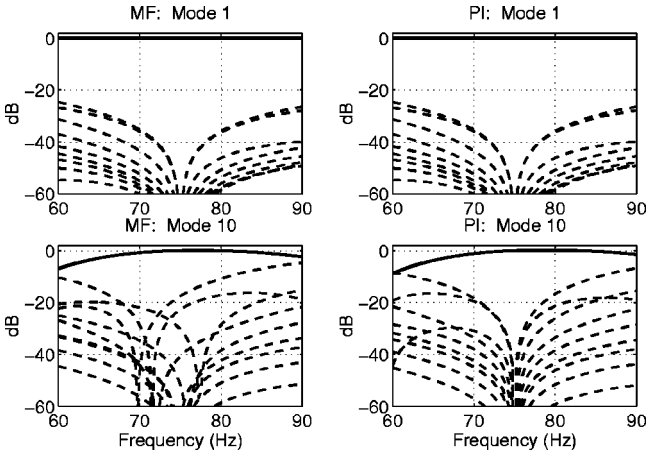


FIG. 9. Frequency-dependent beampatterns for the matched filter and 10-mode pseudo-inverse filter. The filters are designed for the 75-Hz bin using the ATOC Hawaii VLA modeshapes.

not orthogonal and some singular values of $\mathbf{E}$ may be very small. As one or more of the singular values approach zero, the white noise gain is substantially reduced and the pseudo-inverse becomes ill-conditioned. There are a number of standard techniques for mitigating problems with small singular values in the pseudo-inverse, e.g., diagonal loading[44] and elimination of small eigenvalues in the inverse.[45] These methods increase robustness at the expense of biasing the mode estimates. For the ATOC analysis we limit the number of modes in the pseudo-inverse rather than using one of these alternative approaches.

Beampatterns provide a useful illustration of the narrowband performance of these two mode processors. In the context of mode filtering, the beampattern is defined as $20\log_{10}(\mathbf{W}^{\mathrm{H}}\mathbf{\Phi})$, where $\mathbf{W}$ is the multidimensional mode filter and $\mathbf{\Phi}$ is the matrix of sampled modeshapes. The mth row of the beampattern matrix corresponds to the projection of the modes into the estimate for mode m.

For the matched filter, the beampattern corresponds to a normalized version of the sampled modeshape correlation matrix, $\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}$. Figure 6 shows the matched filter beampattern for the first 25 modes at 75 Hz, using the 40-element ATOC VLA at Hawaii. This filter has excellent crosstalk rejection for modes up to 8; above 8 the sampled modeshapes are obviously correlated. As a result, energy from one mode leaks into estimates of adjacent modes. Performance of the matched filter in this environment degrades significantly for modes above 10, which is not surprising since the array was designed to spatially resolve the first 10 modes.

Figure 7 shows the 75-Hz beampatterns for two pseudo-inverse filters, designed for 10 modes and 15 modes, respectively. These plots confirm that the PI filter for mode m has nulls at the locations of all the other modes included in the estimate, resulting in the beampattern's diagonal structure. The beampatterns also illustrate how higher order modes (that are not included in the pseudo-inverse) project into the lower modes. As these plots show, the amount of crosstalk depends on the number of modes in the filter, i.e., the 10-mode filter is significantly better at rejecting energy from higher modes than the 15-mode filter. This is due to the fact that the amount of crosstalk is governed by the conditioning of the pseudo-inverse, which degrades as more modes are included.

Figure 8 demonstrates how adding constraints (i.e., additional modes) increases a filter's sensitivity to noise. The plot shows white noise gain as a function of mode number for the matched filter and five pseudo-inverse filters, designed for 10, 12, 13, 14, and 15 modes, respectively. Note that the results are normalized by N, the maximum possible white noise gain. The plot confirms that the matched filter achieves the optimal gain and also shows that the gain of the 10-mode pseudo-inverse filter is equivalent to the matched filter. Curves for the 12-, 13-, 14-, and 15-mode PI filters indicate that there is a significant loss in white noise gain as additional null constraints are added to the design. The reason is that modes above 10 are not adequately sampled by the 40-element ATOC VLA, therefore the smallest singular values of $\mathbf{E}$ decrease dramatically as these modes are included.

C. Broadband performance analysis

In designing a short-time Fourier mode processor, it is crucial to have a measure of how well a narrowband mode filter designed for particular frequency performs on the modes at neighboring frequencies. The operating bandwidth of the narrowband mode filters determines the required frequency resolution of the lowpass filter, which in turn defines the temporal resolution of the processor. This section addresses the broadband performance issues associated with the STFT approach by examining the frequency and noise responses of the MF and PI mode filters.

The frequency-dependent beampattern characterizes the frequency response of the STFT processor. For the kth bin, this beampattern is defined as $20\log_{10}(\mathbf{W}_k^{\mathrm{H}}\mathbf{\Phi}[\omega])$, where $\mathbf{W}_k$ is the narrow-band spatial filter designed with the modeshapes at the center frequency of the bin. Figure 9 shows the broadband beampatterns for modes 1 and 10, generated with MF and PI filters for the ATOC Hawaii VLA. The pseudo-inverse filter includes the first ten modes. Each of the spatial filters is designed for a center frequency of 75 Hz, and the results are shown for the 30 Hz (± 15 Hz) band around that frequency. The solid lines in the plots represent the response in the desired mode and the dashed lines represent the crosstalk from neighboring modes (from 1 to 10) into the desired mode. As Fig. 9 indicates, both filters have a flat

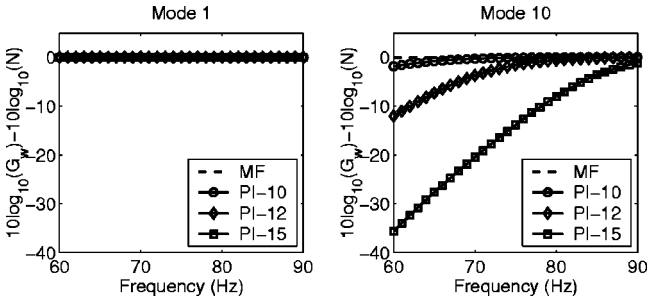


FIG. 10. White noise gain (normalized by N) as a function of frequency for the matched filter and three different pseudo-inverse filters (for 10, 12, and 15 modes). Results are shown in dB.

TABLE I. Simulation parameters for the California–Hawaii adiabatic propagation example.

Source range		700 m
	Center frequency	75 Hz
	Pulse	triangular-windowed sinusoid
	Pulse duration	0.11 s ($\approx$30 Hz bandwidth)
Receiver depth		3515.2 km
	No. of receivers	40
	Element spacing	35 m
	Span	330–1695 m
	Sample rate	300 Hz
Modes	No. included	40

response in mode 1, which is reasonable considering that the first mode's shape does not vary significantly with frequency (see Fig. 4). At 75 Hz there are nulls in the response of mode 1 to input from modes 2–10, but crosstalk increases substantially as ω deviates from the design frequency. The plots for mode 10 indicate that frequency mismatch can affect the gain in the desired mode as well as the crosstalk rejection, e.g., the gain in mode 10 is down by 10 dB at 60 Hz. This is due to the significant changes in the shape of mode 10 as a function of frequency, illustrated in Fig. 4. As expected from the narrowband analysis of the previous section, the frequency-dependent beampattern shows that the MF beamformer does not prevent crosstalk at the center frequency in mode 10. In contrast the PI filter for mode 10 is constrained to have nulls at 75 Hz for modes 1–9. Figure 9 demonstrates that these constraints are not satisfied for frequencies other than the design frequency.

Based on the MF and PI beampatterns shown, frequency mismatch affects crosstalk rejection much more than it affects the response in the desired mode. Thus the allowable crosstalk defines the operating bandwidth of the mode filters. For example, to guarantee a maximum crosstalk level over the band of less than -20 dB, the bandwidth of the ten-mode PI filter is ± 2.5 Hz. Using this criteria, the cutoff frequency of the lowpass filter (H_{LP}) in the STFT processor for ATOC should be ± 2.5 Hz. In addition, since the crosstalk levels in Fig. 9 are increasing away from the center frequency, it is desirable for the temporal filter to have decreasing (rather than constant) sidelobes. To meet these specifications, the lowpass filter chosen for the ATOC data is a 0.4-s Hanning window, which has a -3 dB bandwidth of ± 2.5 Hz and sidelobes that fall off at a rate of $1/\omega^3$. Recall from Sec. IV A that the temporal resolution of the processor is determined by the length of the lowpass filter. The Hanning window taper is such that it can begin to temporally resolve arrivals when they are farther apart than half its length, thus the resolution of the ATOC short-time Fourier processor is on the order of 0.2 s.

The pseudo-inverse filter clearly provides better crosstalk rejection than the matched filter, however it is subject to problems with the conditioning of the matrix inverse. As discussed in Sec. IV B, white noise gain is a convenient measure of processor sensitivity. Figure 10 shows the white noise gain, normalized by N, of the filters for modes 1 and 10 as a function of center frequency. Each of the plots have four curves: one for the matched filter and three for different realizations of the pseudo-inverse filter that are designed for 10, 12, and 15 modes, respectively. For mode 1 all three pseudo-inverse filters achieve the optimal (matched filter) white noise gain of $10\log_{10}(N)$ regardless of frequency. The plot for mode 10 illustrates an important characteristic of the white noise gain for the higher-order modes: white noise gain decreases as frequency decreases. This is due to the fact that the modeshapes at lower frequencies occupy a greater portion of the water column (see Fig. 4), thus are not spanned as well by the array, resulting in a sampled modeshape matrix with small singular values. Note that for the 10-mode PI filter white noise gain is almost constant across the 30-Hz band of interest.

Based on the discussion above, the pseudo-inverse filter for ten modes provides a better combination of noise gain and interference rejection than either the matched filter or PI filters designed for higher numbers of modes. The next section illustrates the characteristics of short-time Fourier mode estimates obtained by using the ten-mode PI filter to process a simulated reception on the ATOC array.

D. STFT processing example

Table I summarizes the parameters for an adiabatic simulation of propagation over the California–Hawaii path of the ATOC experiment. The environment contains a broadband point source at 700-m depth, located 3515.2 km away from a 40-element receiving array with 35-m spacing (identical to the ATOC VLAs). In this example, the source transmits a single windowed-sinusoidal pulse with approximately 30-Hz bandwidth, at a center frequency of 75 Hz. The time series for the receivers is synthesized from the first 40 modes,[46] using range-averaged wavenumbers calculated from Levitus winter climatology and range-varying bathymetry for the path. Figure 11 shows the received pressure field for the adiabatic simulation. At 3515.2-km range, the modes are dispersed enough that it is possible to identify individual modes in the pressure time series, e.g., mode 1 is the strong final arrival,[47] and mode 5 is associated with the five strong peaks lined up right after 2374.5 s.

Figure 11 also shows the corresponding short-time Fourier mode estimates, as a function of time and frequency, for modes 1, 5, and 10. The estimates were computed using processing parameters identical to those used on the ATOC data in the next section: a 0.4-s Hanning window lowpass

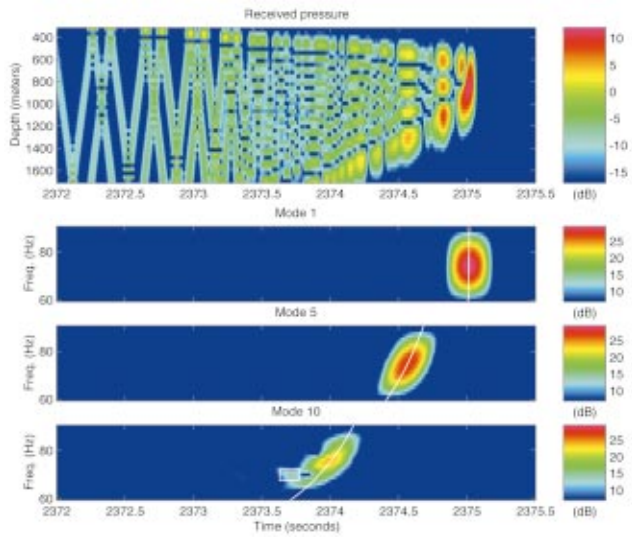


FIG. 11. STFT processing of adiabatic simulation of California–Hawaii path. Top plot is the synthesized pressure field on 40-element array. Bottom plots are the short-time Fourier estimates for modes 1, 5, and 10. White lines in the mode estimates represent the adiabatic arrival times. The white box indicates the components in the mode 10 signal that are due to crosstalk from modes 11 and 12.

filter and a ten-mode pseudo-inverse spatial filter. For this example the bin spacing of the STFT filterbank is 1.25 Hz. Since the lowpass filter has a bandwidth of approximately ± 2.5 Hz, neighboring bins are highly correlated. A 2.5-Hz bin spacing is all that is required to adequately represent the underlying signals filtered by the Hanning window. Oversampling by a factor of 2 in frequency improves the appearance of the plots, but does not affect the overall resolution of the processor. The solid white line in each plot corresponds to the mode arrival time computed using adiabatic group velocities.

These plots demonstrate that the STFT mode processor is working as expected. First, the arrivals in the estimated modal time series correctly line up with the appropriate peaks in the pressure time series and the adiabatic predictions. Second, the dispersion characteristics of each mode are visible in the output. Mode 1 shows all frequencies arriving at the same time, meaning it is undispersed, whereas modes 5 and 10 clearly show the lower frequencies arriving first, as expected in deep water. Third, with the exception of a small amount of crosstalk (indicated by the white box in the bottom plot) from modes 11 and 12 into earliest arrivals of mode 10, the STFT processor effectively filters out the higher-order modes. Note that this crosstalk is predicted by the beampattern in Fig. 7. Finally, the temporal smearing caused by the 0.4-s low-pass filter is evident in the broader arrival peaks of the mode time series, as compared to the arrivals in the simulated VLA data.

V. STFT ANALYSIS OF RECEPTIONS AT 3515 km

This section presents an analysis of the ATOC receptions on the Hawaii vertical line array using the short-time Fourier techniques described in Sec. IV. The discussion begins with an overview of the experimental data set, including a brief description of the source/receiver configuration and preprocessing algorithms. Following that, Sec. V B describes the short-time Fourier mode estimates for the ATOC receptions and compares them to the results for simulated receptions. Sections V C and V D address the issues of mode coherence and temporal variability, respectively. Finally, Sec. V E discusses mode arrival time statistics and analyzes trends over the course of the experiment.

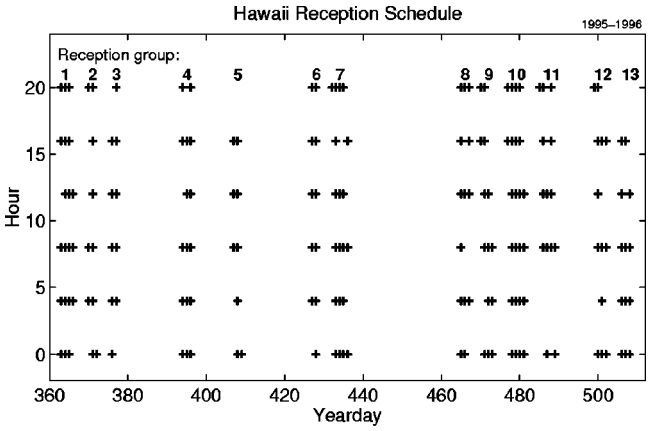


FIG. 12. ATOC transmission schedule through yearday 509. Crosses mark the time of each good reception; receptions with bad channels were eliminated from the data set. The line of numbers above the crosses indicates the division into 13 subgroups for postprocessing.

A. ATOC Hawaii VLA data set

As noted in the Introduction, the ATOC experiment employed a bottom-mounted source at Pioneer Seamount (off California) and two vertical arrays, located near Hawaii and Kiritimati in the Northeastern Pacific. A paper by the ATOC Instrumentation Group provides a thorough description of the source and receiver hardware.[48] What follows is a brief review of the relevant details.

The source transmitted phase-encoded pseudo-random sequences at a center frequency of 75 Hz, with a -3 dB bandwidth of 37.5 Hz. Each transmission consisted of 44 repetitions of the 27.28-s (1023 digits) M-sequence, corresponding to a transmission length of approximately 20 min. The source and the two VLAs were deployed in the fall of 1995, and regular transmissions began in December of that year. Over the course of the experiment, the source transmitted signals every 4 h during periods set by the ATOC Marine Mammal Research Program.

This paper describes the analysis of data recorded on the 40-element array near Hawaii, located at a range of 3515.2 km from the source. Figure 12 shows the schedule of reception times for this array from 28 December 1995 (yearday 362) to 23 May 1996 (yearday 509). Although the array was not recovered until August 1996, the deepest 20 hydrophones failed sometime after day 509. Since mode processing is much more difficult with only the shallow half of the array (due to inadequate sampling of the modeshapes), this study is limited to receptions recorded with the full array. There were 229 transmissions between yeardays 362 and 509. Forty-one of these were eliminated from the analysis due to incomplete or corrupted time series, leaving the 188 receptions shown in Fig. 12.

For each transmission, the 40-element array recorded the

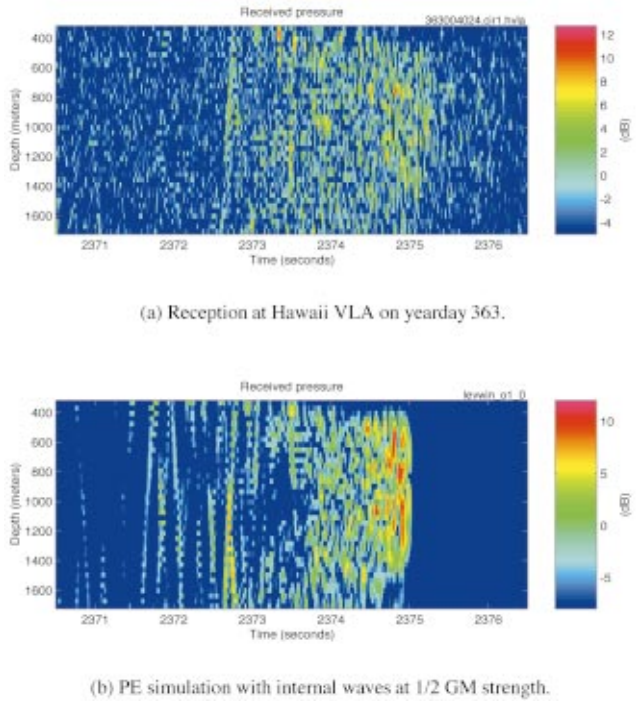


FIG. 13. Comparison of experimental and simulated receptions for the Hawaii VLA. (a) Reception at Hawaii VLA on yearday 363. (b) PE simulation with internal waves at $\frac{1}{2}$ GM strength.

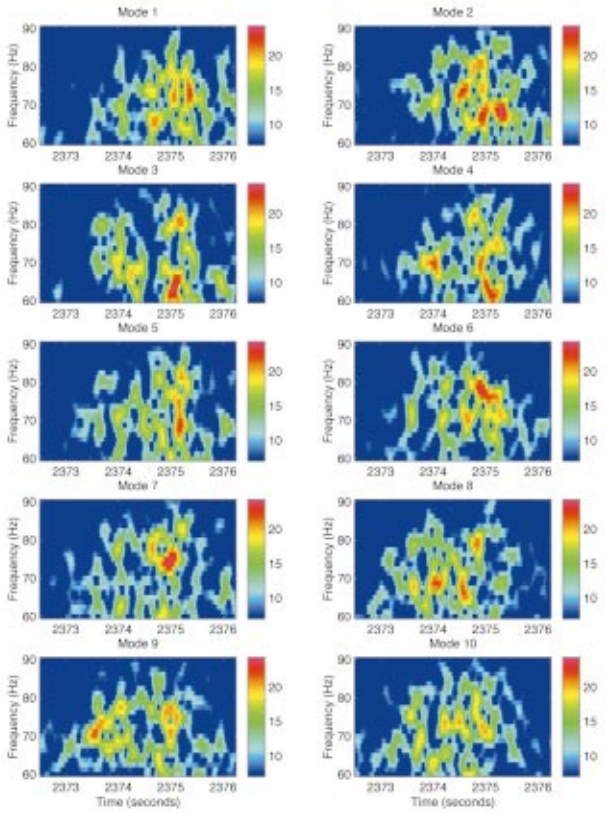


FIG. 14. Short-time Fourier mode estimates for the ATOC reception in Fig. 13(a). Color scale is in dB.

received signal on each hydrophone, averaged over four periods of the pseudo-random sequence. Ten such averages, spanning an 18.2-min interval, were recorded for each transmission.[49,50] Subsequently, the time series for each sensor was demodulated and matched-filtered to achieve pulse compression. For all the numerical results presented in this paper, the received pressure time series consists of these matched-filtered demodulates.

During the experiment, the position of the VLA was tracked using a long-baseline acoustic navigation system consisting of four transponders deployed on the bottom and several interrogator hydrophones on the array. Navigation data was recorded immediately before and after each reception. Array positions for each of the ten four-period averages were determined by interpolating between the beginning and ending locations of the array.

In the results presented below, estimates of the noise level for each reception were used to determine plotting and detection thresholds. These estimates were obtained from spectral analysis of noise-only segments of data at the beginning and end of each reception. Further details of the noise analysis for the Hawaii VLA are discussed by Wage.[51]

B. ATOC processing examples

The purpose of this section is to highlight important features of the short-time mode spectra by comparing the results for an ATOC reception with those for a simulated reception. Figure 13(a) is a plot of a demodulated pressure time series (one four-period average) recorded on the Hawaii array in late December 1995. Figure 13(b) shows the time series for a simulation of propagation over the Pioneer–Hawaii path. The simulation environment consists of Levitus background soundspeed profiles perturbed by $\frac{1}{2}$ Garrett-Munk (GM) internal waves, as described in Sec. II. In the pressure time series plots, the 0-dB level corresponds to the estimated noise floor in the 75Hz bin (the center frequency). Since the simulation does not contain any additive noise components, it is assumed that there is a noise floor 12 dB below the peak in the pressure field, which is consistent with the ATOC measurements.

Based on Fig. 13, the experimental and simulated data are quite similar. Unlike the adiabatic example considered in Sec. IV D, there are no immediately identifiable modes contained in the late-arriving energy of either the measured or simulated reception. The most striking difference between the PE simulation and the real data is that the simulated reception has a sharp cutoff at 2375 s while the ATOC reception exhibits no discernible cutoff. Absence of a sharp cutoff in the ATOC data is attributed to interaction with the steep bottom slope near the Pioneer Seamount source, which the simulation does not model.[51,52]

The STFT processor for the measured and simulated receptions uses a pseudo-inverse filter for ten modes in conjunction with a 0.4-s lowpass filter (Hanning window), which has the required ± 2.5-Hz bandwidth. Figure 14 shows the short-time Fourier mode estimates (as a function of time and frequency) for the first ten modes of the ATOC reception, and Fig. 15 shows the corresponding estimates for the first ten modes in the internal wave simulation. For modal time series plots the 0-dB level corresponds to the estimated noise level in mode 1 at 75 Hz for the reception.

There are a number of observations to make about the

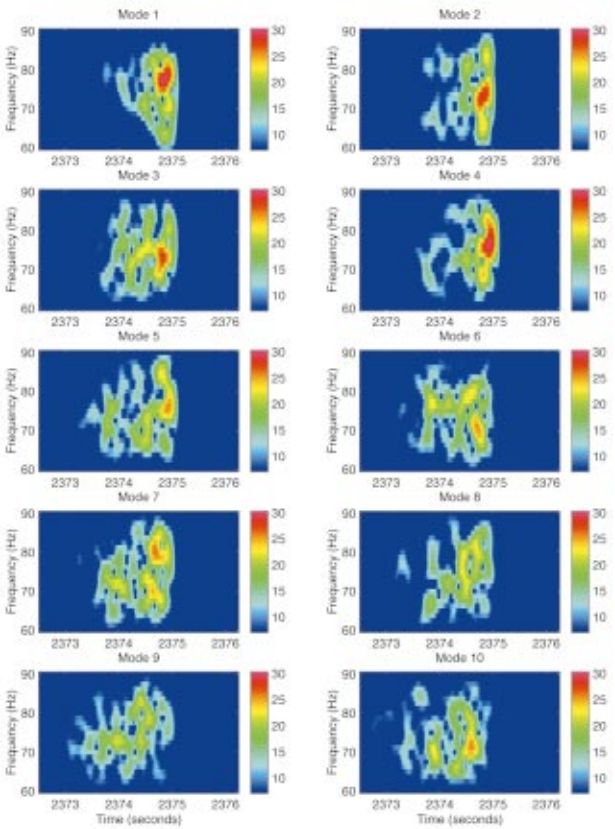


FIG. 15. Short-time Fourier mode estimates for the PE simulation with internal waves in Fig. 13(b). Color scale is in dB.

experimental results in Fig. 14. First, each mode consists of multiple arrivals, spread over 2–3 s. This is in stark contrast to the single, dispersive arrivals that characterized the adiabatic propagation example discussed in Sec. IV D. Second, note that the individual arrivals in each mode are on the order of 0.2 s wide, much wider than a typical ray arrival. This is due to the time-smearing inherent in the STFT processor, rather than to a fundamental difference between rays and modes. Third, recall that in the adiabatic case, the modes arrive in descending order and the highest modes are temporally separable from the lowest modes. For this ATOC reception, there is no obvious ordering of the arrivals and the spread of the signals is such that the first ten modes overlap in time. Fourth, the estimated modal time series contain faded arrivals. Frequency-selective fading occurs when two signals with different phase characteristics arrive simultaneously, i.e., within the time window used to compute the transform. Destructive interference of the signals results in deep fades of the spectral amplitude. If a short-time transform temporally resolves two signals, then only nonfaded arrivals (such as those found in the adiabatic example) are observed. The ATOC STFT processor begins to resolve signals at a separation of 0.2 s, thus the presence of faded arrivals in the Hawaii data suggests that the processor is measuring the interference pattern associated with multiple arrivals separated by less than 0.2 s.

A comparison of Figs. 14 and 15 indicates that the results of the $\frac{1}{2}$ GM internal wave simulation qualitatively agree with the experimental data. Each of the time-varying

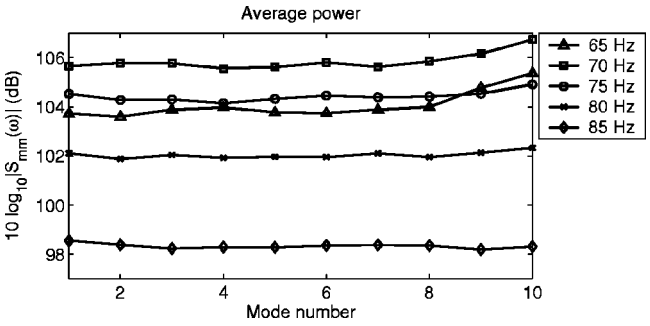


FIG. 16. Average power as a function of mode number in the 65-, 70-, 75-, 80-, and 85-Hz bins.

mode spectra for the PE simulation contains a series of faded arrivals. The sharp cutoff, which is a consequence of ignoring bathymetric effects near the source, is clearly evident in the mode estimates. Although the similarities between experiment and simulation are encouraging, the ATOC data does show significantly more time-spread than the PE data.

C. Mode coherence

According to Dozier and Tappert,[3,6] internal-wave-induced scattering decorrelates the mode signals. In this section we examine the coherence of the first ten modes in the ATOC receptions using magnitude-squared coherence (MSC) as a metric. The MSC of two random processes is defined as[53]

$$\mathrm{MSC}(\omega)=\frac{|S_{mn}(\omega)|^2}{S_{mm}(\omega)S_{nn}(\omega)}, \tag{16}$$

where S_{mn} is the cross power spectral density and S_{mm} and S_{nn} are the auto power spectral densities. In practice, the cross power spectrum is estimated by averaging over L measurements in each frequency bin, e.g., an estimate of the cross spectrum for modes m and n is

$$\hat{S}_{mn}(\omega)=\frac{1}{L}\sum_{l=1}^{L}\hat{a}_m(\omega,l)\hat{a}_n^*(\omega,l), \tag{17}$$

where $\hat{a}_m$ and $\hat{a}_n$ are the amplitude estimates for modes m and n and* denotes the complex conjugate. Setting $m=n$ in Eq. (17) produces auto spectrum estimates. For the ATOC analysis, we obtain the measurements in each bin by subsampling the output of the STFT processor, taking one sample every 0.15 s. Since the processing window (i.e., the filter) is 0.4 s long, this subsampling corresponds to a 62.5% overlap between neighboring windows. To compute the spectral estimates, we use samples from a two-second interval (2373.4 to 2375.4) in each reception and average over all receptions.

Figure 16 shows the average power estimates ($10\log_{10}|\hat{S}_{mm}|$) as a function of mode number for the first ten modes in the 65-, 70-, 75-, 80-, and 85-Hz bins. Note that for these low modes the average power is approximately constant in each bin, which is consistent with Dozier and Tappert's prediction of an equipartitioning of energy among the modes. The slight increase for modes 9 and 10 at 65, 70, and 75 Hz is attributed to crosstalk from higher order modes

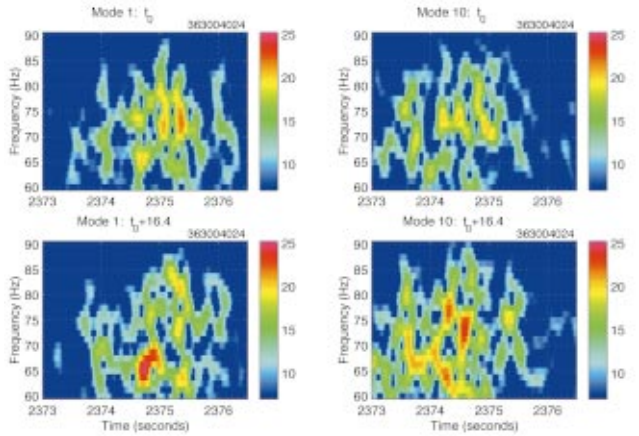


FIG. 17. Comparison of modes 1 and 10 for the first (top plots) and last (bottom plots) periods of a source transmission. Color scale is in dB.

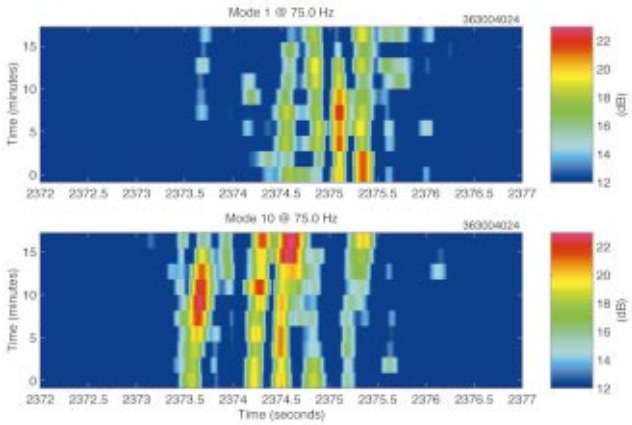


FIG. 18. Variability of modes 1 and 10 across a single transmission (18.2 min) in the 75-Hz bin.

not included in the estimate (recall the discussion of Fig. 7). Differences in absolute levels from bin to bin are a function of the source spectrum.

By definition, the MSC lies between 0 and 1. For the ATOC receptions, the maximum cross-mode coherence for any of the mode pairs in the 65-, 70-, 75-, 80-, and 85-Hz bins is 0.05. These experimental measurements indicate that the signals in the first ten modes are incoherent at a range of 3515 km. This result supports Dozier and Tappert's claim that modes decorrelate with range due to internal-wave effects.

D. Temporal variability

The purpose of this section is to examine the temporal variability of the mode arrivals at megameter range. For deep water environments, Colosi *et al.*[54] (citing the work of Flatte and Stoughton[55]) indicate that acoustic coherence times are on the order of tens of minutes, whereas the coherence times for internal waves are on the order of hours. In analyzing data from the ATOC Engineering Test (range = 3250 km), Worcester *et al.* used 12.7-min coherent averages for the ray arrivals, since the SNR did not increase for longer averaging times.[56] For that same data set, Colosi *et al.* concluded that time fluctuations in the wavefronts show no coherence at 2-h lag times.[54] Note that in the long-range experiments cited above, the focus was primarily on analyzing the ray arrivals, rather than the late-arriving modes. The rest of this section considers the temporal variations of the mode arrivals in ATOC.

Recall that for each source transmission, the VLA recorded ten four-period averages of the 27.28-s pseudo-random sequence. The results discussed in Sec. V B were computed using the first four-period average of a transmission received by the Hawaii array in late December 1995. Figure 17 compares mode estimates for the first four-period average and the last four-period average of that transmission. The top plots are the estimated spectra for modes 1 and 10 from the first four periods and the bottom plots correspond to the last four periods. Note that the time difference between the end of the first reception and the start of the tenth is approximately 14.5 min. The plot clearly demonstrates that the mode signals change significantly over that time interval: some arrivals drop out entirely, and new arrivals emerge. Using the center frequency bin, Fig. 18 provides a clearer picture of the temporal variability of the mode signals. The two plots in the figure are stacks of the received signals in mode 1 (top) and mode 10 (bottom) at successive 1.82-min (four-period) lags. While some of the peaks are consistent across the full transmission interval, others fade in/out suddenly.

There are a variety of ways to quantify the temporal variability of signals. In this paper, we consider two approaches. The first approach consists of computing the MSC between the first four-period average in a transmission and the successive four-period averages. Yang *et al.* used a similar method to analyze coherence times for a recent shallow water internal wave experiment.[57] Figure 19 shows the MSC estimate for mode 1, computed using the Hawaii VLA data. Each curve represents the mean over all transmissions. Note that the MSC decreases rapidly, reaching 0.5 at approximately 4.5 min and becoming effectively equal to zero for lag times greater than 14 min. The figure indicates a mild dependence on frequency: mode 1 at 85 Hz has a slightly shorter coherence time than mode 1 at 65 Hz. This apparent dependence may be an artifact due to the lower signal levels

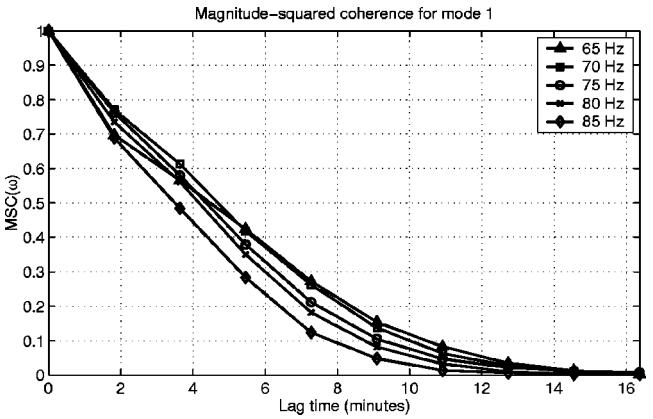


FIG. 19. Magnitude-squared coherence for mode 1 as a function of the lag time between the first four-period average and successive four-period averages in a single transmission. These curves depict the mean over all transmissions.

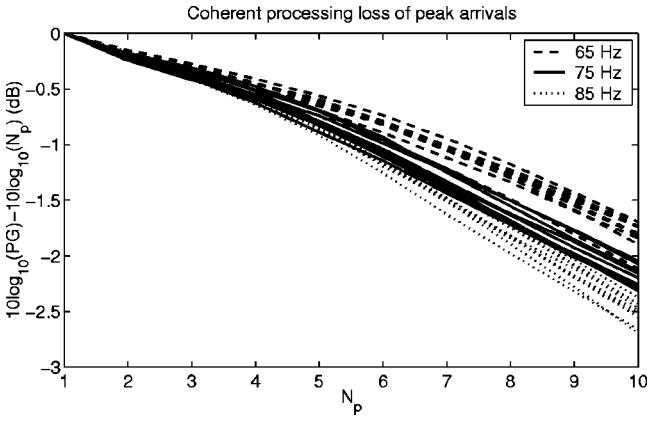


FIG. 20. Coherent processing loss for the peak arrivals in modes 1–10 for the 65-Hz, 75-Hz, and 85-Hz bins.

at the higher frequencies. The MSC estimates for mode 1 are representative of the results for modes 2–10.

The second approach to quantifying temporal variability focuses on the arrival peaks in the bandpass mode estimates. In previous work, Tang and Tappert used processing gain as a measure of the temporal coherence of pulses propagating through internal waves in a shallow (200 m) water environment at ranges of 20 km.[58] They studied pulse coherence times without addressing the issue of modes specifically. Processing gain is defined as a ratio of coherent to incoherent power. Consider the processing gain, PG, associated with a peak arrival in a mode:

$$\mathrm{PG}=\frac{|\Sigma_{i=1}^{N_p} a(t_{\mathrm{peak}},i)|^2}{\Sigma_{i=1}^{N_p}|a(t_{\mathrm{peak}},i)|^2}, \tag{18}$$

where $a(t_{\mathrm{peak}},i)$ is the estimated complex mode amplitude at time t_{peak} in the ith four-period average. N_p is the number of four-period averages included in the calculation. Note that the maximum value of PG is equal to N_p and is achieved when the signal is perfectly coherent. For the Hawaii data set, the following procedure is used to obtain an average processing gain for the peaks in each mode. The initial step consists of finding the peak arrival times for the first four-period average in each reception using a detection threshold of 12 dB above the estimated noise floor. Then the processing gain for each peak is calculated by computing the ratio of coherent and incoherent sums at the peak time. The number of periods (N_p) to include in the sums varies from 1 to 10. The resulting gains are averaged over all peaks obtained for that mode in the Hawaii VLA data set (188 receptions). Figure 20 shows the coherent processing loss for the peaks, which is defined as $10\log_{10}(\mathrm{PG})-10\log_{10}(N_p)$. This quantity measures how close the average processing gain is to the ideal gain achieved by a perfectly coherent signal. The dashed, solid, and dotted lines on the plot correspond to the average coherent processing loss for modes 1–10 at 65, 75, and 85 Hz, respectively. Recall that each four-period average represents 1.82 min of data. Based on the figure, processing gain is within 0.5 dB of the ideal gain for N_p up to 3. In other words the peaks are predominantly coherent up to 5.5 min. Beyond that, the average processing gain diverges from that of a perfectly coherent signal. Similar to the MSC estimates,

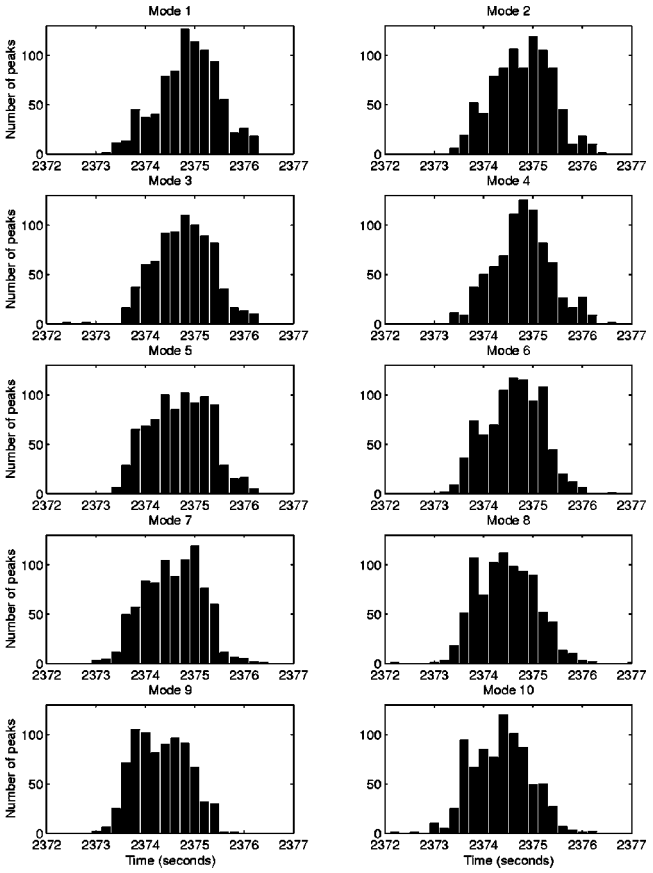


FIG. 21. Histogram of peak arrivals in the 75-Hz bin for the first ATOC reception group. Detection threshold was set at 12 dB above the estimated noise floor.

these curves indicate a mild dependence on frequency: modes at 65 Hz have slightly longer coherence times than the modes at 75 or 85 Hz. At a single frequency, there is no obvious ordering of the data by mode number. This is not surprising given the amount of mode coupling that occurs over the 3515-km propagation path.

It is important to note that Fig. 20 provides an average measure of temporal coherence. Based on the plot in Fig. 18, some of the mode arrivals are coherent over the reception period (18.2 min). Further analysis of the processing gain data reveals that approximately 9% of all arrivals are coherent over the entire reception (where coherence is defined to be $\mathrm{PG}>9$).

E. Arrival time statistics

On days that the ATOC source was operating, there were transmissions at 4-h intervals. Given the variability of the mode signals over a single 18.2-min reception, it is not surprising that the locations of the arrival peaks change substantially after 4 h. To get an idea of how the peak arrival times are distributed, consider the histograms for modes 1–10, shown in Fig. 21. These plots were compiled by picking peaks in the first group of ATOC receptions (yeardays 363–366, containing 200 four-period averages) using a detection threshold of 12 dB above the estimated noise floor for each mode. The results demonstrate that there is not a dominant arrival time in each mode; rather, the peaks are distributed

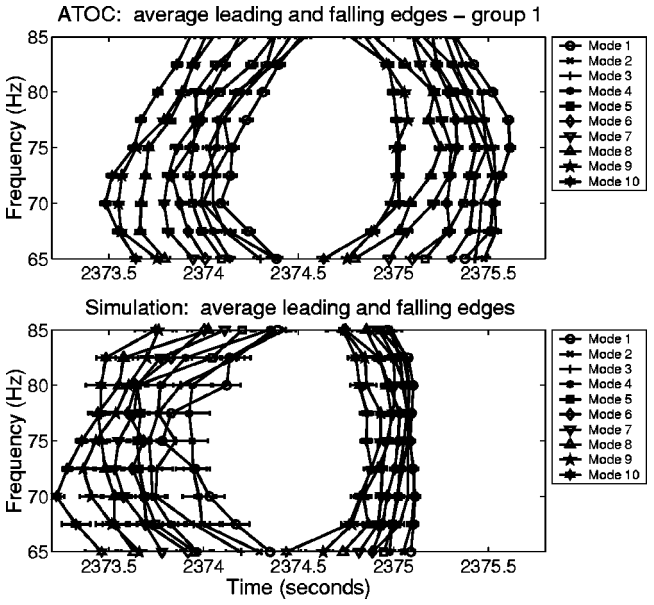


FIG. 22. Comparison of average leading and falling edges for the first group of ATOC receptions and a simulated data set. The latter includes ten simulated receptions through independent realizations of a $\frac{1}{2}$ Garrett-Munk internal wave field.

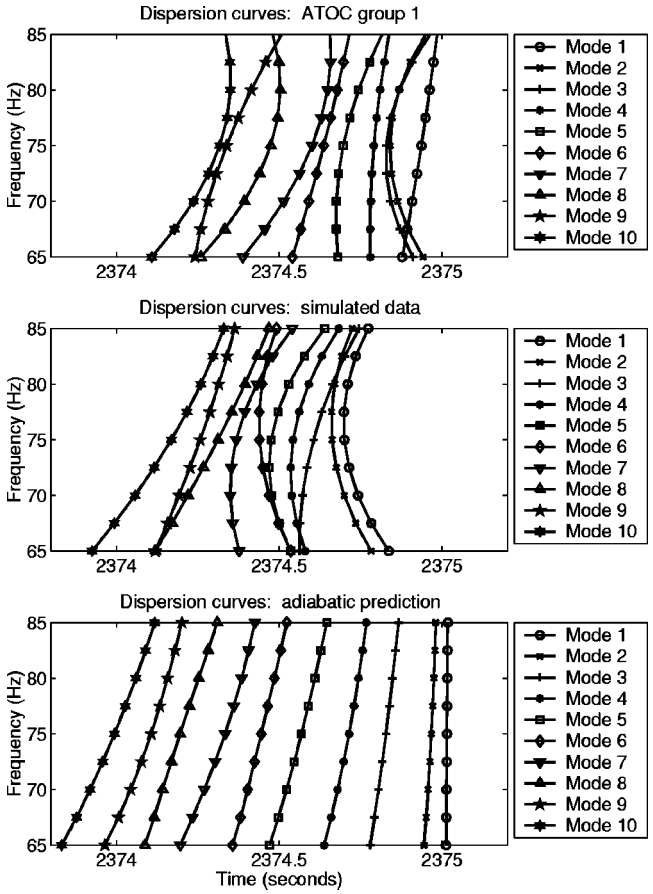


FIG. 23. Comparison of estimated dispersion curves for the first group of ATOC receptions, the simulated receptions, and the adiabatic predictions.

between 2373 and 2376 s for all of the modes. Note that the distribution of the higher modes (e.g., mode 10) is skewed towards the early part of the 3-s interval while the low modes are concentrated towards the latter part of the interval. This is consistent with deep water dispersion, where higher modes arrive first.

Since there is not a dominant peak in each mode that can be tracked, it is necessary to consider average travel time statistics for the mode signals. For the ATOC data, the leading edge, falling edge, and centroid of the short-time Fourier mode estimates provide a useful characterization of the arrival structure. Leading and falling edges are defined as the the first and last (respectively) time indices where the complex envelope of the mode estimate exceeds a threshold. The centroid is defined to be the center of mass of the portion of the complex envelope that is higher than the threshold value. For all of the results presented below, the threshold was set at 12 dB above the noise floor in each mode.

Figure 22 shows the results of averaging over all the receptions in the first group (a total of 200 four-period averages) to obtain the leading and falling edges as a function of frequency for the first 10 modes. For reference, the figure also includes the average leading and falling edges for the mode estimates of ten simulated receptions, which result from ten independent realizations of a $\frac{1}{2}$ Garrett-Munk internal wave field. The error bars on these plots represent the standard error, i.e., the sample standard deviation divided by the square root of the number of samples included in the average.[59] The ATOC statistics reveal several significant features. First, the average spread between leading and falling edges is on the order of 1.5 s. Second, the high modes have earlier arrival times, while low modes have later arrival times, as would be expected in a deep water channel. Although there is some crossover between neighboring modes, the leading and falling edges show that there are statistically significant differences in arrival time among the modes, e.g., compare modes 1 and 10. Finally, the plot indicates that the signals at the center frequency (75 Hz) are more spread than those at either end of the band. This effect may be partially due to the fact that the same detection threshold is used for all frequency bins, while the source spectrum rolls off as a function of frequency.

Since the simulated data is averaged over only ten receptions, the resulting curves in Fig. 22 are not as smooth as the ATOC data and the error bars are larger. Nevertheless, the simulated data is comparable to the real data in two important respects: the leading edges show similar behavior (as a function of frequency) to the ATOC data, and the arrival times are also comparable. Unlike the falling edges in the ATOC data, however, the falling edges in the simulated data are much more concentrated. This sudden cutoff has been noted above and is likely the result of ignoring downslope propagation effects near the source.

A second-order least squares fit to the centroid data has been used to estimate average mode arrival times as a function of frequency, i.e., dispersion curves for each mode. Figure 23 compares the estimated dispersion curves for the first group of ATOC receptions with the dispersion curves obtained from the ten simulated receptions. Predicted adiabatic dispersion curves are included for reference. The experimental data shows the modes arriving in descending order (though there is some crossover among nearest neighbors). The simulated data show good agreement: the mean arrival

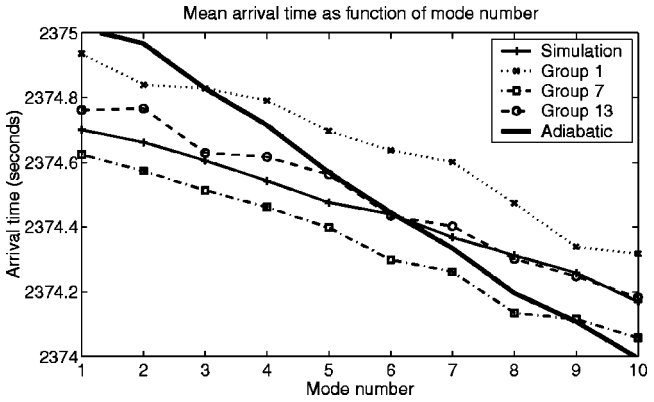


FIG. 24. Mean arrival time at 75 Hz as a function of mode number for several groups of ATOC receptions and a group of simulated receptions. An adiabatic prediction (heavy solid line) is included for reference.

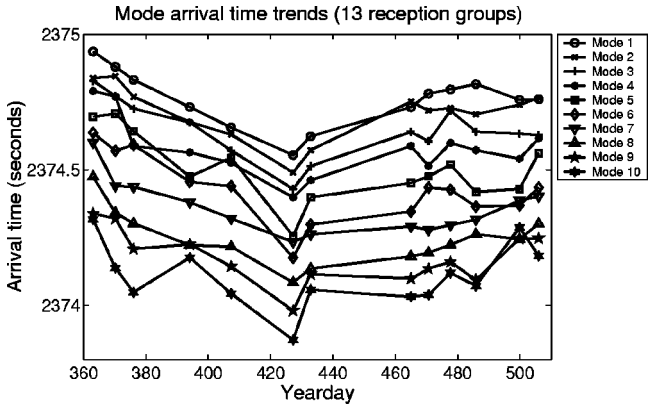


FIG. 25. Mean arrival time at 75 Hz for modes 1–10 as a function of yearday.

times extend over the same interval, approximately 0.8 s. The key difference between the ATOC data and the simulated data in Fig. 23 is a shift in the mean arrival times: the curves for the first group of ATOC data occur approximately 0.2 s later than the centroids of the simulated receptions. Although Fig. 23 shows an obvious difference between experiment and simulation in the curvature of mode 1's dispersion characteristic, the plots for other ATOC reception groups indicate that this is not a consistent feature in the data. Comparing the ATOC and simulated data to the predicted adiabatic arrival times indicates that the intermode dispersion is larger for the adiabatic case than for the measured or simulated data, i.e., the difference in the mean arrival times of modes 1 and 10 is larger for the adiabatic prediction than for the estimated dispersion curves. This is expected because coupling of energy from mode to mode tends to drive the average arrival times closer together. Note that although the intermode dispersion is larger, the dispersion within each mode (intramode) is obviously smaller for the adiabatic case, as Figs. 2 and 3 demonstrate.

Figure 24 shows the mean arrival time at 75 Hz (obtained from the dispersion curves) as a function of mode number. There are five lines on the plot, which correspond to the results for ATOC reception groups 1, 7, and 13, the PE simulation result, and the adiabatic prediction. The standard errors[58] for the mean arrival time estimates range from 7 to 18 ms for the ATOC data and 18 to 32 ms for the simulated data. Figure 24 illustrates several important points. The trend of decreasing travel time with increasing mode number is obvious from all of the curves. The three ATOC groups shown represent the beginning, middle, and end of the Hawaii data set. It is clear that there are shifts in travel time over the course of the experiment, however the slope of the arrival time versus mode number curve remains roughly constant. Note that the slope of the line for the $\frac{1}{2}$ Garrett-Munk PE simulation data also appears to agree with the experimental results. In contrast, the adiabatic mode prediction has a much steeper arrival-time versus mode number curve.

The Hawaii data set contains 5 months worth of data, taken between the end of December 1995 and May of 1996. Figure 25 illustrates how mean arrival time in the 75-Hz bin changes over the course of the experiment. Note that the minimum travel time for all of the modes occurs around yearday 427, which corresponds to the beginning of March 1996. The trend of decreasing mode arrival time from the start of the experiment until March and increasing afterwards agrees with the trend observed in the ray arrivals for the Hawaii VLA.[7] The difference in the arrival times between the first group of receptions (yearday 363) and the sixth group of receptions (yearday 427) is between 0.3 and 0.5 s for the first ten modes. A *t*-test confirms that these time differences are statistically significant at the 1% level. Based on the 1998 *Science* article by the ATOC Consortium,[7] the ray arrivals exhibit somewhat smaller time shifts, on the order of 0.25 s over that same period. The *Science* article notes that the trend observed at the Hawaii array is not in agreement with the expected seasonal trend (which would have shown increasing travel times in winter and decreasing in the summer) and postulates that this is due to a subsurface warming near the receiver that offsets the winter surface cooling layer near the source.

VI. CONCLUSION

In this paper we presented a short-time Fourier framework for broadband mode estimation. Using this flexible framework, we explored the time-and frequency-domain characteristics of the two most common modal beamforming algorithms and designed a mode processor for the ATOC experiment. Our analysis of five months of receptions at the Hawaii VLA produced a number of results regarding the mode arrival structure at megameter range. First, we demonstrated that the low mode signals at 3515 km consist of a series of arrivals, rather than the single dispersive arrival that typifies adiabatic propagation, confirming the predictions of numerous simulation studies. Second, we showed that the first ten modes have roughly equal average powers and that the cross-mode coherence is effectively zero. These results agree with Dozier and Tappert's conclusions based on theoretical and numerical work.[3,6] Third, we examined the temporal variability of the mode signals and determined that the average coherence times of the peaks in each bin are on the order of 5.5 min; coherence times of the overall signal (not just the peaks) are even shorter. Fourth, we computed centroid statistics that reveal mode-dependent trends in arrival

time and highlight mean shifts in the environment over 5 months of the experiment. We noted that the slope of the centroid arrival time versus mode number curve derived from the experimental data agrees with the slope derived from numerical simulations of propagation through $\frac{1}{2}$ Garrett-Munk strength internal waves.

This research has important implications for acoustic tomography using normal modes. Noting that the individual mode arrivals are not stable over time, we conclude that inversions must be based on the statistics of the mode field. Additional work is needed to investigate the dependence of mode statistics on the background environment and the parameters of the internal wave field.

ACKNOWLEDGMENTS

This work was supported in part by the Strategic Environmental Research and Development Program through Defense Advanced Research Projects Agency (DARPA) Grant No. MDA972-93-1-0003 and Office of Naval Research (ONR) Grant No. N00014-97-1-0788. K. Wage gratefully acknowledges additional support from an Armed Forces Communications and Electronics Association Postdoctoral Fellowship, and an ONR Ocean Acoustics Young Faculty Award. This paper is Woods Hole Oceanographic Institution contribution #10445. The authors thank the anonymous reviewers for their suggestions.

[1] W. Munk and C. Wunsch, ''Ocean acoustic tomography: Rays and modes,'' Rev. Geophys. Space Phys. **21(4)**, 777–793 (1983).
[2] S. M. Flatte, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge U. P., Cambridge, England, 1979).
[3] L. B. Dozier and F. D. Tappert, ''Statistics of normal mode amplitudes in a random ocean. I. Theory,'' J. Acoust. Soc. Am. **63**, 353–365 (1978).
[4] E. Yu. Gorodetskaya, A. I. Malekhanov, A. G. Sazontov, and N. K. Vdovicheva, ''Deep-Water Acoustic Coherence at Long Ranges: Theoretical Prediction and Effects on Large-Array Signal Processing,'' IEEE J. Ocean. Eng. **24**(2), 156–171 (1999).
[5] J. A. Colosi and S. M. Flatte, ''Mode coupling by internal waves for multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620 (1996).
[6] L. B. Dozier and F. D. Tappert, ''Statistics of normal mode amplitudes in a random ocean. II. Computations,'' J. Acoust. Soc. Am. **64**, 533–547 (1978).
[7] The ATOC Consortium, ''Ocean Climate Change: Comparison of Acoustic Tomography, Satellite Altimetry, and Modeling,'' Science **281**, 1327–1332 (1998).
[8] L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer-Verlag, New York, 1991).
[9] Y. Desaubies, ''A uniformly valid solution for acoustic normal mode propagation in a range varying ocean,'' J. Acoust. Soc. Am. **76**, 624–626 (1984).
[10] Y. Desaubies, C. S. Chiu, and J. H. Miller, ''Acoustic mode propagation in a range-dependent ocean,'' J. Acoust. Soc. Am. **80**, 1148–1160 (1986).
[11] S. Levitus and T. P. Boyer, *World Ocean Atlas 1994 Volume 4: Temperature*, 1994, NOAA Atlas NESDIS 4.
[12] S. Levitus, R. Burgett, and T. P. Boyer, *World Ocean Atlas 1994 Volume 3: Salinity*, NOAA Atlas NESDIS 3, 1994.
[13] NOAA, National Geophysical Data Center, Boulder, CO, Data Announcement 88-MGG-02, *Digital relief of the Surface of the Earth*, 1988.
[14] The environment contains the actual bathymetry along the path, with the exception of one simplification: the steep downslope near the source has been eliminated. In the simulations the source is at 939.5-m depth (the depth of the actual seamount), and the bottom depth at the source is 3317 m. The effects of the near-source downslope propagation are discussed briefly in Sec. V B of this paper and by Wage.[51]
[15] M. D. Collins, *User's Guide for RAM*, Naval Research Laboratory, Washington, DC.
[16] J. A. Colosi and M. G. Brown, ''Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields,'' J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[17] J. A. Colosi, S. M. Flatte, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).
[18] J. A. Colosi and the ATOC Group, ''A Review of Recent Results on Ocean Acoustic Wave Propagation in Random Media: Basin Scales,'' IEEE J. Ocean. Eng. **24**(2), 138–155 (1999).
[19] A. G. Sazontov and V. A. Farfel, ''Matched filtering of a narrowband pulse signal transmitted through a random waveguide channel,'' Sov. Phys. Acoust. **38**(6), 591–595 (1992).
[20] E. Yu. Gorodetskaya, A. I. Malekhanov, A. G. Sazontov, and V. A. Farfel, ''Effects of Long-Range Propagation of Sound in a Random Inhomogeneous Ocean on the Gain Loss of a Horizontal Antenna Array,'' Acoust. Phys. **42**(5), 543–549 (1996).
[21] A. G. Sazontov, and V. A. Farfel, ''Fluctuation characteristics of the response of a horizontal array in a randomly inhomogeneous ocean with short-time averaging,'' Sov. Phys. Acoust. **37**(5), 514–518 (1991).
[22] N. K. Vdovicheva, E. Yu. Gorodetskaya, A. I. Malekhanov, and A. G. Sazontov, ''Gain of a Vertical Antenna Array in a Randomly Inhomogeneous Oceanic Waveguide,'' Acoust. Phys. **43**(6), 669–675 (1997).
[23] A. G. Sazontov ''Quasiclassical solution of the radiation transport equation in a scattering medium with regular refraction,'' Acoust. Phys. **42**(4), 487–494 (1996).
[24] M. G. Brown, J. Viechnicki, and F. D. Tappert, ''On the measurement of modal group time delays in the deep ocean,'' J. Acoust. Soc. Am. **100**, 2093–2102 (1996).
[25] W. Menke, *Geophysical Data Analysis:* Discrete Inverse Theory (Academic, New York, 1989).
[26] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 1968).
[27] J. R. Buck, J. C. Preisig, and K. E. Wage, ''A unified framework for mode filtering and the maximum *a posteriori* mode filter,'' J. Acoust. Soc. Am. **103**, 1813–1824 (1998).
[28] R. H. Ferris, ''Comparison of measured and calculated normal-mode amplitude functions for acoustic waves in shallow water,'' J. Acoust. Soc. Am. **52**, 981–988 (1972).
[29] F. Ingenito, ''Measurements of mode attenuation coefficients in shallow water,'' J. Acoust. Soc. Am. **53**, 858–863 (1973).
[30] C. T. Tindle, K. M. Guthrie, G. E. J. Bold, M. D. Johns, D. Jones, K. O. Dixon, and T. G. Birdsall, ''Measurements of the frequency dependence of normal modes,'' J. Acoust. Soc. Am. **64**, 1178–1185 (1978).
[31] R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, J. Apel, M. Badiey, C. S. Chiu, S. Finette, M. Orr, B. Pasewark, L. Turgot, S. Wolf, and D. Tielbuerger, ''Acoustic normal mode fluctuation statistics in the 1995 swarm internal wave scattering experiment,'' J. Acoust. Soc. Am. **107**, 201–220 (2000).
[32] P. Casey, ''Mode Extraction for Long Range Underwater Acoustic Signals,'' M. S. thesis, University of Auckland, February 1995.
[33] C. S. Chiu, C. W. Miller, and J. F. Lynch, ''Optimal Modal Beamforming of Bandpass Signals Using an Undersized Sparse Vertical Hydrophone Array: Theory and a Shallow-Water Experiment,'' IEEE J. Ocean. Eng. **22**(3), 522–533 (1997).
[34] T. C. Yang, ''Broadband source localization and signature estimation,'' J. Acoust. Soc. Am. **93**, 1797–1806 (1993).
[35] P. Sutton, W. M. L. Morawitz, B. D. Cornuelle, G. Masters, and P. F. Worcester, ''Incorporation of acoustic normal mode data into tomographic inversions in the Greenland Sea,'' J. Phys. Oceanogr. **99**(C6), 12487–12502 (1994).
[36] K. D. Heaney and W. A. Kuperman, ''Very long-range source localization with a small vertical array,'' J. Acoust. Soc. Am. **104**, 2149–2159 (1998).
[37] H.-Y. Chen and I-T. Lu, ''Localization of a broadband source using a matched-mode procedure in the time-frequency domain,'' IEEE J. Ocean. Eng. **19**(2), 166–174 (1994).
[38] J. B. Allen and L. R. Rabiner, ''A Unified Approach to Short-Time Fourier Analysis and Synthesis,'' Proc. IEEE **65**(11), 1558–1564 (1977).
[39] S. H. Nawab and T. F. Quatieri, ''Short-Time Fourier Transform,'' in *Advanced Topics in Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1988), pp. 289–337.
[40] Note that the block diagram in Fig. 5 implicitly assumes that the array is vertical. When the array is tilted, the processor must incorporate some sensor-dependent timing corrections as discussed by Wage.[51]

[41] F. J. Harris, ''On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform,'' Proc. IEEE **66**(1), 51–83 (1978).
[42] H. Cox, R. M. Zeskind, and M. M. Owen, ''Robust adaptive beamforming,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**(10), 1365–1376 (1987).
[43] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. (Harcourt Brace Jovanovich, San Diego, CA, 1988).
[44] A. G. Voronovich, V. V. Goncharov, A. Yu. Nikol'tsev, and Yu. A. Chepurin, ''Comparative analysis of methods for the normal mode decomposition of a sound field in a waveguide: numerical simulation and full-scale experiment,'' Sov. Phys. Acoust. **38**(4), 365–370 (1992).
[45] T. C. Yang, ''A method of range and depth estimation by modal decomposition,'' J. Acoust. Soc. Am. **82**, 1736–1745 (1987).
[46] Higher-order modes are propagating in the waveguide, but their arrivals occur prior to the start-time of the simulation window.
[47] Unlike the simulation in Fig. 2, mode 1 is excited by the source in this example.
[48] The ATOC Instrumentation Group: B. M. Howe, S. G. Anderson, A. Baggeroer, J. A. Colosi, K. R. Hardy, D. Horwitt, F. W. Karig, S. Leach, J. A. Mercer, Jr., K. Metzger, L. O. Olson, D. A. Peckham, D. A. Reddaway, R. R. Ryan, R. P. Stein, K. von der Heydt, J. D. Watson, S.-L. Weslander, and P. F. Worcester, ''Instrumentation for the Acoustic Thermometry of Ocean Climate (ATOC) Prototype Pacific Ocean Network,'' in *OCEANS '95 Conference Proceedings*, San Diego, CA, October 1995, pp. 1483–1500.
[49] Although the source transmitted 44 periods of the M-sequence, the array only recorded 40 periods. The start-time for recording was chosen so that sampling began approximately two periods after the start of the reception.[50]
[50] P. Worcester, ''ATOC95: Autonomous Vertical Line Arrays Experiment Plan,'' 1995.
[51] K. E. Wage, ''Broadband Modal Coherence and Beamforming at Megameter Ranges,'' Ph.D. thesis, Massachusetts Institute of Technology/Woods Hole Oceanographic Institution, February 2000.
[52] K. D. Heaney, ''Inverting for Source Location and Internal Wave Strength Using Long Range Ocean Acoustic Signals,'' Ph.D. thesis, University of California San Diego, 1997.
[53] G. C. Carter, ''Tutorial overview of coherence and time delay estimation,'' in *Coherence and Time Delay Estimation*, edited by G. Clifford Carter (IEEE, New York, 1993), pp. 1–27.
[54] J. A. Colosi, E. K. Scheer, S. M. Flatte, B. D. Cornuelle, M. A. Dzieciuch, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[55] S. M. Flatte and R. B. Stoughton, ''Predictions of internal-wave effects on ocean acoustic coherence, travel-time variance, and intensity moments for very long-range propagation,'' J. Acoust. Soc. Am. **84**, 1414–1424 (1988).
[56] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Höwe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[57] T. C. Yang, K. Yoo, and M. Siderius, ''Internal waves and its effect on signal propagation on the Adventure Bank,'' in *Proceedings of the 8th International Congress on Sound and Vibration*, Hong Kong, July 2001, pp. 3001–3008.
[58] X. Tang and F. D. Tappert, ''Effects of internal waves on sound pulse propagation in the Straits of Florida,'' IEEE J. Ocean. Eng. **22**(2), 245–255 (1997).
[59] S. A. Teukolsky, W. H. Press, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran* (Cambridge U.P., Cambridge, 1992), Chap. 14, p. 610.

The balanced electromagnetic separation transducer: A new bone conduction transducer

Bo E. V. Håkansson[a)]
Department of Signals and Systems, Chalmers University of Technology, S-412 96 Göteborg, Sweden

(Received 22 March 2002; revised 19 November 2002; accepted 19 November 2002)

Conventional bone conduction transducers, which are relatively large, suffer from poor performance at low frequencies. A new type of electro-dynamic transducer, the balanced electromagnetic separation transducer (BEST), was developed to improve the performance of the conventional transducers. By using a balanced suspension principle, the quadratic distortion forces, as well as the static forces between the vibrating parts, are principally counterbalanced. Both the distortion and the size of the transducer can therefore be considerably reduced. Moreover, the static and dynamic magnetic fluxes are separated, except in the air gap regions, giving a more efficient transducer. For example, in comparison with a conventional B71 transducer, a prototype of the BEST has: Lower total harmonic distortion (THD), by 20–25 dB, and improved sensitivity by 10–20 dB for 100 to 1000 Hz and by 2–10 dB for 1 to 10 kHz. From a clinical point of view, the BEST offers a chance to measure bone thresholds, at 250 and 500 Hz, which are reliable at hearing levels not possible before. For example, at 250 Hz the BEST has 23 dB higher sensitivity than the B71; the THD is improved from 61% (B71) to 3.3% (BEST) at 40 dB HL (ISO 389-3, 1994). © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1536633]

PACS numbers: 43.38.Ar, 43.38.Dv, 43.66.Ts [SLE]

I. INTRODUCTION

Hearing by bone conduction as a phenomenon, i.e., hearing sensitivity to vibrations induced directly or via skin or teeth to the skull bone, has been known since the 19th century. The interest in bone conduction was initially based on its usefulness as a diagnostic tool. In particular, it is used in hearing threshold testing to determine the sensorineural hearing loss or, indirectly, to determine the degree of conduction hearing loss by noting the difference between the air and the bone thresholds. A comprehensive description of bone conduction threshold testing can be found in Dirks (1994).

In recent decades the use of bone conduction hearing aids has increased. One reason for this increase is the promising results achieved with the bone-anchored hearing aid (BAHA). The BAHA uses a percutaneous titanium implant system to attach the transducer directly to the skull behind the ear. The principal design of the BAHA system is extensively presented by Tjellström *et al.* (2001).

Although bone conduction transducers are frequently used in hearing threshold testing and hearing aids, there is still a great need for improvements. A bone conduction transducer has the purpose to efficiently transform the electrical signal energy into mechanical vibratory energy without distorting the original (information carrying) electrical signal. Deterioration of the electrical signal can occur in the electromechanical conversion caused by nonlinearities; nonlinear distortion products then contaminate the mechanical vibratory output. Low distortion of these transducers is a prerequisite in obtaining reliable bone conduction threshold data and also to achieve high sound quality in bone conduction hearing aids. Erroneous hearing threshold data, caused by distortion in the transducer, could lead to an incorrect diagnosis of a patient's hearing etiology and jeopardize subsequent hearing aid fitting procedures.

It is well known that conventional audiometric transducers, as well as transducers for bone conductions hearing aids, have some drawbacks associated with their generic design. These drawbacks arise because the force generated in the air gap(s) is approximately proportional to the magnetic flux squared. To make such transducers usable as a hearing transducer, a static magnetic flux must be superimposed on the signal to obtain an acceptable, although not perfect, linear behavior. This static magnetic flux gives the transducer a bias point in the magnetic flux to force characteristics. When the signal flux is operating around this bias point, along the slightly bent quadratic curve, nonlinear distortion is generated, in particular at low frequencies and high signal levels. Moreover, in bone conduction hearing aids, poor efficiency and large size are other limiting factors that originate from the generic design of conventional bone conduction transducers.

In this study a new transducer design, which simultaneously meets the demands of high linearity, high efficiency, and small size, is presented. The new design is called the balanced electromagnetic separation transducer (BEST). Its performance in terms of frequency response and total harmonic distortion is compared with the widely used conventional audiometric bone conduction transducer, the B71, from Radioear Corporation, in the USA.

II. PRINCIPLE DESIGN

A. Background

Transducers for bone conduction sound generation can be constructed by different technologies, such as electro-

a)Electronic mail: boh@s2.chalmers.se; http://www.chalmers.se

dynamic, piezoelectric, or magnetostrictive. However, piezoelectric and magnetostrictive transducers are often considered inconvenient and are not commercially available, basically because of poor low-frequency response.

Electrodynamic transducers of the moving coil type, which are the most common in ordinary loudspeakers, are not considered desirable in hearing aid applications; their electrical impedance is too low. The low impedance is a consequence of that the driving voice coil: (1) is suspended to the sound radiating diaphragm and therefore should have a low weight; and (2) it should have a small radial (bundle) thickness to allow a high transversal magnetic flux density.

Electromagnetic transducers of the variable reluctance type combine properties such as small size, wide frequency range, high impedance, and efficient energy transformation; hence, they are almost exclusively used in hearing aid applications. In simple terms variable reluctance type transducers function according to the horseshoe magnet principle where there is a small air gap between the armature (basically the permanent magnet) and the yoke. By superimposing a signal magnetic flux (generated by a coil whose dimensions are not so critical) the force in the air gap, between the yoke and the armature, will vary accordingly. A more detailed description of present designs is presented below.

In hearing threshold testing, the B71 is today the most frequently used bone conduction transducer. A similar design is used in the BAHA system, in which the most conspicuous difference is that the internal suspension spring system has inherent damping (Håkansson *et al.*, 1990; Håkansson and Carlsson, 1985). Both of these transducers, the B71 and the BAHA, work according to the variable reluctance principle. However, even if these designs are well functioning there are some generic shortcomings in the conventional variable reluctance type transducers: these are related to poor frequency response and high level of distortion at low frequencies. The new balanced electromagnetic separation transducer (BEST) offers a significant improvement, especially in the application of bone conduction hearing threshold testing. The technical design differences between the B71 and the BEST are described below.

It should be noted that in transducers for bone conduction threshold testing, the consumption of current is of minor importance since they are driven by audiometers that are connected to a power line supply. This is probably the reason that there is a moving coil design as an alternative to the B71 for bone conduction threshold testing: the KH70 from Grahnert Präcitronic GmbH, Germany. The KH70 has a better linearity and radiates less aerial sound than the B71. On the other hand, the KH70 has some disadvantages, such as a bulkiness (the weight is five times and its height is 2.5 times that of the B71); as a consequence it is more difficult to attach it to the skull without it touching the external ear (the pinnae). This is probably the reason that the KH70 is quite uncommon in the audiology clinics in most countries, and it is not dealt with in this paper.

B. The conventional design—the B71

The B71 transducer has a plastic housing with a 1.75-cm^2 circular attachment surface toward the head, as shown in Fig. 1. With a steel-spring headband, the transducer is pressed with a total force of approximately 5–6 Newton against the mastoid area behind the ear. The transducer consists of an armature, a yoke, and a small but essential air gap which disrupts the magnetic flux path. The magnetic flux is composed of the static flux generated by the permanent magnet and the dynamic flux generated by the current in two coils. The total weight of the B71 is 19.9 g. In the cross section in Fig. 1, the spring suspension that connects the yoke to the armature, thereby maintaining the essential air gap between them, cannot be shown in full.

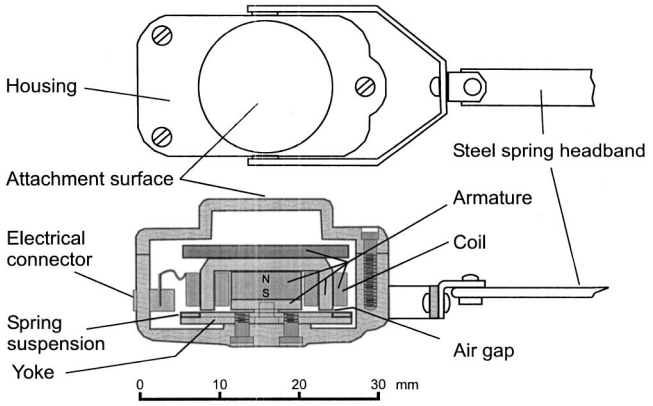


FIG. 1. External view and a cross-sectional sketch of the B71 transducer.

Some details of the static and dynamic magnetic circuits of the B71 are shown in Fig. 2. According to a first-order analysis of the electromagnetic circuit, assuming that the leakage of magnetic flux is negligible and assuming that the dependence between the magnetic field and the magnetic flux in the air gap is linear, it can be shown that the total force between the yoke and the armature is approximately proportional to the total magnetic flux squared.

$$F_{\text{tot}} \propto (\phi_0 + \phi_\sim)^2 = \phi_0^2 + 2 \cdot \phi_\sim \cdot \phi_0 + \phi_\sim^2, \tag{1}$$

where ϕ_0^2 represents the static force from the permanent magnet, $2 \cdot \phi_\sim \cdot \phi_0$ represents the desired signal force, and $\phi_\sim^2$ represents an undesired distortion force.

From Eq. (1) it is clear that in order to achieve a reasonably good linearity in the electro-mechanic transformation, a high static magnetic flux is required. By introducing a high static (biasing) magnetic flux ϕ_0 such that $\phi_0 \gg \phi_\sim$, it is possible to achieve a fairly linear behavior. However, when ϕ_0 is increased, the static force also increases, and the suspension spring compliance C (symbolically depicted in Fig.

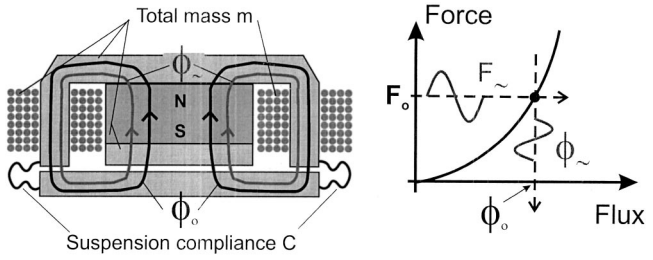


FIG. 2. Close-up of the magnetic circuit of the B71 and a corresponding flux to force characteristic.

2) must therefore be stiffer to avoid a collapse of the air gap. The consequences of a stiffer C (smaller value of C) will be described in what follows.

The seismic mass m, that serves as a counterweight, consists of a coil, a magnet, a soft iron armature, and an additional mass body (not shown in Fig. 2). This counterweight is essential in these designs for two reasons. First, if the counterweight m is very small, this means that very little of the force generated in the air gap will be transmitted to a vibration fed into the skull. Second, the counterweight mass m interacts with the suspension compliance C and generates a resonance frequency f_r

$$f_r \approx 1/(2\pi\sqrt{m \cdot C}) \text{ Hz}. \quad (2)$$

A transducer is normally designed so that the resonance frequency falls at a frequency slightly above the lowest frequency of interest, i.e., in the range of 500–1000 Hz for a BAHA and in the range of 250–500 Hz for a transducer used in bone conduction hearing threshold testing. This resonance gives an essential boost at low frequencies in variable reluctance transducers.

With this background some drawbacks of conventional variable reluctance type transducers can be pointed out. The first drawback is evident from Eq. (2): a low f_r requires a large mass m for a given compliance C. If the low-frequency response has to be improved and C cannot be more compliant (to prevent air gap collapse), the counterweight mass must be increased in spite of the undesired greater weight and size.

Moreover, a second drawback of the conventional variable reluctance transducer is that the magnetic circuit is designed so the magnetic signal flux follows or coincides with the static magnetic flux (Fig. 2). As a consequence, the properties of the electrodynamic conversion suffer since high quality permanent magnets have inherently poor properties for conducting the dynamic magnetic flux.

A third drawback of a conventional variable reluctance transducer is the vulnerability related to the existence of the static force, F_0 (Fig. 2). This force, generated by the permanent magnet, strives continuously to collapse the essential air gap between the yoke and the armature. This is critical, as this magnetic contraction force grows as the reciprocal of air gap length as air gap length decreases, whereas the counteracting suspension spring force grows linearly. A stable balance between the permanent magnet force and the force from the suspension spring is a prerequisite to maintaining this air gap. Furthermore, aging of the spring or severe external force/impact on the transducer can adversely affect the spring suspension. This is quite common in bone conduction hearing aids where such external forces can deform the spring suspension and the air gap might collapse partially or fully. If the air gap collapses, the sound from the transducer is severely distorted; the transducer must be repaired.

Finally, and perhaps most critically, since the total force output is approximately proportional to the total magnetic flux squared, a high level of primarily second harmonic distortion is generated. At low frequencies and at high signal levels, this distortion severely limits the use of conventional variable reluctance transducers in bone conduction hearing threshold testing, especially at 250 and 500 Hz. For example, a patient being tested at 250 Hz may hear the second harmonic at 500 Hz instead. This is especially serious, as the normal-hearing threshold is better by 9 dB at 500 Hz than at 250 Hz (ISO 389-3, 1994). This distortion, described in several studies, severely limits the use of the B71 transducer at lower frequencies (Dirks and Kamm, 1975; Dolan and Morris, 1990; Parving and Elberling, 1982).

It is understandable that the second harmonic distortion of the B71 is high. The primary reason for this, as mentioned, is the quadratic flux dependence, but a contributing factor is that the resonance frequency is located near 500 Hz. This means that the second harmonic component, generated in the electromagnetic conversion, is further enhanced by some 12 dB when the fundamental frequency is 250 Hz (12 dB is the approximate difference in the frequency response of the B71 between 250 and 500 Hz). To overcome this, the Radioear Company has developed another transducer model, the B72, which has a lower resonance frequency. This was mainly achieved by increasing the mass and, as a consequence, also the size. Larger size implies not only that it is more difficult to attach the transducer to the skull, but also that more airborne sound is radiated. The airborne sound may then be heard as an aerial sound via normal air conduction, instead of via bone conduction, and the result will be misleading (Lightfoot and Hughes, 1993). Due to these drawbacks the B72 is not widely used (at least not to the author's knowledge).

C. The balanced electromagnetic separation transducer (BEST)

It is clear that in the optimization procedure of a conventional variable reluctance type transducer some aspects are counterproductive. To achieve low distortion, a high static flux is needed: a high static flux requires a stiff suspension, and finally, a stiff suspension requires a heavy counterweight mass to avoid reducing the low-frequency response. This means that a design having low distortion often has, as consequence, a poor low-frequency response. The opposite is also true, i.e., a design with a good low-frequency response suffers from high distortion. The basic idea of the new transducer design (BEST) is to counterbalance or cancel the static forces, thus avoiding the strong requirement of a stiff spring suspension and a high mass to get an acceptably low distortion and good low-frequency response.

This counterbalance of the static forces is accomplished by introducing a second, opposing, air gap. A solution to implementing two opposing air gaps with only one magnet is shown in Fig. 3. The transducer in this figure has circular symmetry; it was the first attempt to implement the principles of the BEST.

It is shown here that the static forces of the upper and lower air gaps cancel or counterbalance each other, as they are of equal magnitude but act in opposite directions. The signal forces are generated in a push–pull fashion as follows. At one instance, the total magnetic flux is decreased from the static value in the upper air gap $[(\phi_0/2)-\phi_x]$, while it is

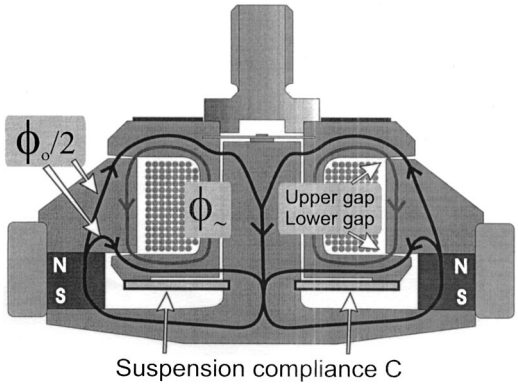


FIG. 3. Cross-sectional view of the BEST, in which one circular magnet and two opposing air gaps are used.

simultaneously increased by the same amount in the lower air gap $[(\phi_0/2)+\phi_\sim]$.

Some prototypes were built and it was proved that the idea underlying the present design worked. This implementation is extremely compact, which might be useful in an implantable bone conduction hearing aid where size is of the utmost importance. However, in these prototypes it was difficult to find the balanced position; also, the components had to be extremely precise, especially the suspension spring system.

A modified design of the BEST, especially adapted for bone conduction hearing threshold testing where size and power consumption is not so critical as in an implantable bone conduction hearing aid, was therefore developed (Fig. 4). This design is slightly larger, as it uses four magnets instead of one. Also, in addition to the two opposing internal upper and lower air gaps in the magnetic signal path, there are two external air gaps that affect only the static magnetic path. These additional external air gaps help to stabilize the static magnetic flux, i.e., to make it independent of the position of the bobbin arms in the air gaps. If the leakage of static magnetic flux is assumed negligible, the total air gap length for the static flux passing through the external and internal air gaps is constant and independent of the actual position of the bobbin relative to the yoke. Hence, it is much easier to maintain the balanced position, with the equal air

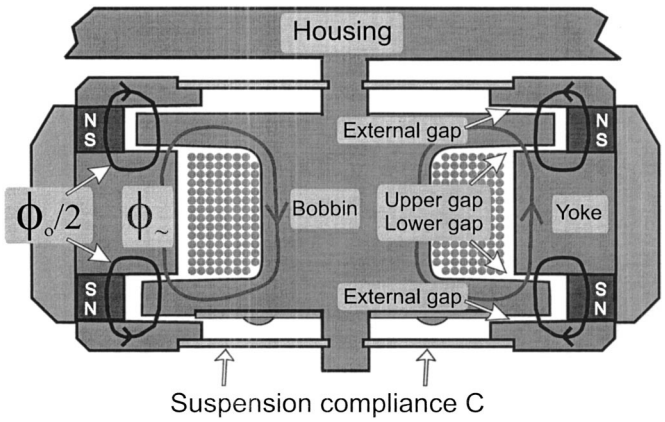


FIG. 4. Cross-sectional view of the BEST, in which four magnets and four internal and four external air gaps are used.

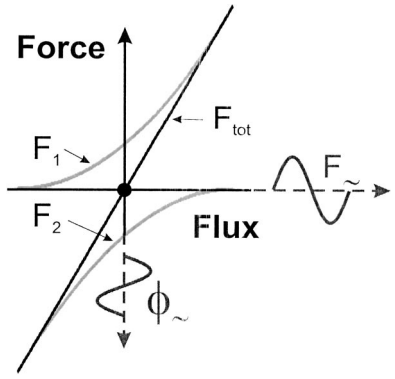


FIG. 5. The force versus magnetic flux characteristics of the BEST.

gaps, as compared with the design presented in Fig. 2. The BEST prototype used in all of the measurements presented in this study were, in principle, of this design.

Since the upper and lower magnets (Fig. 4) each generate the static magnetic flux $\phi_0/2$, the forces F_1 (lower gaps) and F_2 (upper gaps) on the bobbin, with reference to the yoke, are approximately following the equations:

$$F_1 \propto \left(\frac{\phi_0}{2}+\phi_\sim\right)^2 = \frac{\phi_0^2}{4}+\phi_\sim\cdot\phi_0+\phi_\sim^2, \tag{3}$$

$$F_2 \propto \left(\frac{\phi_0}{2}-\phi_\sim\right)^2 = \frac{\phi_0^2}{4}-\phi_\sim\cdot\phi_0+\phi_\sim^2. \tag{4}$$

The total force F_{tot} can be obtained by subtracting F_1 from F_2, which clearly shows that not only the static forces, but also the quadratic distortion forces, are cancelled or counterbalanced.

$$F_{\text{tot}}=F_1-F_2=F_{\text{signal}}\propto 2\cdot\phi_0\cdot\phi_\sim. \tag{5}$$

That the quadratic magnetic flux to force characteristic of the upper and lower air gaps appear as a linear curve after summation is also illustrated in Fig. 5. It should be noted in Fig. 5 that there is no remaining static force between the bobbin and the armature in the balanced position, as is also shown by Eq. (5). It is important to point out that the ideal behavior described by Eq. (5) requires that the assembly has perfect symmetries; i.e., equal air gaps, equal magnetization of magnets, bobbin exactly centered, etc. In a final implementation the degree of asymmetry will determine the amount of harmonic distortion and static force imbalance.

In addition to the principle of balanced suspension, there is another important feature of the BEST. The static and the dynamic magnetic fluxes are separated except in the vicinity of the air gaps where they must be superimposed. This separation of the static and dynamic magnetic fluxes (except in the air gap regions) has two specific advantages. First, the dynamic flux does not need to pass the permanent magnet, which generally has very poor dynamic properties. Second, as seen in Eq. (5), a high static flux is important in obtaining a high gain factor in the electro-mechanical conversion. However, a high static flux can result in local saturation of the soft iron material in the signal flux path. If there is a local saturation somewhere in the soft iron material, the dynamic

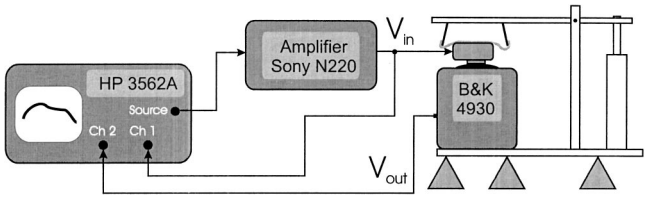


FIG. 6. Measurement setup for the frequency response and distortion measurements.

properties may deteriorate significantly. This problem is minimized in the BEST design, since the static and dynamic fluxes are separated.

Finally, it should be noted that there are some further minor design differences between the B71 and the BEST. To give the resonance of the BEST a better-controlled shape, some internal damping of the mechanical system is used. Furthermore, to improve efficiency and to reduce heat generation, the magnetic circuit is designed to minimize the magnetic losses.

In summary, the *balanced* suspension principle and the *separation* of the fluxes are key features of the new transducer shown in its name, the balanced electromagnetic separation transducer. An international patent application of this new transducer principle has been approved (Håkansson, 2001).

Finally, it should also be pointed out that Hunt (1982) described the general principles of a ''balanced armature'' transducer design. This principle design has also been successfully used in subminiature sound transducers for hearing aids; see, for example, Sebesta and Carlisle (1969).

III. METHODS

The setup that was used in the frequency response and distortion measurements is shown in Fig. 6. The equipment used includes: a two-channel signal analyzer (Hewlett Packard 3562A), a power amplifier (Sony TA-N220), and an artificial mastoid (Brüel & Kjær type 4930).

The B71 transducer was chosen arbitrarily from one of the audiometry test rooms at Sahlgren hospital (Göteborg, Sweden). The BEST was manufactured according to the previous description (see Fig. 4) and was incorporated into the housing of a B71 transducer. Hence, from the outside there was no visible difference between the BEST and the B71, but the weight of the BEST transducer element (without housing) was lighter: 10 g versus 15.3 g.

A. Frequency response

In the frequency response measurement, a swept sinusoid from 0.1 to 10 kHz with peak amplitude of 100 mV was used. The frequency response function $G_{\text{measured}}(j\omega)$ was calculated with the Hewlett Packard 3562A as

$$G_{\text{measured}}(j\omega)=G_{12}/G_{11}, \tag{6}$$

where G_{11} is the power spectrum of the input voltage to the transducer, and G_{12} is the cross-power spectrum between output voltage from the artificial mastoid and the input voltage to the transducer.

The artificial mastoid has a bone conductor attachment area or bilaminar rubber pad simulating skin and subcutaneous tissue properties. As the force gauge of the artificial mastoid is located below the bilaminar rubber pad, some corrections have to be made. To calculate the true frequency response function (force output/voltage input) of a transducer attached to the artificial mastoid, the measured frequency response function was corrected for the transmission through the rubber pad, and for the force to voltage sensitivity of the force gauge of the artificial mastoid. This correction required the frequency response function $A(j\omega)$ calculated as the output voltage of the artificial mastoid divided by the input force level to the surface of the rubber pad; and it is also required a measurement of the mechanical impedance (force/velocity) of the artificial mastoid rubber pad $Z(j\omega)$. The mechanical impedance $Z(j\omega)$ is needed to be compensated for the mass $m\approx 1.1$ g over the force gauge in the impedance head Brüel & Kjær 8000 which was used to measure $A(j\omega)$. The total response $G_{\text{real}}(j\omega)$, that is the output force level at the pad of the artificial mastoid divided by input voltage to the transducer, was then calculated as

$$G_{\text{real}}(j\omega)=G_{\text{measured}}(j\omega)\cdot\frac{Z(j\omega)}{Z(j\omega)+j\omega m}\cdot\frac{1}{A(j\omega)}. \tag{7}$$

B. Distortion

In the measurements of nonlinear distortion, a sinusoid of fixed frequency at 250, 315, 500, 750, 1000, 1500, 2000, 3000, and 4000 Hz was used. For each frequency the power spectrum of the input voltage and output force were measured and the Hewlett Packard 3562A calculated the total harmonic distortion (THD), where the harmonics up to 12 kHz was taken into account

$$\text{THD}=\frac{\text{Total harmonic power}}{\text{Fundamental power}}\cdot 100\%. \tag{8}$$

C. Electrical input impedance

The electrical input impedance, Z_{in}, of both transducers was also measured, using the Brüel & Kjær artificial mastoid type 4930 as the load. By using a resistor $R=10\ \Omega$ in series with the transducer and measuring the voltage at both the source side (V_S) and the transducer side (V_T) of the resistor, R, the impedance, Z_{in}, was calculated from a swept sinusoid measurement from 100–10k Hz

$$Z_{\text{in}}=\frac{V_T}{V_S-V_T}\cdot R=\frac{V_T}{V_S}\cdot\frac{1}{1-\dfrac{V_T}{V_S}}\cdot R, \tag{9}$$

where V_T/V_S was the measured frequency response function.

IV. RESULTS

A. Frequency response

The frequency response functions of the B71 and the BEST are shown in Fig. 7 as the magnitude of $G(j\omega)$ in dB. From this figure, it is clear that the resonance frequency is

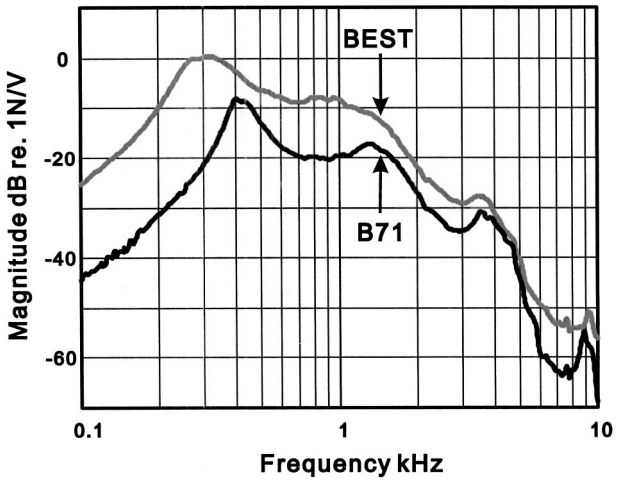


FIG. 7. Magnitude of the frequency response function $G(j\omega)$ of the B71 and the BEST.

reduced from approximately 420 Hz for the B71 to approximately 300 Hz for the BEST. Note that this decrease by a factor of 1.4 corresponds to a change by the same factor squared, in terms of mass, according to Eq. (2). As the BEST is actually lighter than the B71 by a factor of 1.53, the total mass factor is approximately threefold. This means that the BEST can be designed, at least theoretically, with a counterweight that is 3 times lighter than that of the B71 and still have the same resonance frequency.

Below 1000 Hz, the sensitivity is improved by 10–20 dB, except near the resonance frequency of the B71, where it is approximately 5 dB. Above 1000 Hz, the sensitivity is improved by 2–10 dB.

B. Distortion

The distortion measurements were made frequency by frequency, by measuring the force output power spectra of the B71 and the BEST, as for example at 250 Hz (fundamental frequency) where the spectra obtained are shown in Fig. 8. At this particular frequency the output force level was 40 dB HL (107 dB *re*: 1 μN). The THD was obtained by adding

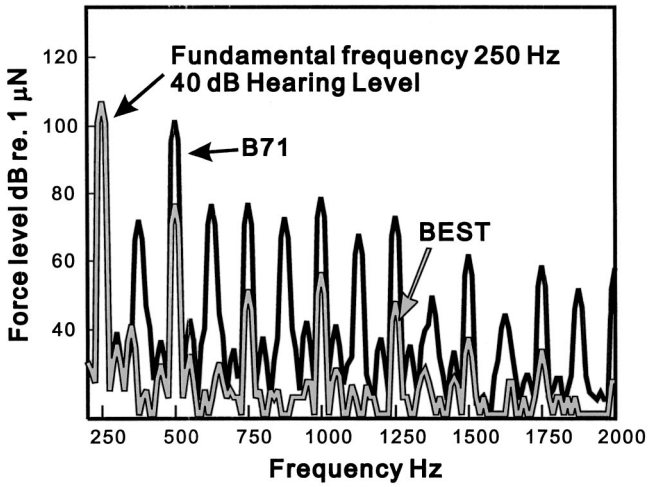


FIG. 8. The power spectrum of the output force when the fundamental frequency of the input signal is 250 Hz.

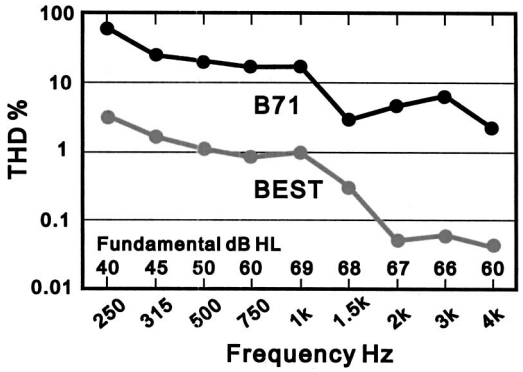


FIG. 9. The total harmonic distortion (THD) of the B71 and the BEST.

the power of all the harmonics (all peaks at multiples of 250 Hz) divided by the power of the fundamental frequency (250 Hz) according to Eq. (8). At 250 Hz, the THD was found to be 61% for the B71 and 3.3% for the BEST.

The major contribution to the THD at 250 Hz for the B71 and the BEST is from the second harmonic peak at 500 Hz, as is clear in Fig. 8. A peculiar phenomenon, appearing for the B71 only, is that additional peaks between the harmonics, i.e., at 375, 625 Hz, etc. can be easily seen. These peaks between the harmonics were present only in the measurement with the fundamental frequency of 250 Hz (Fig. 8). Also found was a small peak at the subfundamental frequency 125 Hz (not shown in Fig. 8). These nonharmonic peaks are obviously not intermodulation components between the harmonic frequencies and the power line frequency (e.g., 250 ± 50 Hz$\neq$125 or 375 Hz). No explanation for this phenomenon has been found so far.

The THDs for all fundamental frequencies tested are presented in Fig. 9. The level of the fundamental (dB hearing level) for each test frequency is shown by the figure above the frequency scale. Although it was originally planned to use the minimum output level requirements for type 1 audiometers (IEC 645-1, 1992), small changes were made at 250 and 500 Hz: a lower level was used to avoid overloading the B71 transducer. At 315 Hz no particular level is specified in the standard but, for the other frequencies, the specified level was used, although some of them were corrected for the transmission through the rubber pad of the artificial mastoid afterwards (at 1000–3000 Hz a reduction of 1–4 dB was used).

From Fig. 9 it can be seen that the THD in percent is more than 10 times lower (reduction by 20–25 dB) for the BEST than for the B71. For verification purposes, the THD of the input voltage fed to the transducers was also measured. It was found that this input THD was lower by a factor of 10 than at the output, for all frequencies. Hence, it can be concluded that all of the distortion presented in Fig. 9 is generated in the transducer.

C. Electrical input impedance

The magnitude of the electrical input impedances of the B71 and the BEST is shown in Fig. 10. At low frequencies the impedances are resistive and mainly determined by the ohmic losses in the coil wires. The dc impedances of the

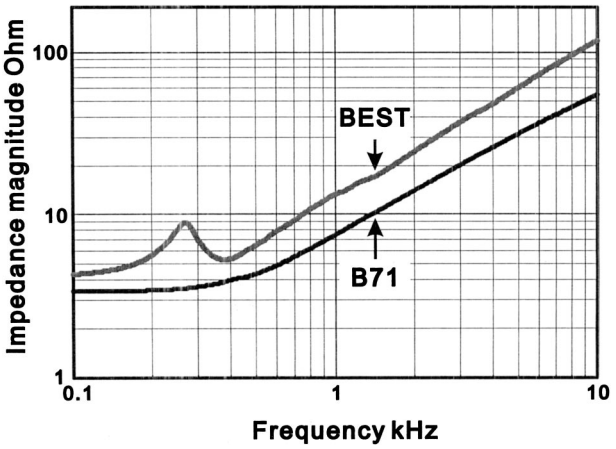


FIG. 10. The magnitude of the electrical input impedance of the B71 and the BEST.

BEST and the B71 are 4.0 and 3.3 ohms, respectively. At 1000 Hz the impedance magnitude of the BEST is some 1.8 times higher than that of the B71; above 1000 Hz, it is slightly more and below it is slightly less.

It is interesting to note that although the BEST has a generally higher sensitivity (Fig. 7), it consumes less current for a given voltage input. Hence, the BEST design is more efficient than the B71 by a factor that may be determined from the level difference in sensitivity in Fig. 7 ($BEST_{dB}$ minus $B71_{dB}$, where dB is defined as $20\times\log$[force/voltage]), plus a level correction for impedance from Fig. 10 that is calculated as $10\times\log$[impedance magnitude ratio BEST/B71]). Although this improvement is of moderate importance in audiometric applications, where the audiometer is powered from a power line, in hearing aids it is a dramatic one. Also, under some conditions, the thermal heating of the transducer element is crucial; this is lower with the more efficient BEST design.

V. DISCUSSION

Even if the improvements presented in Figs. 7–10 look promising, there are other requirements that have to be fulfilled to assure that the BEST is a successful new design. First, the transducer must be reliable over time. Although this has not yet been proven, one can assume that the balanced design, with static forces between moving parts counterbalanced, has the potential to be as reliable as the conventional designs. Furthermore, the coil in the BEST design is placed at the side opposite to that of the counterweight mass, in contrast to in the conventional design. The reason for doing so is that the counterweight side of the transducer is the most vibrant one. Consequently, there is a risk of fatigue in the electrical connection wires, if the coil is not rigidly attached to the driving side (skull side) of the transducer.

Second, it must be possible to manufacture the transducer at a reasonable cost and with a sufficient degree of reproducibility. To evaluate these aspects, a larger number of transducers must be manufactured. From the limited experience we have so far, it seems that the BEST is easier to assemble with respect to the air gap stability. In particular, the use of both internal and external air gaps, stabilizing the static magnetic flux, appears to provide a robust design and individual adjustments were not needed in the assembling of the prototypes.

It should also be noted that the present design is optimized for bone conduction hearing threshold measurements. Specifically, the present design is optimized to produce high output forces at low frequencies. An important aspect taken into account is that the transducer must have large enough air gaps to allow the larger deflections that occur at lower frequencies.

The balanced electromagnetic separation transducer (BEST) can be used in other applications and the optimization may then be different. Model parameters that may be modified are related to: air gap size, coil wires, mechanic and magnetic material properties. For example, designs for hearing threshold testing purposes (present study), implantable bone conduction hearing aids, ear level BAHA, and body-worn BAHA would all be optimized differently.

Planned and suggested future research includes: a clinical evaluation of the BEST for hearing threshold testing, development of better test and calibration methods, and finally, further improvement of the models of the transducer and the calibration equipment. The aim with this bone conduction methodological project is to significantly reduce the prevailing relatively large intra- and intersubject variability as well as the large interclinic variability in bone threshold testing.

In a concurrent project at our department, dealing with bone conduction physiology, attention is directed to understanding the underlying bone conduction mechanisms in an effort to improve the interpretation of bone conduction hearing threshold data.

VI. CONCLUSION

A new transducer design, the Balanced Electromagnetic Separation Transducer (BEST), is presented and evaluated. The results show that this transducer has the following advantages over the conventional B71 transducer:

(i) Lower total harmonic distortion (THD) by 20–25 dB;
(ii) Lower counterweight mass by a factor of 3 for the same resonance frequency;
(iii) Improved sensitivity by 10–20 dB for 100 to 1000 Hz and by 2–10 dB for 1 to 10 kHz; and
(iv) Improved efficiency, in addition to improved sensitivity, since the electrical input impedance is higher by some 1.8 times.

The BEST offers the opportunity to measure bone thresholds at 250 and 500 Hz with sufficiently good accuracy at hearing levels not possible before. For example, at 250 Hz the BEST has 23 dB higher sensitivity than the B71, and the THD is improved from 61% (B71) to 3.3% (BEST) at 40 dB HL for the transducers used in this study.

A clinical evaluation of the BEST and the development of improved test methods are the next steps in this project. The final goal is to improve accuracy so as to reduce the patient and calibration influenced variability in bone conduction hearing threshold data.

ACKNOWLEDGMENT

This study was supported by a grant from Stingerfonden, Göteborg, Sweden.

Dirks, D. (**1994**). ''Bone conduction threshold testing,'' in *Handbook of Clinical Audiology*, 4th ed., edited by J. Katz (Lippincott Williams & Wilkins, Philadelphia), pp. 132–146.

Dirks, D., and Kamm, C. (**1975**). ''Bone-vibrator measurements: Physical characteristics and behavioral thresholds,'' J. Speech Hear. Res. **18**(2), 242–260.

Dolan, T., and Morris, S. (**1990**). ''Administering audiometric speech tests via bone conduction: A comparison of transducers,'' Ear Hear. **11**(6), 446–449.

Hunt, F. V. (**1982**). ''Electroacoustics: The analysis of transduction and its historical background,'' published by the American Institute of Physics for the Acoustical Society of America, 1954 (reprinted 1982), pp. 213–235.

Håkansson, B. (**2001**). Electromagnetic vibrator, International Patent Application No. PCT/SE01/00484 (priority from 9 March 2000, SE 0000810-2) issued 13 September 2001.

Håkansson, B., Tjellström, A., and Carlsson, P., (**1990**). ''Percutaneous vs transcutaneous transducers for hearing by direct bone conduction,'' Otolaryngol. Head & Neck Surg., **102**, 339–344.

Håkansson, B., and Carlsson, P. (**1985**). ''Anordning vid för en vibrator till en hörapparat eller annan för ljudstimulering av ben anordnad vibrationsalstrare avsedd fjäder,'' Swedish Patent No. SE 8502426, (filed 15 May) issued 22 December 1985, Title in English: ''Arrangement of a spring suspension for hearing aid transducers.''

IEC 645-1 (**1992**). International standard, ''Audiometers, Part 1: Pure-tone audiometers.''

ISO 389-3 (**1994**). International standard, ''Acoustics—Reference zero for the calibration of audiometric equipment, Part 3: Reference equivalent threshold force levels for pure tones and bone vibrators.''

Lightfoot, G. R., and Hughes, J. B. (**1993**). ''Bone conduction errors at high frequencies: implications for clinical and medico-legal practice,'' J. Laryngol. Otol. **107**(4), 305–308.

Parving, A., and Elberling, C. (**1982**). ''High-pass masking in the classification of low-frequency hearing loss,'' Scand. Audiol. **11**(3), 173–178.

Sebesta, G. J., and Carlisle R. W. (**1969**). Sub-miniature sound transducers, U. S. Patent No. 3,432,622 (filed 10 May 1965) issued 11 March 1969.

Tjellström, A., Håkansson, B., and Granström, G. (**2001**). ''The bone-anchored hearing aids: Current status in adults and children,'' Otolaryngol. Clin. North Am. **34**(2), 337–364.

Numerical homogenization techniques applied to piezoelectric composites

Eve Lenglet
ONERA-Lille, Solid and Damage Mechanics Department, 5 Boulevard Paul Painlevé, 59045 Lille Cedex, France

Anne-Christine Hladky-Hennion[a] and Jean-Claude Debus
IEMN, ISEN Department (UMR 8520 CNRS), 41 Boulevard Vauban, 59046 Lille Cedex, France

(Received 29 January 2002; revised 13 September 2002; accepted 22 November 2002)

With the recent availability of piezoelectric fibers, the design and the analysis of piezoelectric composites needs new modeling tools. Therefore, a numerical homogenization technique has been developed, based on the ATILA finite element code, that combines two techniques: one relying upon the representative volume element (RVE) the other relying upon the wave propagation (WP). The combination of the two methods allows the whole tensor of the homogenized properties of the piezoelectric composite to be found. Considering a fiber embedded in epoxy, the numerical results are compared to the results obtained using previous analytical models, thus validating the models. Even if the method is presented in a particular case, its extension to any piezoelectric composite is straightforward. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1537710]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.20.Bi, 43.35.Cg [SLE]

I. INTRODUCTION

Piezoelectric ceramics, such as lead zirconate–titanate (PZT), are widely used for transducers in sonar, underwater communications, underwater or medical imaging applications, as well as for control application. Even if their properties make them interesting, they are often limited, first by their weight, that can be a clear disadvantage for shape control, and, as a consequence, by their high specific acoustic impedance, which reduces their acoustic matching with the external fluid domain.

For the last 20 years, composite piezoelectric materials have been developed by combining piezoceramics with passive nonpiezoelectric polymers.[1–3] Superior properties have been achieved by these composites by taking advantage of the most profitable properties of each of the constituents and a great variety of structures have been made. Many models are available to describe these piezocomposites.[4–6]

Recently, due to the miniaturization of the piezocomposites and the use of PZT fibers instead of piezoelectric bars,[7–10] new applications toward electromechanical sensors and actuators have become possible. But, because they are now much smaller than the wavelength, homogenization techniques are necessary to describe the behavior of piezocomposites.

Even if analytical and semianalytical models have been developed to homogenize piezocomposites, they are often reduced to specific cases. Numerical models seem to be a well-suited approach to describe the behavior of these materials, because there is no restriction on the geometry, on the material properties, on the number of phases in the piezocomposite, and on the size. Therefore, the finite element method, with the help of the ATILA code,[11] has been used to determine the effective properties of the piezocomposites and is the basis of the simulation tool used in this paper.

After a brief description of previous homogenization techniques, the numerical homogenization technique is presented, that combines two techniques: one relying upon the representative volume element (RVE), the other relying upon the wave propagation (WP). The combination of the two methods allows the whole tensor of the effective properties of the piezocomposite to be found, without any complicated mesh or complicated boundary conditions. Then, as a validation, considering a fiber embedded in epoxy, the results are compared with the results obtained with the help of previous homogenization techniques. Even if the method is presented in a specific case, its extension to any piezoelectric composite is straightforward.

II. BRIEF DESCRIPTION OF PREVIOUS HOMOGENIZATION TECHNIQUES

Considering a piezocomposite material, containing a piezoelectric part and a polymer part, the homogenization allows the material to be considered as a piezoelectric material with equivalent properties, called effective properties. In this case, 11 constants have to be determined: 6 elastic constants ($s_{11}^{E\,\mathrm{eff}}$, $s_{12}^{E\,\mathrm{eff}}$, $s_{13}^{E\,\mathrm{eff}}$, $s_{33}^{E\,\mathrm{eff}}$, $s_{44}^{E\,\mathrm{eff}}$, and $s_{66}^{E\,\mathrm{eff}}$); three piezoelectric constants (d_{15}^{eff}, d_{31}^{eff}, and d_{33}^{eff}); and two dielectric constants ($\varepsilon_{11}^{T\,\mathrm{eff}}$ and $\varepsilon_{33}^{T\,\mathrm{eff}}$):

[a]Electronic mail: anne-christine.hladky@isen.fr

$$\begin{Bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ D_1 \\ D_2 \\ D_3 \end{Bmatrix} = \begin{bmatrix} s_{11}^{E\,\mathrm{eff}} & s_{12}^{E\,\mathrm{eff}} & s_{13}^{E\,\mathrm{eff}} & 0 & 0 & 0 & 0 & 0 & -d_{31}^{\mathrm{eff}} \\ s_{12}^{E\,\mathrm{eff}} & s_{22}^{E\,\mathrm{eff}} & s_{13}^{E\,\mathrm{eff}} & 0 & 0 & 0 & 0 & 0 & -d_{31}^{\mathrm{eff}} \\ s_{13}^{E\,\mathrm{eff}} & s_{13}^{E\,\mathrm{eff}} & s_{33}^{E\,\mathrm{eff}} & 0 & 0 & 0 & 0 & 0 & -d_{33}^{\mathrm{eff}} \\ 0 & 0 & 0 & s_{44}^{E\,\mathrm{eff}} & 0 & 0 & 0 & -d_{15}^{\mathrm{eff}} & 0 \\ 0 & 0 & 0 & 0 & s_{44}^{E\,\mathrm{eff}} & 0 & -d_{15}^{\mathrm{eff}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{66}^{E\,\mathrm{eff}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & d_{15}^{\mathrm{eff}} & 0 & \varepsilon_{11}^{T\,\mathrm{eff}} & 0 & 0 \\ 0 & 0 & 0 & d_{15}^{\mathrm{eff}} & 0 & 0 & 0 & \varepsilon_{11}^{T\,\mathrm{eff}} & 0 \\ d_{31}^{\mathrm{eff}} & d_{31}^{\mathrm{eff}} & d_{33}^{\mathrm{eff}} & 0 & 0 & 0 & 0 & 0 & \varepsilon_{33}^{T\,\mathrm{eff}} \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \\ E_1 \\ E_2 \\ E_3 \end{Bmatrix}, \tag{1}$$

where S_i is the strain, D_i is the electrical displacement, T_i is the stress, and E_i is the electrical field, using the condensed notation.

Using other systems for the equations of piezoelectricity, one can determine other constants, such as the c^E stiffness constants, the e piezoelectric constants, and the ε^S dielectric constants. It is easy to transform one set of constants into another.[12] Many homogenization techniques are presented in this section.

First, the Schulgasser's method,[13] relying upon the rule of mixture, is the simplest method but only gives five constants and a relation between two others. Then, Schulgasser[13] has improved the model; thus three more constants are known by using the Milgrom–Shtrikman results.[14] But the $c_{66}^{E\,\mathrm{eff}}$ coefficient is still unknown. These methods do not take into account the shape of the inclusion.

Newnham *et al.*[15] have developed an analytical formula to calculate the piezoelectric constants and a dielectric constant in a 1–3 piezocomposite. In the same way, Smith *et al.*[4] have determined some of the effective properties of a 1–3 piezocomposite when the thickness mode is excited.

Hashimoto and Yamaguchi[16] suggest that some field components are uniform along the length of a one-dimensional periodic multilayered structure and then the macroscopic properties of the structure could basically be characterized by the average value of the field quantities.

Two models are proposed by Bent and Hagood[7] The first one relies upon the Uniform Field Method (UFM) and its originality is that it takes into account the electrodes. Thus, it is particularly well suited to actuators. The effective properties of the actuators are determined using parallel and series additions. It has to be noted that this approach violates compatibility and equilibrium at some material interfaces. Only four elastic constants, two piezoelectric constants and one dielectric constant are obtained and the shape of the inclusion is simplified as a cube. The second model relies upon the finite element method, with specific boundary conditions. The same volume is meshed and the same coefficients are determined. The results obtained by the two methods agree well.

Using the finite element method, Poizat[17] has meshed the unit cell of the material and by applying specific boundary conditions, he has determined two piezoelectric coefficients.

Then, Benveniste and Dvorak[18] consider a binary composite medium. The constituents are transversely isotropic and the phase boundaries are cylindrical. He expresses the strain and the electric intensity with the influence functions and uniform fields. The effective constants can be explained with the volume fraction and the concentration factors, which are the averages of the influence functions.

For given geometry of the inclusion, two methods give the concentration factors using Eshelby's tensor:[19] the dilute method and Tanaka's method.[20] The dilute method is valid for a small volume fraction of inclusion. Tanaka's method is valid for all volume fractions. Therefore, in this paper, it will be the reference for the comparison of the results.

Another class of homogenization technique is the periodic homogenization. Abgossou *et al.*[21,22] have developed a numerical method based on the variational formulation. Additional macroscopic degrees of freedom, related to the averaged displacement and electrical potential, are introduced

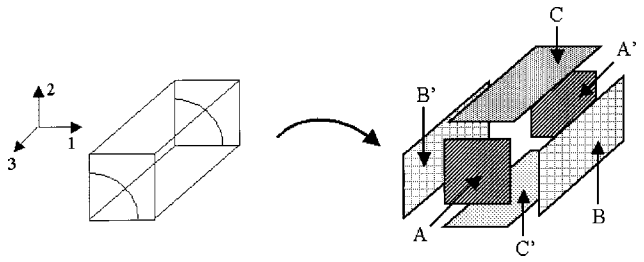


FIG. 1. One-quarter of the representative volume elementary (RVE) and description of the six faces delimiting the RVE. Symmetry conditions are applied on faces A′, B′, and C′.

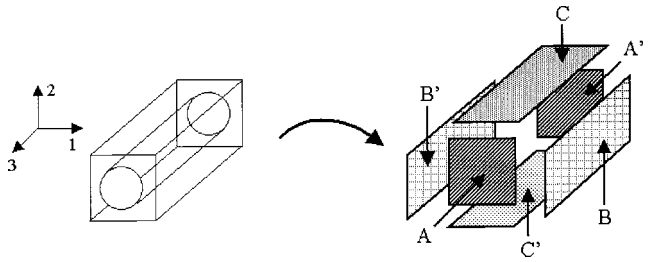


FIG. 2. Whole representative volume elementary (RVE) and a description of the six surfaces delimiting the RVE. There are no symmetry conditions in the RVE.

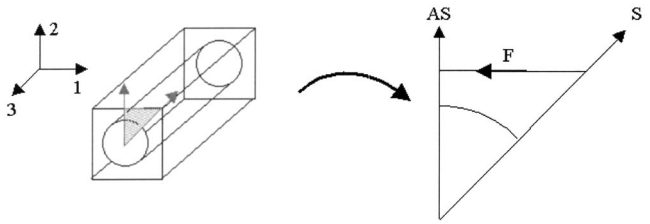


FIG. 3. One-eighth of the cross section of the representative volume elementary (RVE). One symmetry plane (S) and one antisymmetry plane (AS) are considered.

into the system, leading to an iterative resolution method. Moreover, the introduction of these additional degrees of freedom can lead to unsatisfactory conditioning of the final system matrix.

III. NUMERICAL HOMOGENIZATION TECHNIQUE

With a view to avoiding the restrictions of the previous methods, two numerical homogenization techniques, RVE and WP, are proposed, that use the ATILA finite element code.[11] Even if the RVE method can be used with other finite element codes, the WP method is available in the ATILA finite element code. Each approach gives a part of the whole tensor and the combination of the two methods allows us to know easily the full set of constants. For the sake of simplicity and as an example, piezoelectric fibers embedded in epoxy are considered. The polarization direction is in the fiber direction, which is named direction 3. In that case, due to the symmetry of the structure, many simplifications can be made and, in particular, reduced meshes can be used in the RVE method. But the numerical methods described here can be used for other structures, made of more than two phases, of any shape, with the superposition of different slices of fibers with a different orientation.

A. Representative volume elementary (RVE) method

In the RVE method, only one unit cell of the structure is considered. Depending on the effective constant that has to be determined, three different RVE or meshes are used (Figs. 1–3). In all the cases, isoparametric elements (hexahedral, prismatic, quadrilateral, or triangular) are used. Thus, complex structures with curved sides or faces can be modeled using a reduced number of elements. The finite element formulation used here relies upon quadratic interpolation functions along element sides. Several meshes have been tested to ensure the convergence of the results. In Fig. 1, a tridimensional mesh is considered. Due to symmetry planes, only a quarter of the unit cell is meshed, using three symmetry planes (faces A′, B′, and C′). Classical symmetry conditions are applied on these faces, whereas displacement on faces A, B, and C are constrained to be uniform, so the faces remain plane and continuity is preserved with the next cell. In Fig. 2, a whole tridimensional mesh of the unit cell is considered. No symmetry planes are applied but the displacement on each of the six faces is constrained to be uniform. Finally, in Fig. 3, a bidimensional mesh is considered and only one-eighth of the cross section of the unit cell is meshed, thanks to one symmetry plane and one antisymmetry plane.

TABLE I. Determination of the effective constants using the RVE method. The column "unit cell" is referred to the unit cell used to determine the constant (Figs. 1–3). Force or electrical potential is applied.

Constant	Force direction	Electrical field direction	Calculation	Unit cell	Electrodes
$s_{11}^{E\,\mathrm{eff}}$	1	⋯	S_1/T_1	1	⋯
$s_{12}^{E\,\mathrm{eff}}$	2	⋯	S_1/T_2	1	⋯
$s_{13}^{E\,\mathrm{eff}}$	3	⋯	S_1/T_3	1	⋯
$s_{33}^{E\,\mathrm{eff}}$	3	⋯	S_3/T_3	1	⋯
$s_{44}^{E\,\mathrm{eff}}$	3	⋯	S_4/T_4	2	⋯
$s_{66}^{E\,\mathrm{eff}}$	1	⋯	S_6/T_6	3	⋯
d_{31}^{eff}	⋯	3	S_1/E_3	2	A, A′
d_{33}^{eff}	⋯	3	S_3/E_3	2	A, A′
d_{15}^{eff}	3	⋯	D_1/T_5	2	⋯
$\varepsilon_{11}^{T\,\mathrm{eff}}$	⋯	1	D_1/E_1	2	B, B′
$\varepsilon_{33}^{T\,\mathrm{eff}}$	⋯	3	D_3/E_3	1	A, A′

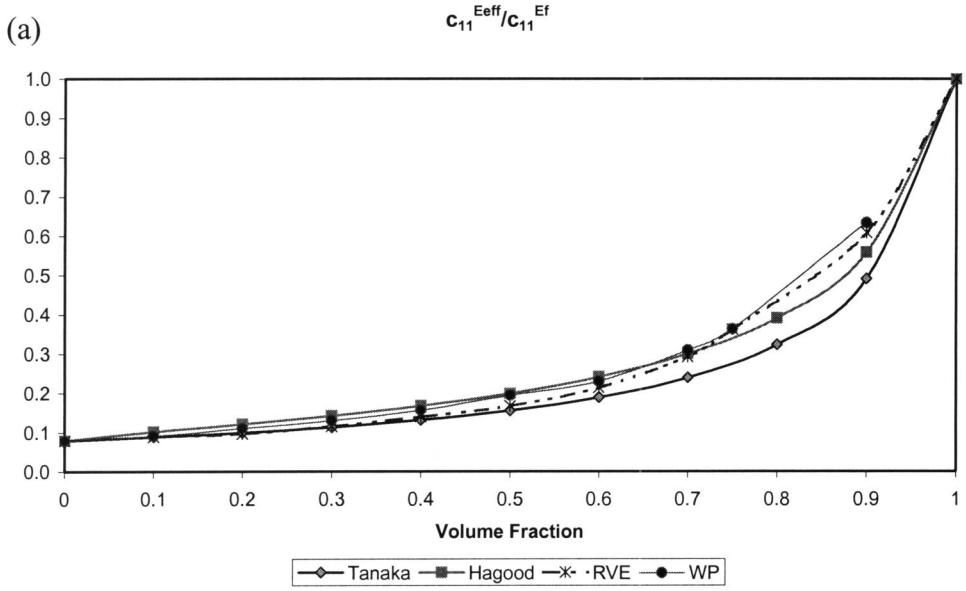
By applying appropriate boundary conditions on the surfaces of the RVE, it is possible to determine the effective constants of the piezocomposite. The first part of Table I presents the applied boundary conditions used to obtain the elastic constants. In this case, the electrical potential ϕ is equal to zero in the whole domain; there are no electrodes. For instance, to calculate $s_{33}^{E\,\mathrm{eff}}$, using the notations of Fig. 1, a F_3 force is applied on the A surface, in direction 3. With the help of the displacement in direction 3, the strain S_3 is calculated. The stress T_3 in direction 3 is known as the ratio between F_3 and the cross-section surface. Then, the $s_{33}^{E\,\mathrm{eff}}$ coefficient is deduced as the ratio between S_3 and T_3. To determine the shear constant $s_{44}^{E\,\mathrm{eff}}$ the whole RVE is considered (Fig. 2), but a bidimensional mesh is enough to evaluate $s_{66}^{E\,\mathrm{eff}}$ (Fig. 3).

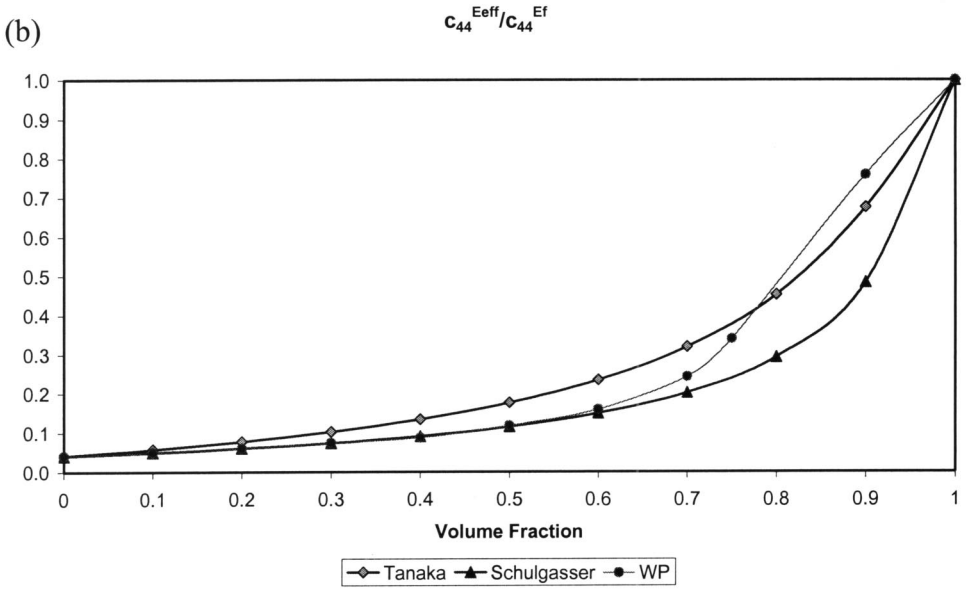
In the same way, Table I also presents the applied boundary conditions to obtain the piezoelectric and the dielectric constants, respectively. In the case of $\varepsilon_{33}^{T\,\mathrm{eff}}$, electrodes are considered on faces A and A′. The ATILA finite element code[11] gives the capacitance C. Then, the electrical displacement D_3 is calculated and $\varepsilon_{33}^{T\,\mathrm{eff}}$ is deduced as the ratio between D_3 and E_3.

Using the RVE method, even if all the constants of the tensor can be determined, it needs the application of boundary conditions on numerous nodes of the mesh, especially for

TABLE II. Identification of the constants using the RVE method and the WP method. X means that the constant is easily determined, 0 means that the constant can be determined but needs complicated boundary conditions, Y means that the constant is obtained using the combination of the RVE method and the WP method.

Constant	$c_{11}^{E\,\mathrm{eff}}$	$c_{12}^{E\,\mathrm{eff}}$	$c_{13}^{E\,\mathrm{eff}}$	$c_{33}^{E\,\mathrm{eff}}$	$c_{44}^{E\,\mathrm{eff}}$	$c_{66}^{E\,\mathrm{eff}}$	d_{31}^{eff}	d_{33}^{eff}	d_{15}^{eff}	$\varepsilon_{11}^{T\,\mathrm{eff}}$	$\varepsilon_{33}^{T\,\mathrm{eff}}$
RVE	X	X	X	X	0	X	X	X	0	X	X
WP	X				X	X			Y		





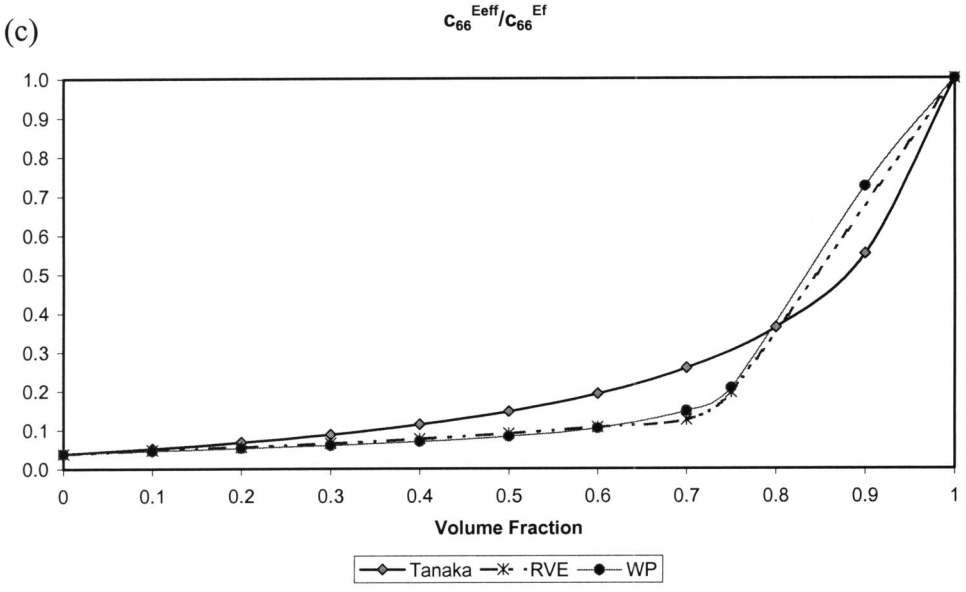
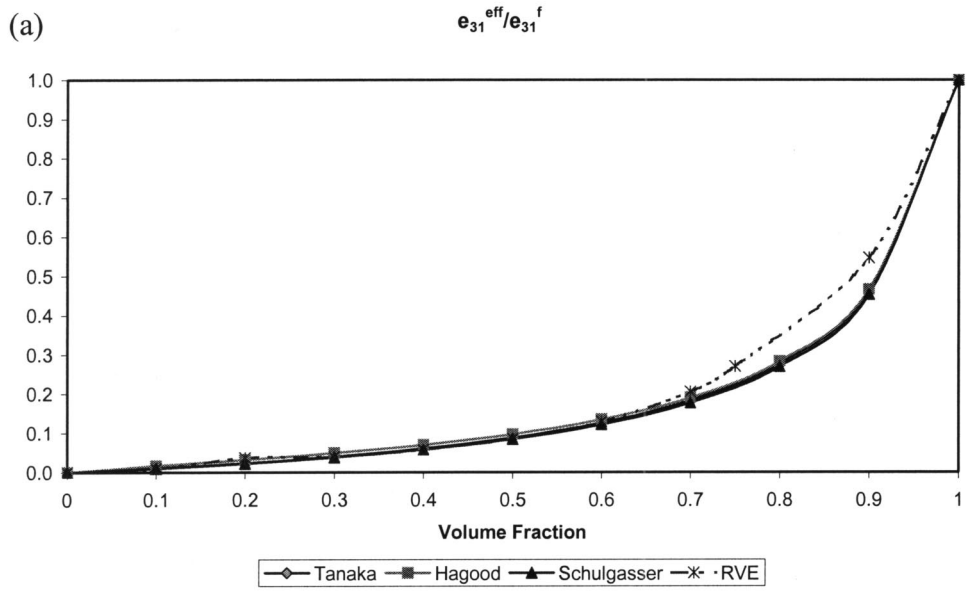
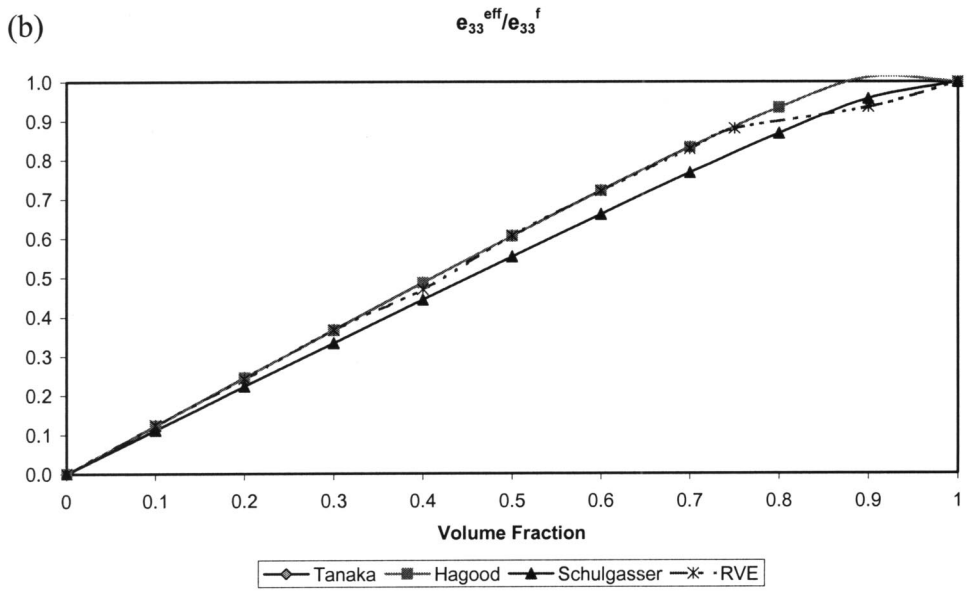


FIG. 4. Variations of the effective elastic constant as a function of the volume fraction of the PZT5A fiber in the RVE. A comparison between numerical results (RVE and WP) and analytical results from Tanaka (Ref. 20), Hagood (Ref. 7), and Schulgasser (Ref. 13). (a) c_{11}^{E}, (b) c_{44}^{E}, (c) c_{66}^{E}. Each coefficient is normalized to the ceramic coefficient.





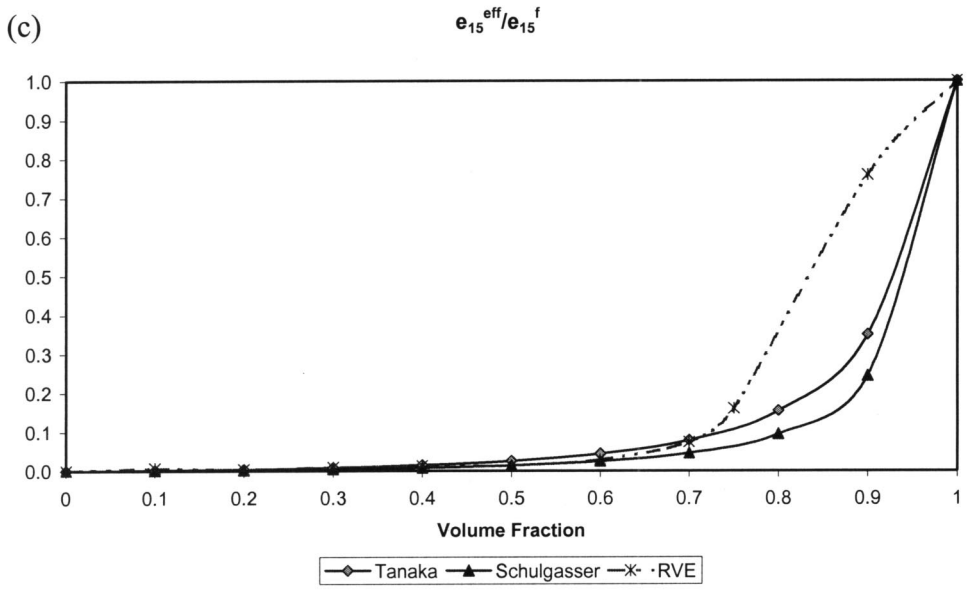


FIG. 5. Variations of the effective piezoelectric constant as a function of the volume fraction of the PZT5A fiber in the RVE. A comparison between numerical results (RVE and WP) and analytical results from Tanaka (Ref. 20), Hagood (Ref. 7), and Schulgasser (Ref. 13). (a) e_{31}, (b) e_{33}, (c) e_{15}. Each coefficient is normalized to the ceramic coefficient.

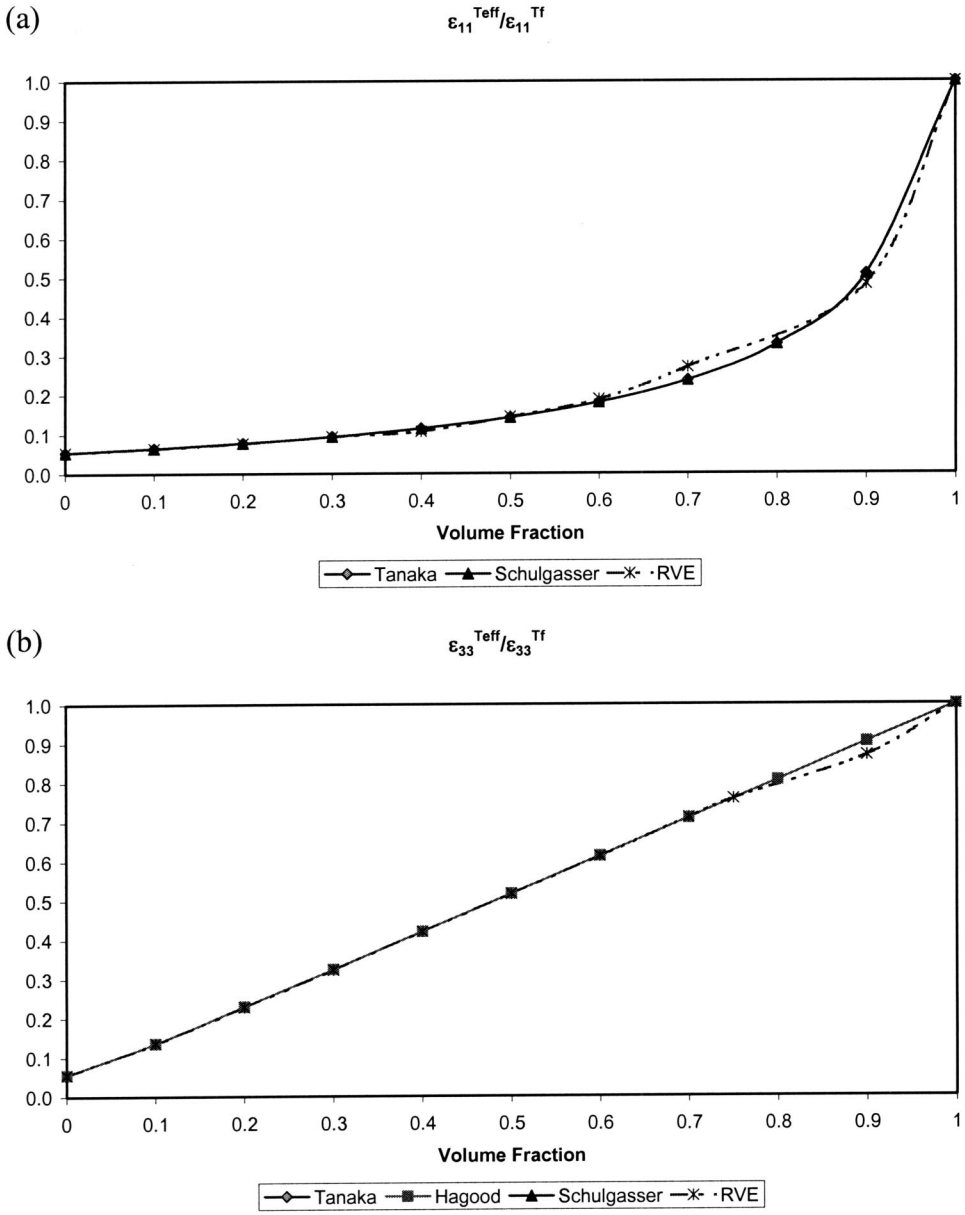


FIG. 6. Variations of the effective dielectric constant as a function of the volume fraction of the PZT5A fiber in the RVE. A comparison between numerical results (RVE) and analytical results from Tanaka (Ref. 20), Hagood (Ref. 7), and Schulgasser (Ref. 13). (a) ε_{11}^{T}, (b) ε_{33}^{T}. Each coefficient is normalized to the ceramic coefficient.

the knowledge of the $s_{44}^{E\,\mathrm{eff}}$ and d_{15}^{eff} coefficients, due to T_4 and T_5. In particular, the applied force has to be weighted on each node of the mesh. This gives a specific treatment of the data and complicated applied boundary conditions. Therefore, for the knowledge of some constants, the wave propagation method is used.

B. Wave propagation (WP) method

The wave propagation method has already been used to homogenize the properties of an elastic periodic structure.[23] Here, the method is generalized to a piezoelectric periodic structure and the corresponding numerical method is available in the ATILA[11] finite element code. A periodic material in three space directions is completely described using one unit cell and three vectors. The Bloch–Floquet theorem allows the modeling of the material to be reduced to only one unit cell, which is meshed using finite elements. For a given wave vector **k**, phase relations between nodes spaced one period are determined and provide boundary conditions between adjacent cells. The resolution of the system gives the eigenvalues and the corresponding eigenvectors, which are the angular frequencies ω and the corresponding displacement field and electrical potential. Then, dispersion curves are obtained by varying the direction and the modulus of the wave vector **k**, and give information on the propagation modes, pass bands, and stop bands. The three low-frequency branches of the curves can be approximated to three straight lines whose slopes are denoted first quasilongitudinal velocity (c_L) and then quasitransverse velocities (c_{T1} and c_{T2}). These velocities depend on the direction of the wave vector **k**. With a view to calculating the effective constants of the material in the large wavelength limit, first the three velocities are calculated using the dispersion curve, and then a system based on classical Christoffel equations has to be solved.[24,25] As an example, considering a wave propagating in the 13 plane, where angle θ is defined with respect to axis 3, it gives

$$\rho c_{T1}^2=\frac{\overline{\Gamma}_{11}+\overline{\Gamma}_{33}}{2}-\frac{1}{2}\sqrt{(\overline{\Gamma}_{11}-\overline{\Gamma}_{33})^2+4\overline{\Gamma}_{13}^2},$$

$$\rho c_{T2}^2=\overline{\Gamma}_{22}, \tag{2}$$

$$\rho c_{L}^2=\frac{\overline{\Gamma}_{11}+\overline{\Gamma}_{33}}{2}+\frac{1}{2}\sqrt{(\overline{\Gamma}_{11}-\overline{\Gamma}_{33})^2+4\overline{\Gamma}_{13}^2},$$

with

$$\overline{\Gamma}_{11}=c_{11}^E(\cos\theta)^2+c_{44}^E(\sin\theta)^2+\frac{(e_{15}+e_{31})^2(\cos\theta)^2(\sin\theta)^2}{\varepsilon_{11}^S(\cos\theta)^2+\varepsilon_{33}^S(\sin\theta)^2},$$

$$\overline{\Gamma}_{22}=c_{66}^E(\cos\theta)^2+c_{44}^E(\sin\theta)^2,$$

$$\overline{\Gamma}_{13}=(c_{13}^E+c_{44}^E)\cos\theta\sin\theta+\frac{(e_{15}+e_{31})\cos\theta\sin\theta[e_{15}(\cos\theta)^2+e_{33}(\sin\theta)^2]}{\varepsilon_{11}^S(\cos\theta)^2+\varepsilon_{33}^S(\sin\theta)^2}, \tag{3}$$

$$\overline{\Gamma}_{33}=c_{44}^E(\cos\theta)^2+c_{33}^E(\sin\theta)^2+\frac{[e_{15}(\cos\theta)^2+e_{33}(\sin\theta)^2]^2}{\varepsilon_{11}^S(\cos\theta)^2+\varepsilon_{33}^S(\sin\theta)^2}.$$

Using relations (2) and (3), it is easy to determine c_{11}^E, c_{66}^E ($\theta=0$), and c_{44}^E ($\theta=90°$). The system gives also the value of $(e_{15})^2/\varepsilon_{11}^S$. With the help of the coefficients previously obtained using the RVE method, e_{15} can be deduced. This illustrates the interest of the combination of the two numerical methods.

C. Short review of obtained constants using the RVE and WP methods

As a conclusion to the presentation of the RVE method and of the WP method, Table II presents for each method which coefficients can be obtained. It clearly shows that the methods are complementary. Moreover, some coefficients can be obtained using both methods; thus it can be used as a validation.

IV. VALIDATION

First, a PZT5A fiber embedded in epoxy is considered. The fiber section is circular. By varying the diameter of the fiber in the unit cell, effective coefficients of the material are numerically obtained and are presented on Figs. 4–6 as a function of the volume fraction. For a low volume fraction of the fiber in the unit cell, the effective coefficients are equal to those of epoxy, whereas for a high volume fraction, the effective coefficients are equal to those of the ceramic. Due to the geometry of the fiber, for a volume fraction greater than 0.78, there is contact between adjacent fibers and the fiber section is no longer circular. For volume fractions close to 0.78, the finite element mesh has to be denser because of the small thickness of epoxy around the fiber. Even if the whole set of constants is determined, only the most important are presented. Figures 4–6 compare the coefficients obtained numerically and using various analytical methods. Each coefficient is normalized to the corresponding ceramic coefficient. Particular attention is given to c_{11}^E and c_{66}^E that can be obtained by both RVE and WP methods (Fig. 4). The two numerical methods used in this paper give very close results. The c_{44}^E coefficient is only obtained with the WP method. e_{15} is obtained using a combination of RVE and WP methods. Figures 4–6 show that analytical and numerical methods give results in good agreement, particularly if the volume fraction is lower than 0.70. It can be noticed that Hagood's model does not give c_{44}^E, c_{66}^E, e_{15}, and ε_{11}^T, whereas Schulgasser's model does not give c_{11}^E, c_{12}^E, and c_{66}^E. Only Tanaka's method together with the numerical model give the whole set of constants. The main limitation of Tanaka's method is that the shape of the inclusion is taken into account with the help of the Eshelby's tensor, whereas any geometry can be considered in the numerical results. Figures 4–6 validate the model.

V. CONCLUSIONS

A numerical model for predicting the homogenized properties of piezocomposites has been presented. It relies upon the finite element method, with the help of the ATILA code. The method uses the combination of the RVE and the WP methods to obtain easily the whole set of the homogenized properties. No restrictive hypotheses are assumed with respect to the structure geometry, material properties, and number of phases in the structure. Moreover, modifying the structure only requires the modification of the mesh, without any new development related to the method. Then, the method has been applied and validated on the classical case of a PZT fiber embedded in epoxy. Now, the method is used for other structures, where the miniaturization implies the use of homogenization techniques, such as in the case of 1–3 piezocomposites or active fiber composites (AFC).[7]

[1] R. E. Newnham, D. P. Skinner, and L. E. Cross, "Connectivity and piezoelectric–pyroelectric composites," Mater. Res. Bull. **13**, 525–536 (1978).

[2] T. R. Gururaja, W. A. Schulze, L. E. Cross, R. E. Newnham, B. A. Auld, and Y. J. Wang, "Piezoelectric composite materials for ultrasonic transducer applications. Part I: resonant modes of vibration of PZT rod–polymer composites," IEEE Trans. Sonics Ultrason. **SU-32**, 481–498 (1985).

[3] T. R. Gururaja, W. A. Schulze, L. E. Cross, and R. E. Newnham, "Piezoelectric composite materials for ultrasonic transducer applications. Part II: evaluation of ultrasonic medical applications," IEEE Trans. Sonics Ultrason. **SU-32**, 499–513 (1985).

[4] W. A. Smith, A. A. Shaulov, and B. M. Singer, "Properties of composite piezoelectric materials for ultrasonic transducers," in *Proceedings of the IEEE Ultrasonics Symposium* (IEEE, New York, 1984), pp. 539–544.

[5] G. Hayward and J. A. Hossack, "Unidimensional modeling of 1–3 composite transducers," J. Acoust. Soc. Am. **88**, 599–608 (1990).

[6] D. Leedom, R. Krimholtz, and G. Matthei, "Equivalent circuits for transducers having arbitrary even- or odd-symmetry piezoelectric excitation," IEEE Trans. Sonics Ultrason. **SU-18**, 128–141 (1971).

[7] A. A. Bent and N. W. Hagood, "Piezoelectric fiber composites with interdigitated electrodes," J. Intell. Mater. Syst. Struct. **8**, 903–919 (1997).

[8] W. Watzka, S. Seifert, H. Scholz, D. Sporn, A. Schönecker, and L. Seffner, "Dielectric and ferroelectric properties of 1–3 composites containing thin PZT-fibers," in *Proceedings of the 10th International Symposium on Applied Ferroelectrics*, ISAF'96, edited by B. M. Kulwicki, A. Amin, and A. Safari (1996), Vol. 2, pp. 569–572.

[9] R. Steinhausen, T. Hauke, H. Beige, W. Watzka, U. Lange, D. Sporn, S. Gebhardt, and A. Schönecker, "Properties of fine scale piezoelectric PZT

fibers with different Zr content,'' J. Eur. Ceram. Soc. **21**, 1459–1462 (2001).
[10] S. S. Livneh, V. F. Janas, and A. Safari, ''Development of fine scale PZT ceramic fiber/polymer shell composite transducers,'' J. Am. Ceram. Soc. **78**, 1900–1906 (1995).
[11] ATILA Finite Element Code for Piezoelectric and Magnetostrictive Transducer Modeling, developed by the Acoustics Department of ISEN, ATILA Version 5.2.2, User's manual, ISEN, Acoustics Department, Lille, France, 2002.
[12] O. B. Wilson, *Introduction to Theory and Design of Sonar Transducers* (Peninsula, Los Altos, 1989).
[13] K. Schulgasser, ''Relationships between the effective properties of transversely isotropic piezoelectric composites,'' J. Mech. Phys. Solids **40**, 473–479 (1992).
[14] M. Milgrom and S. Shtrikman, ''Linear response of two phase composites with cross moduli: exact numerical relations,'' Phys. Rev. A **40**, 1568–1575 (1989).
[15] R. E. Newnham, L. J. Bowen, K. A. Klicker, and L. E. Cross, ''Composite piezoelectric transducers,'' Mat. Eng. **2**, 96–106 (1980).
[16] K. Y. Hashimoto and M. Yamaguchi, ''Elastic, piezoelectric and dielectric properties of composite materials,'' IEEE Ultrason. Symp., 697–702 (1986).
[17] C. Poizat, ''Numerical modelling of composites and structures with embedded piezoelectric fibers (Modélisation numérique de matériaux et structures composites à fibres piézoélectriques),'' Ph.D. thesis, Université de Technologie de Troyes, France, 2000.
[18] Y. Benveniste and G. J. Dvorak, ''Uniform fields and universal relations in piezoelectric composites,'' J. Mech. Phys. Solids **40**, 1295–1312 (1992).
[19] J. D. Eshelby, ''The determination of the elastic field of an ellipsoidal inclision and related problems,'' Proc. R. Soc. London, Ser. A **241**, 376–396 (1957).
[20] M. L. Dunn and M. Taya, ''Micromechanics predictions of the effective electroelastic moduli of piezoelectric composites,'' Int. J. Struct. **30**, 161–175 (1993).
[21] A. Agbossou, H. N. Viet, and J. Pastor, ''Homogenization techniques and application to piezoelectric composite materials,'' Int. J. Appl. Electromagn. Mech. **10**, 391–403 (1999).
[22] J. Pastor, ''Homogenization of linear piezoelectric media,'' Mech. Res. Commun. **24**, 145–150 (1997).
[23] P. Langlet, A. C. Hladky-Hennion, and J. N. Decarpigny, ''Analysis of the propagation of plane acoustic waves in passive periodic materials using the finite element method,'' J. Acoust. Soc. Am. **98**, 2792–2800 (1995).
[24] E. Dieulesaint and D. Royer, *Elastic Waves in Solids—Application to Signal Processing* (Ondes Élastiques Dans les Solides—Application au Traitement de Signal) (Ed Masson, Paris, 1974).
[25] P. Langlet, ''Analysis of the progagation of acoustic waves in periodic materials with the help of the finite element method (Analyse de la propagation des ondes acoustiques dans les matériaux périodiques à l'aide de la méthode des éléments finis),'' Ph.D. thesis, Université de Valenciennes et du Hainaut-Cambrésis, 1993.

Reduced models for the medium-frequency dynamics of stochastic systems

Roger Ghanem[a)]
Department of Civil Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

Abhijit Sarkar[b)]
Department of Mechanical Engineering, McGill University, 817 Sherbrooke Street, Montreal, Quebec, Canada

(Received 23 January 2002; accepted for publication 10 November 2002)

In this paper, a frequency domain vibration analysis procedure of a randomly parametered structural system is described for the medium-frequency range. In this frequency range, both traditional modal analysis and statistical energy analysis (SEA) procedures well-suited for low- and high-frequency vibration analysis respectively, lead to computational and conceptual difficulties. The uncertainty in the structural system can be attributed to various reasons such as the coupling of the primary structure with a variety of secondary systems for which conventional modeling is not practical. The methodology presented in the paper consists of coupling probabilistic reduction methods with dynamical reduction methods. In particular, the Karhunen–Loeve and Polynomial Chaos decompositions of stochastic processes are coupled with an operator decomposition scheme based on the spectrum of an energy operator adapted to the frequency band of interest. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1538246]

PACS numbers: 43.40.Cw, 43.40.Qi, 43.58.Ta [RLW]

I. INTRODUCTION

The problem of linear structural dynamics has been the subject of extensive research. Except for geometrically simple systems for which analytical solutions are attainable, the dynamical analysis of structural systems typically requires the modelization of the structure in a reduced coordinate space. The hallmark of a good reduced model is its ability, while neglecting a number of features of the behavior of the system, to capture its salient characteristics in the regime of operation that is of interest to the application at hand. A number of successful reduced models have been developed that are particularly adapted to the analysis of structural vibrations in various frequency bands.

The framework of the finite element method (FEM), for instance, in conjunction with the normal mode expansion, is a very efficient tool to construct a reduced dynamical model of the system in the low-frequency range, where the frequency response functions typically exhibit distinct modes with negligible modal overlap due to dissipation effects. The efficacy of the modal reduction procedure relies on the fact that a small number of generalized degrees of freedom contribute to the total response. The reduced model is effectively constructed by solving a well-stated generalized eigenvalue problem. The dominant eigenspace of the structure, constructed from just its first few eigenvalues and corresponding eigenvectors are necessary to synthesize the reduced model according to the normal mode approach.

Statistical Energy Analysis (SEA) is another well-established reduction method that is appropriate for the high-frequency response analysis of linear structural systems. In this approach, the system is modeled as a collection of subsystems, each of which is assigned a single response attribute, namely its total average energy. In SEA, total system response is characterized by the energy exchange among the subsystems. It is generally recognized that the validity of the approach necessitates that each subsystem contains a large number of resonant modes, manifested through the presence of uniformly high modal density. Physically, the vibration wavelength of each subsystem is of the same order of magnitude or less than the characteristic dimension of the subsystem. This approach avoids dealing with a large amount of information on individual participating modes. The sensitivity of the high-frequency vibration to minor changes in system parameters is accounted for by assuming that the natural frequencies of the subsystem can be adequately modeled by a Poisson point process in the frequency domain, with the natural frequencies as independent and identically distributed (Lyon, 1969).

In the medium frequency band, neither of the above two reduction strategies provides a suitable alternative for the dynamical analysis of structural systems. A large number of natural vibration modes are significant in propagating the motion of the structure in this frequency band. The wavelength of system deformation, in general, decreases with increasing driving frequency and the response pattern becomes increasingly complex. The traditional approaches based on FEM necessitate a prohibitively large computational model to capture the short-wavelength vibration features of the system. A large number of natural vibration modes participate in the calculation of the total response, rendering the modal analysis impractical. On the other hand, the presence of nonuniform modal density precludes the application of SEA.

[a)]Telephone: 410 516 7647. Electronic mail: ghanem@jhu.edu
[b)]Telephone: 514 398 8546. Electronic mail: asarkar@mecheng.mcgill.ca

This is due to the fact that the various subsystems show dissimilar dynamic characteristics through the presence of both short- and long-wavelength deformations.

Another difficulty arises from the fact that in the medium-frequency range, the system response becomes increasingly sensitive to geometrical details and imperfections as well as discontinuities due to the short-wavelength vibration patterns. Uncertainty regarding the nature of the coupling among the components of a complex system as well as the frequent lack of modelization of the components themselves further complicates the analysis. Thus, it becomes difficult to construct a precise mathematical model that captures the complex response behavior. Consequently, a very detailed deterministic mathematical model based on nominal system parameters may lead to unreliable response predictions in the medium-frequency band.

The effect of parameter variability is perhaps most dramatic in the spatially periodic structures such as bladed assemblies in turbomachineries, multispan structures, and aircraft fuselages. The frequency response functions of such spatially periodic structures exhibit alternate sequences of stop bands and pass bands. The pass bands are characterized by dense clusters of system natural frequencies, where vibration energy propagates freely without attenuation. For irregularities breaking the periodicity, the vibration energy is confined to a specific part of the structure due to wave reflections arising from disordered parts of the system exhibiting the well-known Anderson mode localization phenomenon (Brillouin, 1946; Weaver, 1996). This effect of disorder is much more pronounced in the higher-frequency vibration (Hodges and Woodhouse, 1983; Cha and Pierre, 1991).

A number of recent papers have reported various methodologies for the medium-frequency vibration analysis. The concept of the dynamic finite element analysis technique based on frequency-dependent shape functions is reviewed in the literature (Fergusson and Pilkey, 1993a, 1993b). The method is specifically suitable for the linear dynamic analysis in the higher-frequency range as the finite element shape functions adapt with the frequency of vibration. The feasibility of the method to formulate the dynamic stiffness matrix of randomly-parametered systems has already been presented (Manohar and Adhikari, 1998; Adhikari and Manohar, 1999). In these investigations, the random fields involving the inhomogeneous system properties are first discretized using weighted integrals. The resulting systems of linear equations are then solved using a number of procedures. These include the random eigenfunction expansion method, the Neumann expansion method, and combined analytical and simulation-based approaches. A hybrid finite element method (Vlahopoulos and Zhao, 1999) has also been developed that combines traditional FEM and SEA approaches for the vibration analysis in the medium-frequency range. The subsystems exhibiting several wavelengths of vibration are modeled by SEA approach; whereas those subsystems exhibiting few wavelengths in their dynamical response are analyzed using the FEM. A methodology to construct a reduced model for a general three-dimensional structural system in the medium-frequency range has been recently presented (Soize, 1998a, 1998b). An energy operator (symmetric positive definite) is defined for a fixed frequency band whose dominant eigensubspace allows the efficient construction of a reduced model in the frequency domain. The approach appears to be very promising for medium-frequency dynamic analysis of large-scale FE models. In part in this paper, we discuss a logical extension of the energy operator approach for dealing with randomly parametered systems.

In the present paper we begin an investigation to deal with the system parameter uncertainties using the stochastic finite element method (SFEM) in the medium-frequency range with applications to a simple system. The final objective of the investigation is to develop a general probabilistic analysis framework that is suitable for the mid-frequency vibration analysis of large-scale systems. This will be achieved in two steps. First, a frequency–domain model reduction strategy is adopted to minimize the computational effort in the mid-frequency range. In this approach, an energy operator approach introduced by Soize is used to construct the reduced model using a Ritz–Galerkin method. In parallel, another approach based on the finite element shape function that adapts with the frequency of vibration is also considered. Second, the effect of parameter uncertainties is investigated on the reduced model in the framework of the stochastic finite element method. Both approaches are essentially based on dynamic stiffness methodology, which avoids the modal superposition method and eliminates the need to determine the joint statistics of natural frequencies and mode shapes, which is a difficult and cumbersome task for general structural assembly.

In the probabilistic approach adopted herein, the uncertain system parameters are described by Gaussian random fields. The well-known Karhunen–Loeve (KL) expansion is used to discretize the random fields. This expansion decomposes a typical random field (defining an uncertain system parameter) by a set of Gaussian random variables. Consequently, a typical response quantity is expanded using a generalized random series, namely the Polynomial Chaos expansion. The deterministic matrix equations governing the unknown coefficients of the Polynomial Chaos expansion is derived and solved. The approach is exemplified through its application to the analysis of a randomly parametered axially vibrating rod. While only the stiffness of the rod is modeled as a random process, the analysis can be readily extended to account for additional dynamical parameters, such as damping, being random. This simple example helps to contrast the energy operator methodology to the frequency-dependent shape function approach in view of both model reduction strategy and system parameter uncertainty analysis.

II. FINITE ELEMENT APPROXIMATION

A finite element discretization of a linear time-invariant distributed parameter system in the frequency domain involves representing the solution with respect to a basis in a suitable functional space leading to the following approximation:

$$u_n(\mathbf{x},\omega)=\sum_{i=1}^{n} q_i(\omega)N_i(\mathbf{x}), \tag{1}$$

which involves an n-dimensional approximation subspace. Here $N_i(\mathbf{x})$ refers to the finite element shape functions and q_i is the ith nodal response quantity. In the finite-dimensional approximation to the distributed parameter system, the governing equation of motion can be rewritten as

$$\mathbf{A}_n\mathbf{q}=\mathbf{f}, \tag{2}$$

where $\mathbf{f}$ refers to external forcing, the dynamic stiffness matrix, $\mathbf{A}_n(\omega)$ is given by

$$\mathbf{A}_n=-\omega^2\mathbf{M}_n+i\omega\mathbf{D}_n+\mathbf{K}_n, \tag{3}$$

and $\mathbf{M}_n$, $\mathbf{D}_n$, and $\mathbf{K}_n$ are the mass, damping, and stiffness matrices, respectively.

III. MEDIUM-FREQUENCY STRUCTURAL DYNAMICS: DETERMINISTIC CASE

In this section, two medium-frequency vibration analysis procedures will be reviewed in the case of deterministic system parameters. Both procedures are based on the finite element approach and aim to construct reduced models for computational efficiency in the medium-frequency range. First, an energy operator approach is reviewed (Soize, 1998a). Second, a methodology based on the frequency-dependent shape functions (adaptive with the vibration wavelength) is reviewed. Only the case of deterministic system parameters are considered in this section. In the subsequent section, the applicability of the procedure in dealing the randomly parametered system will be considered.

A. Energy operator approach

In this section, a symmetric positive definite energy operator (Soize, 1998a) adapted to a fixed medium frequency band is defined. The dominant eigensubspace of the energy operator allows the construction of a reduced model using a Ritz–Galerkin method. This operator can be perceived as optimally condensing the energy of the system, in the frequency bandwidth of interest. This is, loosely speaking, an extension of the stiffness operator to the mid-frequency range, and it permits the synthesis of a solution to the problem by a mere calculation of the dominant eigenspace of the energy operator. This is the distinction with standard approaches requiring the numerically intensive task of computing higher eigenmodes of some operator. The methodology is briefly described below.

1. Definition of an energy operator

Let $\mathcal{M}$ be the mass operator and the operator-valued frequency response function $\mathcal{T}$ be defined as $\mathcal{A}^{-1}$ for a linear time-invariant distributed parameter system. The energy operator over frequency band B, $\mathcal{E}_B$, is then defined as

$$\mathcal{E}_B=\frac{1}{\pi}\int_B \omega^2\Re[\mathcal{T}^*(\omega)\mathcal{M}\mathcal{T}(\omega)]d\omega. \tag{4}$$

This energy operator is a positive-definite symmetric operator having a countable set of decreasing positive eigenvalues. Its eigenfunctions, satisfying the equation

$$\mathcal{E}_B e_i=\lambda_i e_i, \tag{5}$$

form a complete basis by which a finite element displacement field can be expanded.

2. Finite element approximation of the energy operator

In general, an explicit determination of the eigenfunctions of the energy operator is not possible. Consequently, one resorts to introduce an n-dimensional approximation of the energy operator denoted by $\mathcal{E}_{B,n}$, with its eigenfunctions denoted by e_i^n. In a FEM framework, these eigenfunctions would be approximated by their projection on the shape functions used in the discretization process,

$$e_i^n(\mathbf{x})=\sum_{j=1}^{n} P_j^i N_j(\mathbf{x}). \tag{6}$$

Thus, $\mathcal{E}_{B,n}$, the projection of $\mathcal{E}_B$ on the n-dimensional subspace spanned by the n shape functions $N_i(\mathbf{x})$, is given by (Soize, 1998a),

$$\mathcal{E}_{B,n}=\sum_{i,j=1}^{n} \mathbf{E}_{n_{ij}}(\cdot,N_j)_H N_i. \tag{7}$$

Thus, for all $g(\mathbf{x})$,

$$\mathcal{E}_{B,n}[g(\mathbf{x})]=\sum_{i,j=1}^{n} \mathbf{E}_{n_{ij}}(g,N_j)_H N_i, \tag{8}$$

with the inner product operation defined over domain Ω as

$$(f,g)_H=\int_\Omega f(\mathbf{x})g(\mathbf{x})dx, \tag{9}$$

and $\mathbf{E}_{n_{ij}}$ is the ijth element of the following matrix:

$$\mathbf{E}_n(\omega)=\frac{1}{\pi}\int_B \omega^2\Re[\mathbf{A}_n^{*-1}\mathbf{M}_n\mathbf{A}_n^{-1}]d\omega. \tag{10}$$

Thus, in a finite-dimensional representation, the standard eigenvalue problem transforms to the following generalized eigenvalue problem:

$$\mathbf{G}\mathbf{E}_n\mathbf{G}\mathbf{P}=\lambda_n\mathbf{G}\mathbf{P}, \tag{11}$$

where

$$\mathbf{G}_{ij}=(N_i,N_j)_H. \tag{12}$$

3. Representing the solution in the medium-frequency range

The energy operator, as discussed above, is a symmetric positive definite operator. Its eigenvectors, thus, are real and span the functional space in which the solution is defined. This operator, however, is such that its dominant eigenspace is adapted to frequency band B, which means that its first few eigenvectors are ideally suited for representing a solution vector in that frequency band. Thus, representing the solution with respect to the first N such eigenvectors,

$$u_n^N(\mathbf{x},\omega)=\sum_{i=1}^{N} U_i(\omega)e_i^n(\mathbf{x}). \tag{13}$$

Equation (2) can be reduced to

$$\mathbf{A}_N(\omega)\mathbf{U}=\mathbf{F}, \tag{14}$$

where the new coordinates $\mathbf{U}$ are related to the coordinates $\mathbf{q}$ through the following relation:

$$\mathbf{q}=\mathbf{P}\mathbf{U}, \tag{15}$$

where $\mathbf{P}$ is the $(n\times N)$ real matrix whose columns are the N eigenvectors corresponding to the N highest eigenvalues in Eq. (11) and the operator $\mathbf{A}_n$ is reduced to $\mathbf{A}_N$ in accordance to

$$\mathbf{A}_N(\omega)=\mathbf{P}^T\mathbf{A}_n(\omega)\mathbf{P}, \tag{16}$$

and

$$\mathbf{F}=\mathbf{P}^T\mathbf{f}. \tag{17}$$

In general, N is much smaller than the original system dimension n, which signifies the advantage of the reduced model.

B. Finite dynamic element method

In the finite dynamic element method (also referred to as the frequency-dependent shape function approach) the finite element discretization of the structure under consideration can be approximated as

$$u_n(\mathbf{x},\omega)=\sum_{i=1}^{n} q_i(\omega)N_i(\mathbf{x},\omega), \tag{18}$$

where $N_i(\mathbf{x},\omega)$ are the frequency-dependent global shape functions. These global shape functions can be written in terms of element shape functions $N_i(\mathbf{x},\omega)$ that are the solutions to the boundary value problem with homogeneous system properties with the prescribed boundary conditions on the nodes that any conventional shape function should satisfy. In case of systems with inhomogeneous properties, the analytical solutions do not exist, in general. Consequently, the finite element method is used in association with the aforementioned frequency-dependent shape functions.

In the traditional FE approach, the short-wavelength vibration associated with higher frequencies of excitation necessitates mesh refinement. The frequency-dependent shape function method avoids this requirement by adapting the shape functions to the frequency of vibration. In contrast to the traditional approach, only a few elements can adequately capture the short-wavelength vibration. Consequently, the approach offers a relief from dealing with the large computational model involving the high-frequency vibration. For the purpose of illustration, consider the case of an axially vibrating rod. The governing partial differential equation for the motion of a rod under external excitation is given by

$$\frac{\partial}{\partial x}\left(EA(x)\frac{\partial u}{\partial x}+C_1(x)\frac{\partial^2 u}{\partial t\,\partial x}\right)=m(x)\frac{\partial^2 u}{\partial t^2}+C_2(x)\frac{\partial u}{\partial t}, \tag{19}$$

with suitable initial and boundary conditions. Here EA, m, C_1, and C_2 are the stiffness, mass, strain-rate-dependent, and velocity-dependent damping per unit length of the rod. The expressions for the shape functions for a typical element are (Manohar and Adhikari, 1998),

$$\begin{Bmatrix} N_1(x,\omega) \\ N_2(x,\omega) \end{Bmatrix}=\begin{bmatrix} \cot(aL) & 1 \\ \operatorname{cosec}(aL) & 0 \end{bmatrix}\begin{Bmatrix} \sin(ax) \\ \cos(ax) \end{Bmatrix}, \tag{20}$$

where

$$a^2=\frac{\overline{m}\omega^2-i\omega\overline{C_2}}{\overline{AE}+i\omega\overline{C_1}}, \tag{21}$$

where an overbar denotes the statistical average of a system property. Consequently, the expressions for the mass and stiffness matrices are

$$[M]_{ij}=\int_0^L m(x)N_i(x,\omega)N_j(x,\omega)dx, \tag{22}$$

$$[K]_{ij}=\int_0^L EA(x)\frac{\partial N_i(x,\omega)}{\partial x}\frac{\partial N_j(x,\omega)}{\partial x}dx. \tag{23}$$

This is in contrast to the traditional finite element approach, where the aforementioned quantities are frequency independent. The higher-order frequency-dependent terms in the system matrix include the extra correction terms, thus providing more accuracy in the dynamic response estimate at the higher frequency of vibrations. Only a few degrees of freedom adequate to model the entire system provide a better convergence rate compared to traditional FEM (Fergusson and Pilkey, 1993a, 1993b).

However, the solutions for the general three-dimensional structural systems, even with uniform system properties does not exist in the majority of cases. Consequently, the approach has a limited application in modeling a large-scale built-up structural assembly.

IV. REVIEW OF PROBABILISTIC CONCEPTS

In this section, a brief review of the probabilistic approach is presented. This approach will be used subsequently for the system parameter uncertainty analysis. The development presented in this paper hinges on the definition of random variables as functions from the space of elementary events to the real line (Ghanem and Spanos, 1991; Ghanem and Red-Horse, 1999a; Ghanem, 1999b). As functions, approximation theory, as developed for deterministic functions, can be applied to random variables. The main question to be addressed concerns methods to characterize the model-based predictions, where some parameters of the model have been represented as stochastic processes. The answer to this question lies in the realization that in the deterministic finite element method, as well as most other numerical analysis techniques, a solution to a deterministic problem is known once its projection on a basis in an appropriate function space has been evaluated. It often happens, in deterministic analysis, that the coefficients in such a representation have an immediate physical meaning, which distracts from the mathematical significance of the solution. Carrying this argument over to the case involving stochastic processes, the solution to the problem is identified with its projection on a set of appropriately chosen basis functions. A random variable is thus viewed as a function of a single variable, θ, that refers to the space of elementary events. Monte Carlo simulation can be viewed as a collocation along this θ dimension. Other ap-

proximations along this dimension are possible, and can be associated with different choices of basis functions in the corresponding space of random variables. This theoretical development is consistent with the identification of the space of second-order random variables as a Hilbert space with the inner product on it defined as the operation of statistical correlation (Loeve, 1977). Second-order random variables are those random variables with finite variance, they are mathematically similar to deterministic functions with finite energy.

A. Mathematical characterization of random processes

A mathematical framework suitable for describing the problem at hand can be presented as follows: The data, modeled as random variables or processes, lives in the Hilbert space $\mathcal{H}_{\mathcal{G}}$. Assuming the data to be well defined in a probabilistic sense provides a full characterization of this space, in which a set of basis functions, ξ, must be identified. This is accomplished in the section below on Karhunen–Loeve expansions. The state of the system, again modeled as a random variable or process, resides in the Hilbert space $\mathcal{H}_{\mathcal{L}}$. A set of basis function, ψ, is also identified in this space, which, in general, is different from the basis ξ since this latter one spans only a subset of the space of second-order random variables, namely those that characterize the data. Identifying a basis for the space $\mathcal{H}_{\mathcal{L}}$ is accomplished in the section below dealing with the Polynomial Chaos expansion.

1. Karhunen-Loeve expansion

The Karhunen–Loeve expansion (Loeve, 1977) of a stochastic process $\alpha(\mathbf{x},\theta)$, is based on the spectral expansion of its covariance function $R_{\alpha\alpha}(\mathbf{x},\mathbf{y})$. Here, $\mathbf{x}$ and $\mathbf{y}$ are used to denote spatial coordinates. The covariance function being symmetrical and positive definite, by definition, has all its eigenfunctions mutually orthogonal, and they form a complete set spanning the function space to which $\alpha(\mathbf{x},\theta)$ be longs. It can be shown that if this deterministic set is used to represent the process $\alpha(\mathbf{x},\theta)$; then the random coefficients used in the expansion are also orthogonal (i.e., uncorrelated). The expansion then takes the following form:

$$\alpha(\mathbf{x},\theta)=\overline{\alpha}(\mathbf{x})+\sum_{i=1}^{\infty}\sqrt{\lambda_i}\xi_i(\theta)\phi_i(\mathbf{x}), \tag{24}$$

where $\overline{\alpha}(\mathbf{x})$ denotes the mean of the stochastic process, and $\{\xi_i(\theta)\}$ forms a set of orthogonal random variables. Furthermore, $\{\phi_i(\mathbf{x})\}$ are the eigenfunctions and $\{\lambda_i\}$ are the eigenvalues, of the covariance kernel, and can be evaluated as the solution to the following integral equation:

$$\int_{\mathcal{D}}R_{\alpha\alpha}(\mathbf{x},\mathbf{y})\phi_i(\mathbf{y})d\mathbf{y}=\lambda_i\phi_i(\mathbf{x}), \tag{25}$$

where $\mathcal{D}$ denotes the spatial domain over which the process $\alpha(\mathbf{x},\theta)$ is defined. The most important aspect of this spectral representation is that the spatial random fluctuations have been decomposed into a set of deterministic shapes multiplying random amplitudes. If the random process being expanded, $\alpha(\mathbf{x},\theta)$, is Gaussian, then the random variables $\{\xi_i\}$ form an orthonormal Gaussian vector. The Karhunen–Loeve expansion is mean-square convergent irrespective of the probabilistic structure of the process being expanded, provided it has a finite variance (Loeve, 1977). The closer a process is to white noise, the more terms that are required in its expansion, while at the other limit, a random variable can be represented by a single term. In physical systems, it can be expected that material properties vary smoothly at the scales of interest in most applications, and therefore only a few terms in the Karhunen–Loeve expansion can capture most of the uncertainty in the process.

2. Polynomial chaos expansion

The covariance function of the solution process is not known *a priori*, and hence the Karhunen–Loeve expansion cannot be used to represent it. Since the solution process is a function of the material properties, nodal solution variables, $S(\theta)$, can be formally expressed as some nonlinear functional of the set $\{\xi_i(\theta)\}$ used to represent the material stochasticity. It has been shown (Cameron and Martin, 1947) that this functional dependence can be expanded in terms of polynomials in Gaussian random variables, referred to as Polynomial Chaos. Namely,

$$u(\theta)=a_0\Gamma_0+\sum_{i_1=1}^{\infty}a_{i_1}\Gamma_1(\xi_i(\theta))+\sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}a_{i_1i_2}\Gamma_2(\xi_{i_1}(\theta),\xi_{i_2}(\theta))+\cdots. \tag{26}$$

In this equation, the symbol $\Gamma_n(\xi_{i_1},\ldots,\xi_{i_n})$ denotes the Polynomial Chaos (Wiener, 1938; Kallianpur, 1980) of order n in the variables $(\xi_{i_1},\ldots,\xi_{i_n})$. These are generalizations of the multidimensional Hermite polynomials to the case where the independent variables are themselves measurable functions (in this case they are random variables). Introducing a one-to-one mapping to a set with ordered indices denoted by $\{\Psi_i(\theta)\}$ and truncating the Polynomial Chaos expansion after the Pth term, Eq. (26) can be rewritten as

TABLE I. Polynomial chaoses and their variances; two terms in the Karhunen–Loeve expansion.

ith polynomial chaos	Order of the homogeneous chaos	Ψ_i	$\langle\Psi_i^2\rangle$
0	0	1	1
1	1	ξ_1	1
2		ξ_2	1
3	2	ξ_1^2-1	2
4		$\xi_1\xi_2$	1
5		ξ_2^2-1	2
6	3	$\xi_1^3-3\xi_1$	6
7		$\xi_1^2\xi_2-\xi_2$	2
8		$\xi_1\xi_2^2-\xi_2$	2
9		$\xi_2^3-3\xi_2$	6
10	4	$\xi_1^4-6\xi_1^2+3$	24
11		$\xi_1^3\xi_2-3\xi_1\xi_2$	6
12		$\xi_1^2\xi_2^2+\xi_1^2-\xi_2^2+1$	4
13		$\xi_1\xi_2^3-3\xi_1\xi_2$	6
14		$\xi_2^4-6\xi_2^2+3$	24

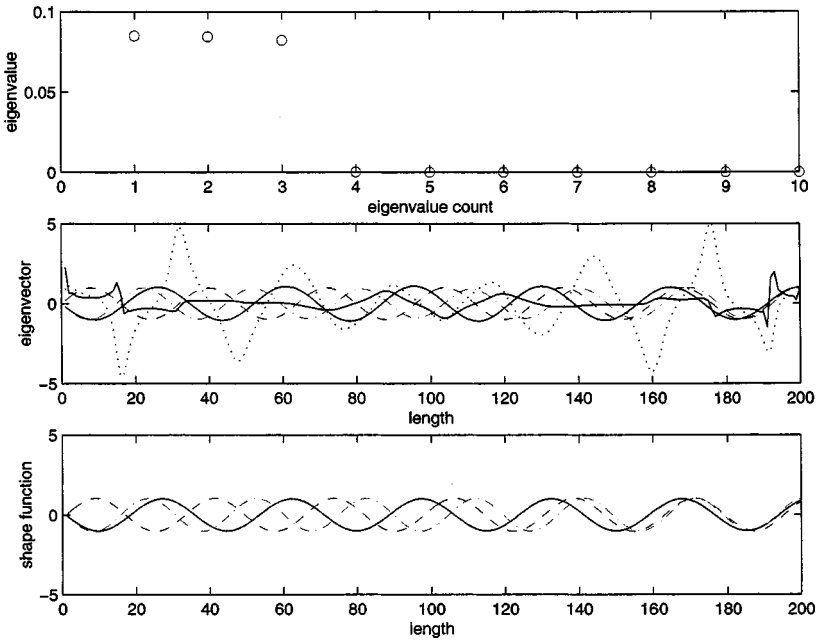


FIG. 1. Top: The distributions of eigenvalues of the energy operator in the frequency band 16–20 rad/s for $C_1=5.0$ N s and $C_2=5.0$ N s/m^2; middle: eigenvectors of the energy operator; bottom: frequency-dependent shape functions.

$$u(\theta)=\sum_{j=0}^{P} u_j \Psi_j(\theta). \tag{27}$$

Table I shows the explicit expressions for the first few of these polynomials. These polynomials are orthogonal in the sense that their inner product $\langle \Psi_j \Psi_k \rangle$, which is defined as the statistical average of their product, is equal to zero for $j \neq k$. Moreover, they can be shown to form a complete basis in the space of second-order random variables. A complete probabilistic characterization of the process $u(\theta)$ is obtained once the deterministic coefficients u_j have been calculated. A given truncated series can be refined along the random dimension either by adding more random variables to the set $\{\xi_i\}$ or by increasing the maximum order of polynomials included in the Polynomial Chaos expansion. The first refinement takes into account higher-frequency random fluctuations of the underlying stochastic process, while the second refinement captures a strong nonlinear dependence of the solution process on this underlying process.

It should be noted at this point that the Polynomial Chaos expansion can be used to represent, in addition to the solution process, stochastic processes that model non-Gaussian material properties. The processes representing the material properties are thus expressed as the output of a nonlinear system to a Gaussian input.

In the next section, the Karhunen–Loeve and the Polynomial Chaos expansions are integrated into the equations of motions derived in the previous section, and procedures are developed for evaluating an expansion of the solution process with respect to the Polynomial Chaos basis.

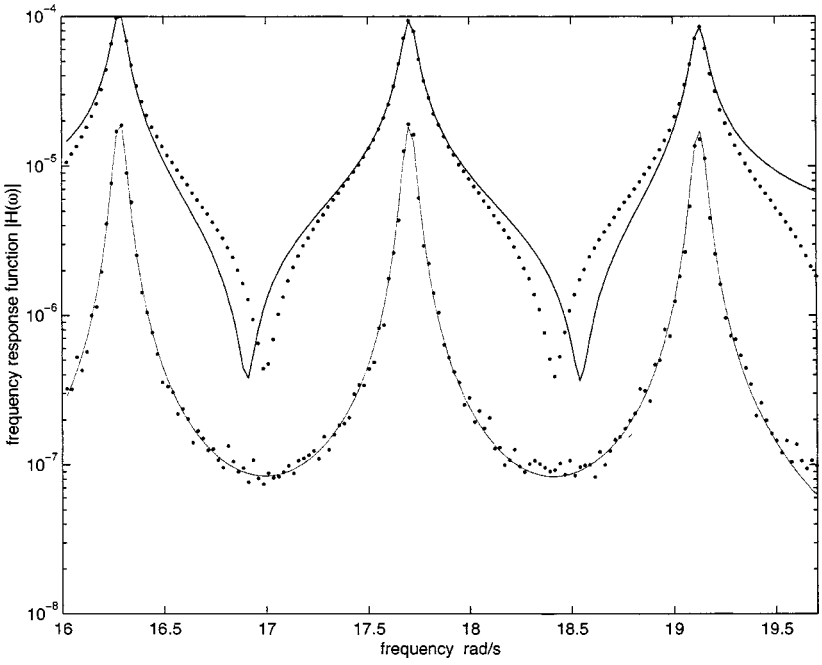


FIG. 2. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the energy operator approach; line plots refer to the Polynomial Chaos approximation with an energy operator; point plots refer to Monte Carlo simulation results on full model; frequency band 16–20 rad/s, $C_1=5.0$ N s, $C_2=5.0$ N s/m^2.

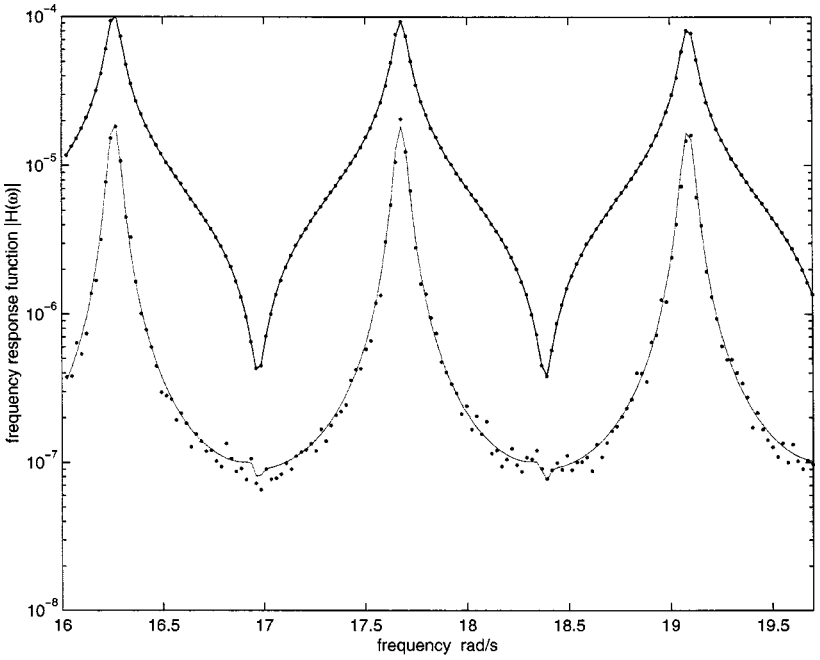


FIG. 3. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the frequency-dependent shape function approach; line plots refer to the Polynomial Chaos approximation with frequency-dependent shape functions; point plots refer to Monte Carlo simulation results on a full model; Frequency band 16–20 rad/s, $C_1=5.0$ N s, $C_2=5.0$ N s/m^2.

V. MEDIUM-FREQUENCY STRUCTURAL DYNAMICS: STOCHASTIC CASE

In a previous section, the mid-frequency dynamic analysis procedures have been described in case of deterministic system parameters. In this section, the modification of the procedures to analyze the randomly parametered systems will be considered.

A. Energy operator approach

In Sec. III, the system has been assumed to be deterministic, resulting in a deterministic energy operator and its associated eigenproblem. In case of a randomly parametered system, the eigensolution computed above can be construed as a first-order approximation to the mean eigenproblem of the energy operator. Given that this operator undergoes an averaging process over frequency band B, it is expected to exhibit milder fluctuations than the underlying material properties, hence justifying this level of approximation. Furthermore, the random fluctuations are expected to influence the higher-order eigenvalues and eigenvectors that are of little importance for the purpose of the present analysis. Once these eigenvectors have been computed, however, the coordinates of the solution with respect to this basis are treated as random quantities. Thus expanding $\mathbf{A}_N$ in its Karhunen–Loeve expansion (synthesized from that of the stiffness, mass, and damping matrices) results in the following expansions:

$$\left[\sum_{k=0}^{nkl} \xi_k \mathbf{A}_N^k(\omega)\right]\hat{\mathbf{U}}=\mathbf{F}, \tag{28}$$

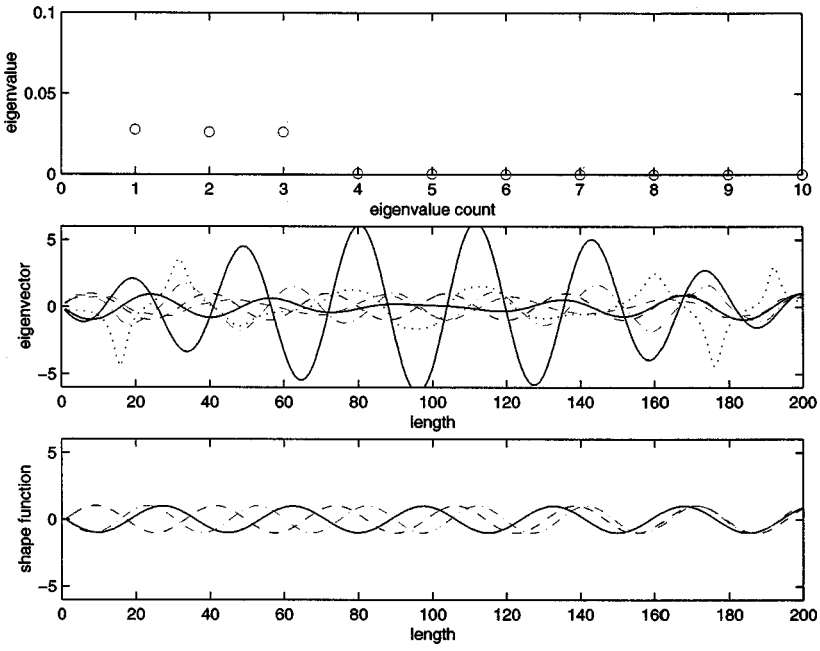


FIG. 4. Top: The distributions of eigenvalues of the energy operator in the frequency band 16–20 rad/s for $C_1=30.0$ N s and $C_2=30.0$ N s/m^2; middle: eigenvectors of the energy operator; bottom: frequency-dependent shape functions.

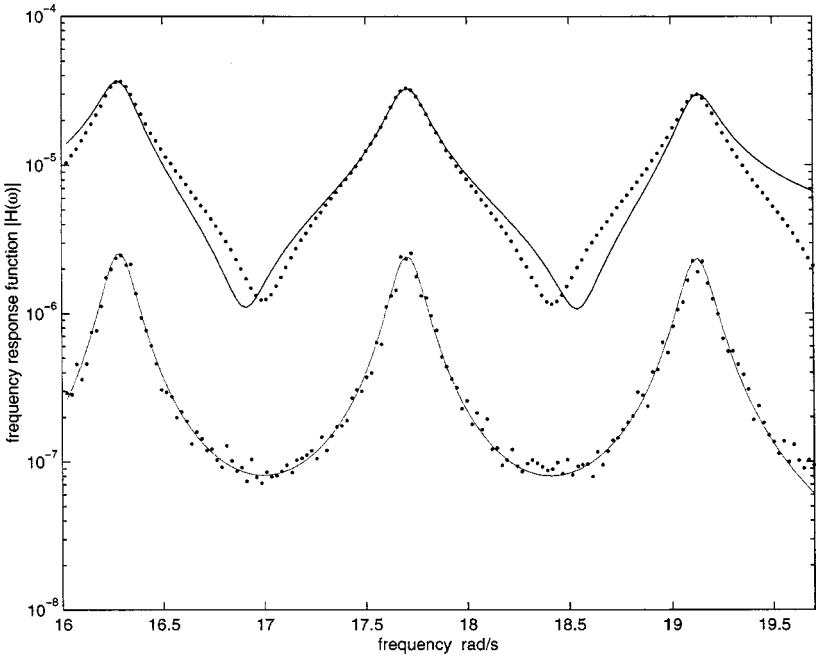


FIG. 5. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the energy operator approach; line plots refer to the Polynomial Chaos approximation with an energy operator; point plots refer to Monte Carlo simulation results on a full model; Frequency band 16–20 rad/s, $C_1=30.0$ N s, $C_2=30.0$ N s/m^2.

where

$$\mathbf{A}_N^k=\mathbf{P}^T\mathbf{A}_n^k\mathbf{P}, \quad k=0,...,nkl \tag{29}$$

Consequently, we expand the solution process in terms of Polynomial Chaos expansion as

$$\hat{\mathbf{U}}=\sum_{j=0}^{npc}\Psi_j(\theta)\hat{\mathbf{U}}_j, \tag{30}$$

where ξ_i are the random variables obtained from the Karhunen–Loeve expansion, and are Gaussian whenever the material properties are modeled as Gaussian processes, and Ψ_i are polynomials that are orthonormal with respect to the probability density function of the ξ. Here *nkl* and *npc* represent the number of terms retained in the Karhunen–Loeve and Polynomial Chaos expansions. Substituting the above expansions into the reduced model equation and carrying on a Galerkin projection, results in

$$\sum_{i=0}^{npc}\sum_{j=0}^{nkl}\langle\xi_i\Psi_j\Psi_k\rangle\mathbf{A}_N^i\hat{\mathbf{U}}_j=\langle\Psi_k\mathbf{F}\rangle, \tag{31}$$

where the deterministic coefficients $\hat{\mathbf{U}}_j$ are computed as the solution of the above deterministic algebraic problem.

B. Dynamic element approach

Similarly to their integration into the energy operator formalism, the Karhunen–Loeve and Polynomial Chaos expansions can be integrated into the frequency-dependent shape functions. Under the assumption of a smooth variation of material properties, the foregoing approximation is ex-

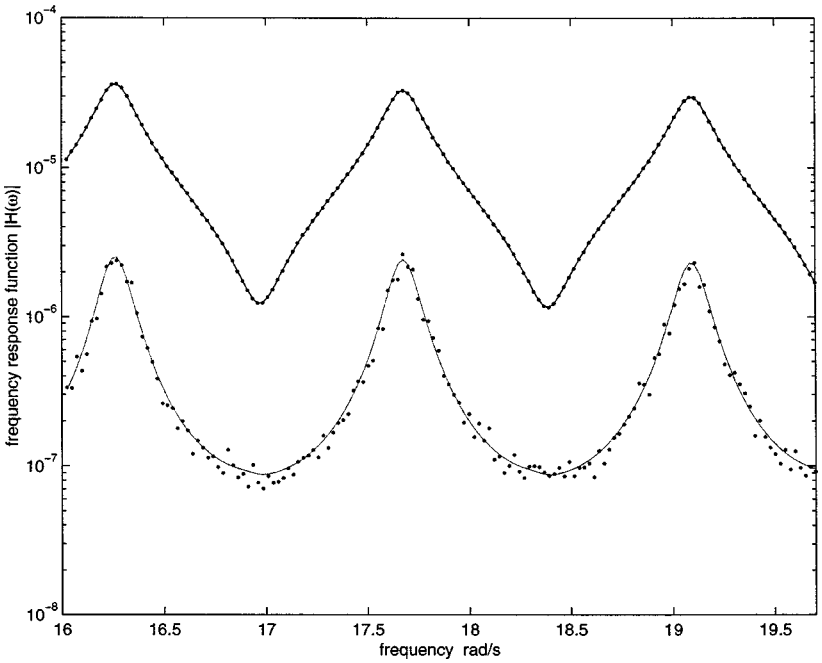


FIG. 6. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the frequency-dependent shape function approach; line plots refer to the Polynomial Chaos approximation with frequency-dependent shape functions; point plots refer to Monte Carlo simulation results on a full model; frequency band 16–20 rad/s, $C_1=30.0$ N s, $C_2=30.0$ N s/m^2.

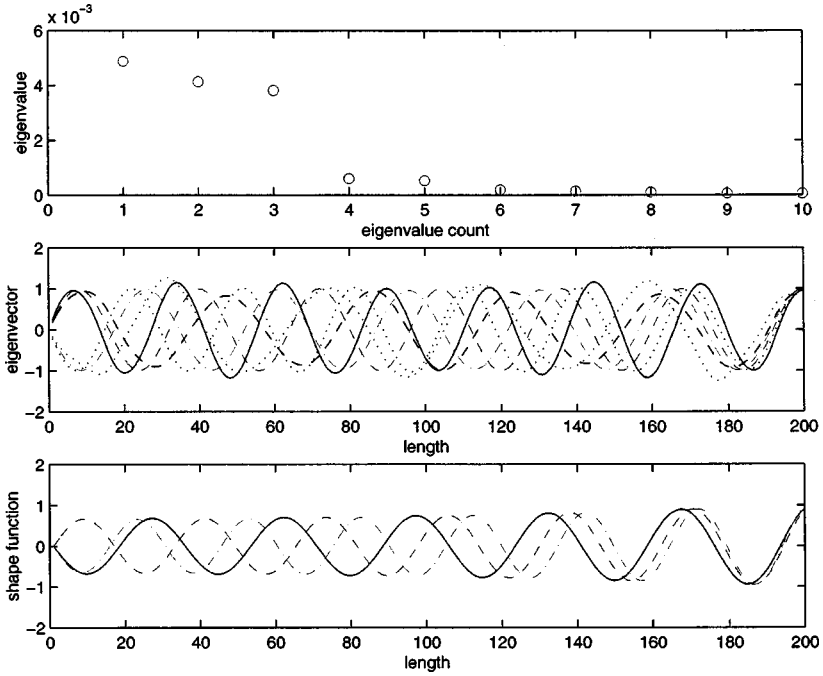


FIG. 7. Top: The distributions of eigenvalues of the energy operator in the frequency band 16–20 rad/s for $C_1=150.0$ N s and $C_2=150.0$ N s/m^2; middle: eigenvectors of the energy operator; bottom: frequency-dependent shape functions.

pected to be reasonable. This assumption is required in order to ensure that the analytical expressions for the shape functions, obtained under the assumption of homogeneous coefficients, is still approximately valid. Once the stochastic material properties are expanded using Karhunen–Loeve expansion, the steady-state dynamic equation of equilibrium in frequency domain can be expressed as Eq. (28). Consequently, a typical response quantity is expanded using Polynomial Chaos expansion. Using a Galerkin projection, the deterministic matrix equations involving the coefficients of the Polynomial Chaos basis are obtained similar to Eq. (31).

VI. NUMERICAL RESULTS

The aforementioned analysis is illustrated by its application to a simple system, namely an axially vibrating rod. The dynamic motion of the rod due to an external excitation is described by Eq. (19). The choice of this simple system provides insight into the problem as it permits the comparison of the energy operator approach and the frequency-dependent finite element shape function methodology. It is worthwhile noting that the natural frequencies of an axially vibrating rod with homogeneous properties are almost equally spaced on the frequency axis. Thus, the manifestation of relatively higher modal density in the medium-frequency range is not typified through this system. As our main purpose in the paper is to gain insight in the foregoing approaches, the authors believe that this simple system is appropriate and adequate for the study without introducing undue mathematical complexity.

The analysis is carried out for a fixed-free axial bar,

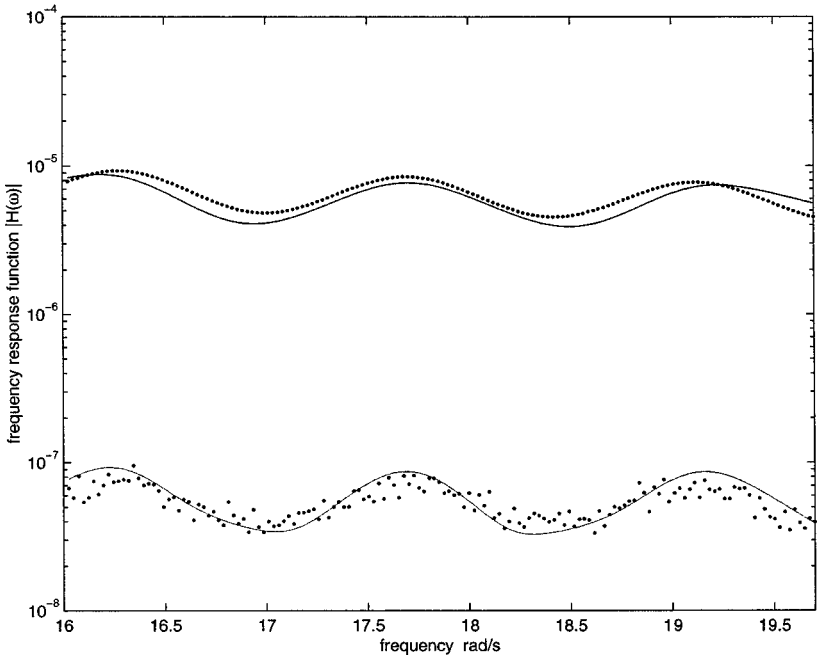


FIG. 8. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free-end of the bar using the energy operator approach; line plots refer to the Polynomial Chaos approximation with energy operator; point plots refer to Monte Carlo simulation results on full model; frequency band 16–20 rad/s, $C_1=150.0$ N s, $C_2=150.0$ N s/m^2.

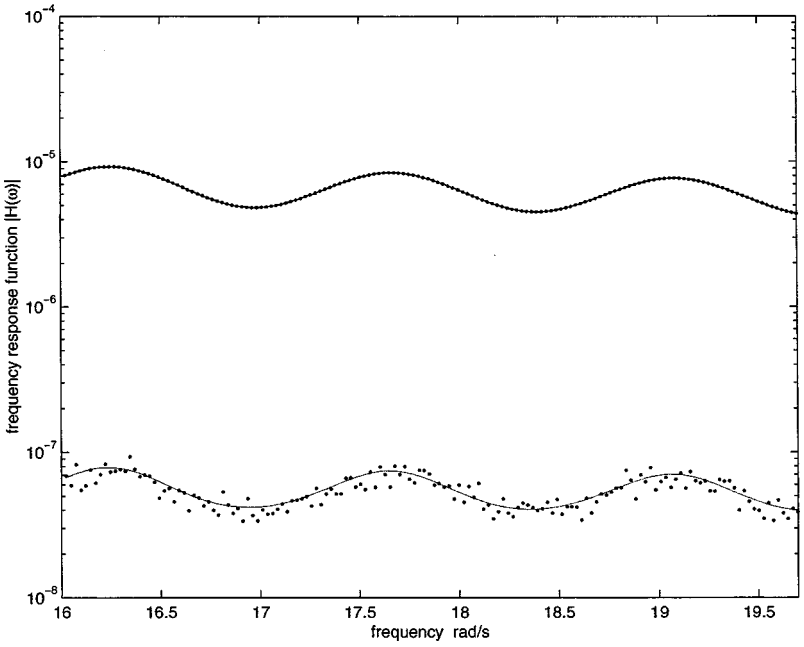


FIG. 9. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the frequency-dependent shape function approach; line plots refer to the Polynomial Chaos approximation with frequency-dependent shape functions; point plots refer to Monte Carlo simulation results on a full model; frequency band 16–20 rad/s, $C_1 = 150.0$ N s, $C_2 = 150.0$ N s/m^2.

excited at the free end, with Young's modulus modeled as a Gaussian stochastic process. Its correlation length is assumed to be half the length of the bar, with a coefficient of variation equal to 0.05. It should be noted that for dynamic problems, the significance of the coefficient of variation should be judged relative to the frequency band and structural damping in which the analysis is being conducted. Thus, small variations in material properties, while unlikely to cause significant perturbations to the lower-order mode shapes and natural frequencies, could significantly modify the higher-order modal properties.

For the numerical investigation, the following numerical values are assumed of the physical parameters of the axial bar: $\overline{EA} = 405\,284.0$ N; $m = 200.0$ kg/m, $L = 100.0$ m; the number of finite elements for the energy operator approach $= 200$; the number of finite elements for a frequency-dependent shape function approach $= 1$. Figure 1(a) shows the distributions of eigenvalues of the energy operator in the frequency band 16–20 rad/s for $C_1 = 5.0$ N s and $C_2 = 5.0$ N s/m^2. The first few eigenvectors of the energy operator are plotted in Fig. 1(b). Figure 1(c) shows a frequency-dependent shape function of the rod computed at the natural frequencies of the rod. Note that the first three eigenvectors of the energy operator plotted in Fig. 1(b) are very similar to the shape functions plotted in Fig. 1(c) at the natural frequencies (approximately at 16.25, 17.75, and 19.25 rad/s) of the rod. In Fig. 1(a), the sharp decrement of the eigenvalues points toward the fact that an efficient reduced model can be constructed using the corresponding eigenvectors as the basis for FEM within the specified frequency band of interest. In this specific case, only eigenvectors corresponding to the first

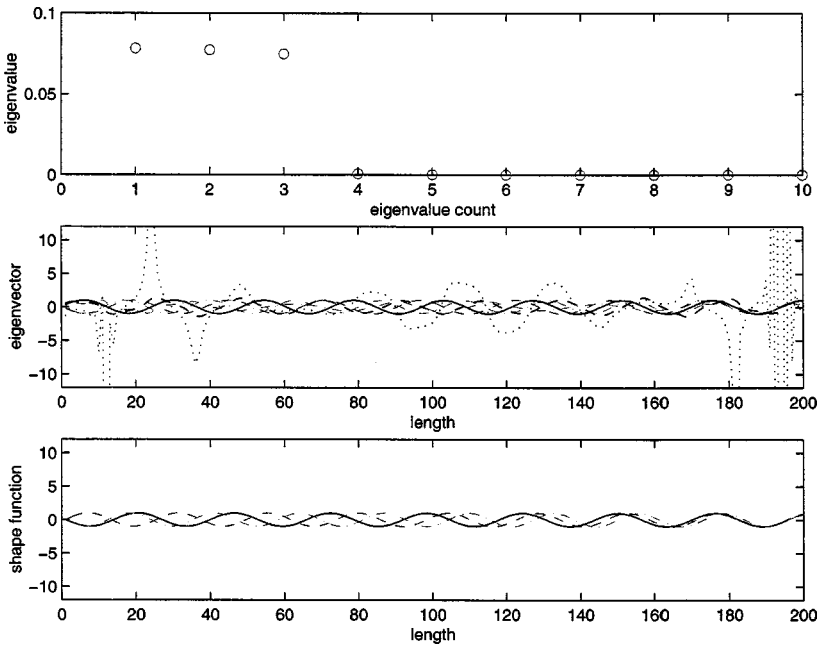


FIG. 10. Top: The distributions of eigenvalues of the energy operator in the frequency band 21.5–25.5 rad/s for $C_1 = 5.0$ N s and $C_2 = 5.0$ N s/m^2; middle: eigenvectors of the energy operator; bottom: frequency-dependent shape functions.

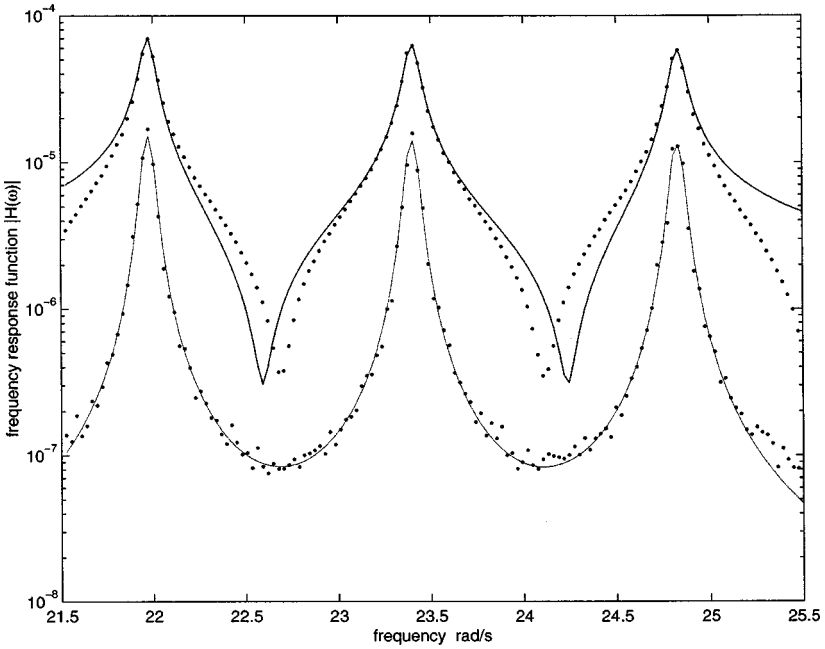


FIG. 11. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the energy operator approach; line plots refer to the Polynomial Chaos approximation with an energy operator; point plots refer to Monte Carlo simulation results on a full model; frequency band 21.5–25.5 rad/s, $C_1=5.0$ N s, $C_2=5.0$ N s/m^2.

three eigenvalues are found to be adequate in the construction of the reduced model. Thus, the reduced model having only three degrees of freedom can effectively reproduce the dynamics of the original system having 200 degrees of freedom. Consequently, the uncertainty analysis is performed on the three degrees-of-freedom system. The autocovariance function of the Young's modulus is chosen to be of the following form: $R(x,y)=\exp-|x-y|/b$, where b is a correlation length, assumed to be half the length of the rod. Only two terms in the Karhunen–Loeve expansion are retained. In the Polynomial Chaos expansion, the first 15 terms corresponding to a fourth-order expansion are used. It is noted here that this functional form for the correlation function results in a stochastic process whose realizations are nondifferentiable. This is the so-called ''Gauss–Markov'' process. Truncating the Karhunen–Loeve representation of the process at the second term, however, results in an approximation that is differentiable. Figure 2 shows the mean (top curve) and standard deviation (bottom curve) of the transfer function at the free-end of the system using the energy operator approach. The results of Monte Carlo simulations (100 realizations) performed on the full-scale system (with 200 degrees of freedom) is represented by the point plots. The result obtained using Polynomial Chaos expansion on the reduced model are in excellent agreement with that of Monte Carlo simulations. It is clear that the level of accuracy of the present analysis degrades around the cutoff frequencies of the frequency band of interest, while the approximation is quite good inside that band, using only three terms in the eigenexpansion of the solution from the energy operator. The agreement is better around the resonant peaks compared to the antiresonance troughs.

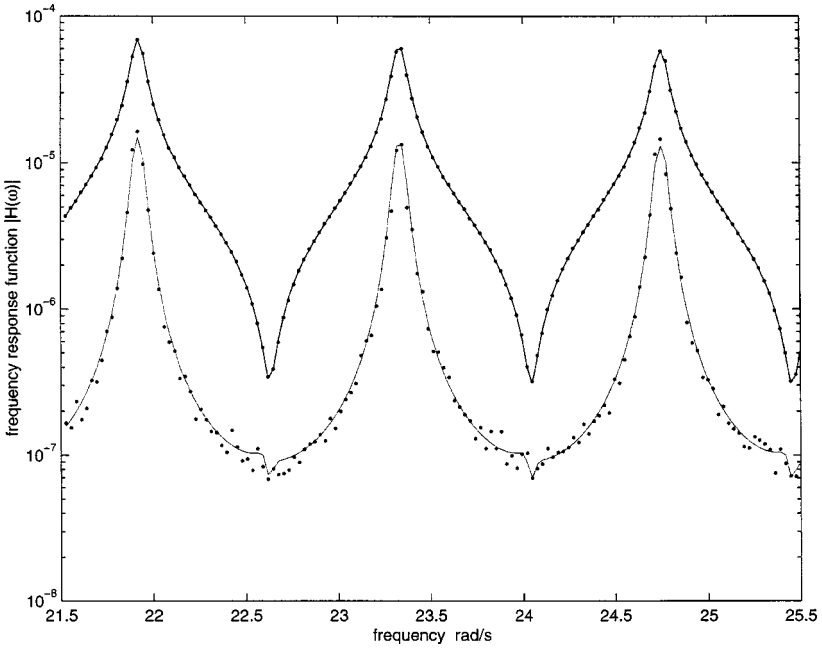


FIG. 12. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the frequency-dependent shape function approach; line plots refer to the Polynomial Chaos approximation with frequency-dependent shape functions; point plots refer to Monte Carlo simulation results on a full model; frequency band 21.5–25.5 rad/s, $C_1=5.0$ N s, $C_2=5.0$ N s/m^2.

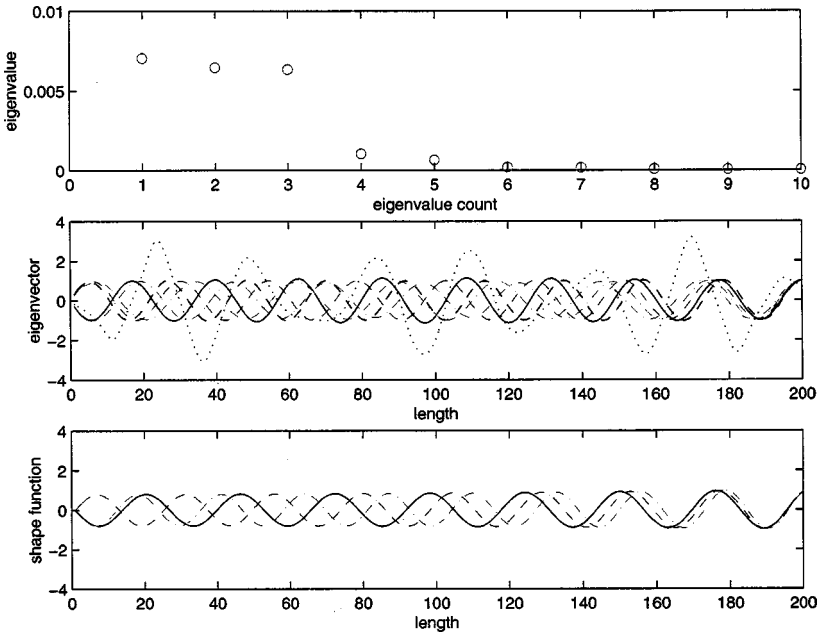


FIG. 13. Top: The distributions of eigenvalues of the energy operator in the frequency band 16–20 rad/s for $C_1 = 150.0$ N s and $C_2 = 150.0$ N s/m^2; middle: eigenvectors of the energy operator; bottom: frequency-dependent shape functions.

It is worthwhile mentioning at this point that the frequency response curve for a fixed value of frequency is a random variable when the system parameters are modeled as random fields. Furthermore, it becomes a complex-valued random process in its frequency evolution due to the presence of damping. The presence of discrete natural frequencies induces strong nonstationary characteristics of the frequency response functions with the standard deviation shooting up significantly near the resonance points, even with small random deviations of system parameters. This may potentially affect the extremes of the frequency response curves and thus, the dynamic response.

The results obtained using the frequency-dependent shape function approach are shown in Fig. 3. In this case, the results obtained using Monte Carlo simulation and a frequency-dependent shape function approach are in perfect agreement. Similar results are presented in Figs. 4–6 for $C_1 = 30.0$ N s and $C_2 = 30.0$ N s/m^2 and Figs. 7–9 for $C_1 = 150.0$ N s and $C_2 = 150.0$ N s/m^2. It is clear from these figures that as damping increases, the size of the dominant eigenspace increases accordingly resulting in a relatively larger size of the reduced model.

In Figs. 10–12, the results corresponding to the frequency band 21.5–25.5 rad/s are presented for the following case: $C_1 = 5.0$ N s and $C_2 = 5.0$ N s/m^2. Clearly, the dominant eigenvectors of the energy operators [see Fig. 10(b)] and the frequency-dependent shape functions [see Fig. 10(c)] represent relatively higher fluctuations adapting with higher-frequency vibration of the system. Similar results are shown in Figs. 13–15 for $C_1 = 150.0$ N s and $C_2 = 150.0$ N s/m^2.

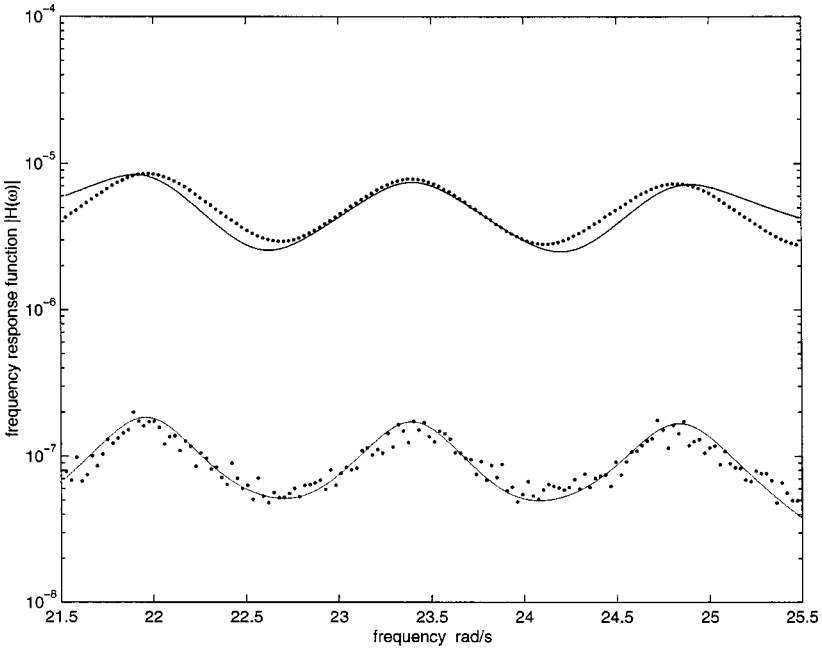


FIG. 14. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the energy operator approach; line plots refer to the Polynomial Chaos approximation with energy operator; point plots refer to Monte Carlo simulation results on full model; frequency band 21.5–25.5 rad/s, $C_1 = 150.0$ N s, $C_2 = 150.0$ N s/m^2.

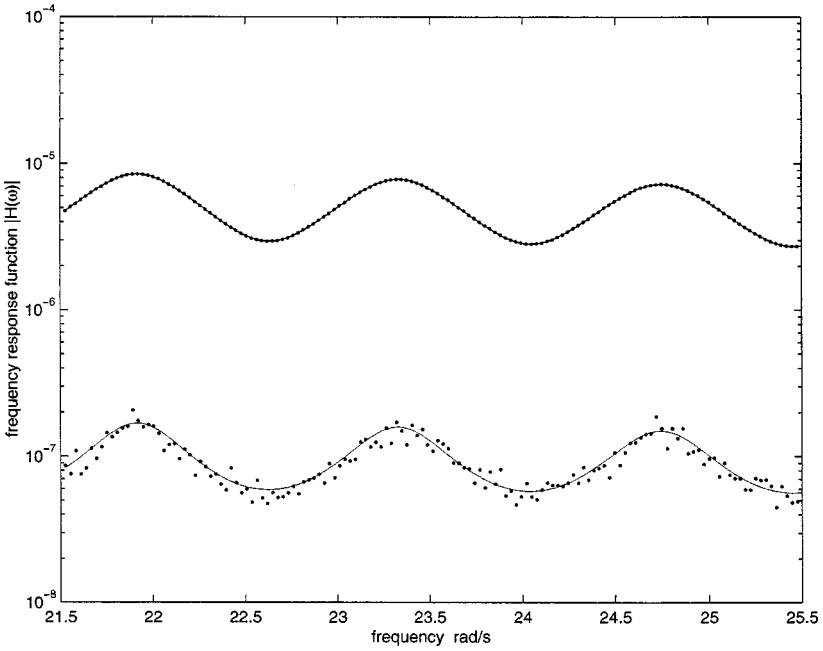


FIG. 15. The mean (top curve) and standard deviation (bottom curve) of the transfer function at the free end of the bar using the frequency-dependent shape function approach; line plots refer to the Polynomial Chaos approximation with frequency-dependent shape functions; point plots refer to Monte Carlo simulation results on a full model; frequency band 21.5–25.5 rad/s, $C_1 = 150.0$ N s, $C_2 = 150.0$ N s/m^2.

VII. CONCLUDING REMARKS

In this paper we investigate the influence of parameter uncertainties on the mid-frequency vibration of linear systems. The mid-frequency vibration analysis procedure developed for the analysis of deterministic systems has been extended to consider the case of systems with stochastically inhomogeneous properties. The paper contrasts an energy operator-based methodology with an approach based on frequency-dependent finite element shape functions.

The inhomogeneities of the system parameters modeled as Gaussian stochastic processes are represented using their Karhunen–Loeve expansion. Consequently, a typical response quantity is expanded using its Polynomial Chaos decomposition. The aforementioned mid-frequency vibration analysis procedures, based on the energy operator approach and the frequency-dependent shape function methodology, elegantly integrate SFEM using the Karhunen–Loeve and Polynomial Chaos formalisms. As reported in a number of other publications, these expansions are not limited to Gaussian processes or small levels of uncertainty (Ghanem and Spanos, 1991; Ghanem and Dham, 1998; Sakamoto and Ghanem, 2002).

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of the Office of Naval Research under Grant No. N000149910900.

Adhikari, S., and Manohar, C. S. (**1999**). ''Dynamic analysis of framed structures with statistical uncertainties,'' Int. J. Numer. Methods Eng. **44**, 1157–1178.

Brillouin L. (**1946**). *Wave Propagation in Periodic Structures* (McGraw Hill, New York).

Cameron, R. H., and Martin, W. T. (**1947**). ''The orthogonal development of nonlinear functionals in series of Fourier–Hermite functionals,'' Ann. Math. **48**, 385–392.

Cha, P. D., and Pierre, C. (**1991**). ''Vibration localization by disorder in assemblies of monocoupled, multimodal componenent systems,'' J. Appl. Mech. **58**, 1072–1081.

Fergusson, N. J., and Pilkey, W. D. (**1993a**). ''Literature review of variants of dynamic stiffness method, Part 1: The dynamic element method,'' Shock Vib. Dig. **25**, 3–12.

Fergusson, N. J., and Pilkey, W. D. (**1993b**). ''Literature review of variants of dynamic stiffness method, Part 2: Frequency dependent matrix and other corrective methods,'' Shock Vib. Dig. **25**, 3–10.

Ghanem, R., and Dham, S. (**1998**). ''Stochastic finite element analysis for multiphase flow in heterogeneous porous media,'' Flows Porous Media **32**, 239–262.

Ghanem, R., and Red-Horse, J. (**1999a**). ''Propagation of uncertainty in complex physical systems using a stochastic finite element approach,'' Physica D **133**, 137–144.

Ghanem, R. (**1999b**). ''Ingredients for a general purpose stochastic finite elements formulation,'' Comput. Methods Appl. Mech. Eng. **168**, 19–34.

Ghanem, R., and Spanos, P. (**1991**). *Stochastic Finite Elements: A Spectral Approach* (Springer-Verlag, New York).

Hodges, C., and Woodhouse, J. (**1983**). ''Vibration isolation from irregularity in a nearly periodic structure: theory and measurements,'' J. Acoust. Soc. Am. **74**, 894–905.

Kallianpur, G. (**1980**). *Stochastic Filtering Theory* (Springer-Verlag, Berlin).

Loeve, M. (**1977**). *Probability Theory*, 4th ed. (Springer-Verlag, New York).

Lyon, R. H. (**1969**). ''Statistical analysis of power injection and response in structures and Rooms,'' J. Acoust. Soc. Am. **45**, 545–565.

Manohar, C. S., and Adhikari, S. (**1998**). ''Dynamic stiffness of randomly parameteredbeams,'' Probab. Eng. Mech. **13**, 39–49.

Sakamoto, S., and Ghanem, R. (**2002**). ''Polynomial chaos decomposition for the simulation of non-Gaussian non-stationary stochastic processes,'' ASCE J. Eng. Mech. **128**, 190–201.

Soize, C. (**1998a**). ''Reduced models in the medium frequency range for general dissipative structural-dynamics systems,'' Eur. J. Mech. A/Solids **17**, 657–685.

Soize, C. (**1998b**). ''Reduced models in the medium-frequency range for general external structural–acoustic systems,'' J. Acoust. Soc. Am. **103**, 3393–3406.

Vlahopoulos, N., and Zhao, X. (**1999**). ''Basic development of hybrid finite element method for midfrequency structural vibrations,'' AIAA J. **37**, 1495–1505.

Weaver, R. (**1996**). ''Localization, scaling, and diffusion transport of wave energy in disordered media,'' Appl. Mech. Rev. **49**, 126–135.

Wiener, N. (**1938**). ''The homogeneous chaos,'' Am. J. Math. **60**, 897–936.

Radial vibrations of orthotropic laminated hollow spheres

Yehuda Stavsky and J. Barry Greenberg[a)]
Faculty of Aerospace Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel

(Received 13 June 2002; revised 26 October 2002; accepted 12 November 2002)

The three-dimensional elasticity problem of the radial vibrations of a composite hollow spherical shell laminated of spherically orthotropic layers is considered. After formulating the equations, the exact determinantal equation from which the frequencies of vibration can be extracted is developed. Some calculated results for combinations of isotropic and orthotropic materials indicate the sensitivity of the frequencies to the geometry and material make up of the shells. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1536625]

PACS numbers: 43.40.Ey [ANN]

I. INTRODUCTION

The radial vibrations of an isotropic *full* sphere were first investigated by Poisson[1] in 1828 and numerical results were obtained by Lamb[2] in 1882. The six lowest roots of the frequency equation are given in Love's treatise[3] on the mathematical theory of elasticity for an isotropic sphere. He also showed the results for the particular case of a very thin *hollow* spherical shell. More recently, Schafbuch *et al.*[4] used Debye's[5] solution to compute the eigenfrequencies of an isotropic elastic sphere with fixed boundary equations.

Moving away from the purely isotropic problem Ding and Chen[6] and Chen *et al.*[7] (see also references therein) looked into the nonaxisymmetric vibrations of a spherically (or transversely) isotropic full sphere under different operating conditions. For both the cases of a shell embedded in an elastic medium and the simpler case of fixed boundary conditions a comparison was made between frequencies calculated for a purely isotropic sphere to those of a transversely isotropic sphere. In an allied problem Wang *et al.*[8] examined the problem of thermal stress focusing in transversely isotropic full spheres but restricted their attention to radial displacements. However, in all the aforementioned references only single full isotropic or transversely isotropic spheres were considered.

Heyliger and Jilani[9] formulated the variational form of equations governing free vibrations of an orthotropic sphere and then used a Ritz approximation to compute the natural frequencies. In a work on electrostatic fields in layered piezoelectric spheres Heyliger and Wu[10] examined the effects of piezoelectric stiffening on the frequencies of transversely isotropic spheres. Using an analytical solution they computed radial vibrational frequencies with and without the piezoelectric coefficients, the latter case yielding the natural frequencies of a hollow elastic sphere.

A review containing an extensive bibliography of research on the vibrations of thin, moderately thick, and thick shallow shells was presented by Liew *et al.*[11] The authors survey the many *approximate* theories that have appeared in the literature but exact analytical solutions of three-dimensional theories for spherical shells are not covered. The use of laminated structures is widespread due to the engineering advantages that accrue from utilizing combinations of materials of different properties whose combined characteristics can transcend those of the individual components. In this vein the three-dimensional radial vibration problem is formulated and solved exactly for the first time for a composite hollow sphere laminated of spherically *orthotropic* and/or isotropic layers, under arbitrary linear boundary conditions.

II. ANALYSIS OF HOMOGENEOUS SINGLE LAYER SPHERE

The displacement field is given by

$$u_r = u_r(r) \equiv u, \quad u_\phi = 0, \quad u_\theta = 0 \tag{1}$$

and the single equation of motion is, in spherical coordinates,

$$\sigma_{r,r} + \frac{1}{r}(2\sigma_r - \sigma_\phi - \sigma_\theta) = \rho \ddot{u} \tag{2}$$

in which the subscript "$,r$" denotes differentiation with respect to the radial coordinate r and the dot notation denotes differentiation with respect to time.

Hooke's law for the spherically orthotropic material is

$$\begin{bmatrix} \sigma_r \\ \sigma_\phi \\ \sigma_\theta \end{bmatrix} = \begin{bmatrix} E_{11} & E_{12} & E_{13} \\ & E_{22} & E_{23} \\ \text{Symmetry} & & E_{33} \end{bmatrix} \begin{bmatrix} \epsilon_r \\ \epsilon_\phi \\ \epsilon_\theta \end{bmatrix}, \tag{3a}$$

where the E_{ij} are the elastic stiffnesses, σ_i are the normal stresses, the ϵ_i are the corresponding strains, and ρ is the mass density. For a spherically isotropic material

$$E_{13} = E_{12} \neq E_{23}, \quad E_{33} = E_{22} \neq E_{11}, \tag{3b}$$

whereas for a three-dimensional isotropic material

$$E_{11} = E_{22} = E_{33} = \lambda + 2G, \quad E_{12} = E_{13} = E_{23} = \lambda,$$
$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}, \quad G = \frac{E}{2(1+\nu)}. \tag{3c}$$

λ and G are the Lame elastic constants and ν is Poisson's ratio.

In view of Eq. (1) the strain-displacement relations are

$$\epsilon_r = u_{,r} \quad \text{and} \quad \epsilon_\phi = \epsilon_\theta = u/r. \tag{4}$$

a)Electronic mail: aer9801@aerodyne.technion.ac.il

For free vibrations,

$$u(r,t)=u^0(r)e^{i\omega t}, \tag{5}$$

where ω is the frequency of vibration, and the equation of motion (2), in view of Eqs. (3) and (4), becomes

$$u^0_{,rr}+\frac{2}{r}u^0_{,r}+\frac{1}{E_{11}}\left(\rho\omega^2-\frac{K}{r^2}\right)u^0=0, \tag{6}$$

where

$$K=E_{22}+E_{33}+2E_{23}-E_{12}-E_{13}. \tag{7}$$

Equation (6) reduces for an isotropic sphere, in view of Eq. (3b), to

$$u^0_{,rr}+\frac{2}{r}u^0_{,r}-\frac{2}{r^2}u^0+\Omega^2_{\rm iso}u=0, \tag{8}$$

where

$$\Omega^2_{\rm iso}=\rho\omega^2/(\lambda+2G) \tag{9}$$

as in Sec. 198 of Love.[3]

The solution of Eq. (6) is obtained as a special case of Eq. 2.162(Ia) in Kamke's collection of solutions of differential equations,[12] namely

$$u=r^{-1/2}Z_\mu(\Omega r)=r^{-1/2}[C_1J_\mu(\Omega r)+C_2Y_\mu(\Omega r)], \tag{10}$$

where

$$\Omega=\omega\sqrt{\frac{\rho}{E_{11}}}, \tag{11}$$

$$\mu=\frac{1}{2}\sqrt{1+4\frac{K}{E_{11}}}. \tag{12}$$

[Parenthetically it is noted that both Eq. (6) and the associated solution Eq. (10) reduce identically to the equation and solution presented by Heyliger and Wu[10] for the particular case of a transversely isotropic material.]

For a single layer free hollow sphere the boundary conditions are taken as

$$\sigma_r(a)=\sigma_r(b)=0, \tag{13}$$

where a is the inner radius of the sphere and b is its outer radius.

The determinantal equation from which Ω is determined takes the form:

$$\begin{vmatrix} E_{11}a\Omega J'_\mu(\Omega a)+(E_{12}+E_{13}-\frac{1}{2}E_{11})J_\mu(\Omega a) & E_{11}a\Omega Y'_\mu(\Omega a)+(E_{12}+E_{13}-\frac{1}{2}E_{11})Y_\mu(\Omega a) \\ E_{11}b\Omega J'_\mu(\Omega b)+(E_{12}+E_{13}-\frac{1}{2}E_{11})J_\mu(\Omega b) & E_{11}b\Omega Y'_\mu(\Omega b)+(E_{12}+E_{13}-\frac{1}{2}E_{11})Y_\mu(\Omega b) \end{vmatrix}=0, \tag{14a}$$

where a prime denotes a derivative with respect to r. For other boundary conditions the general solution (10) can be readily used.

It is of interest to note that for the particular case of a very thin spherical shell for which $a\to b$ Love's analysis can be generalized to yield the following equation that replaces Eq. (14a):

$$\frac{\partial}{\partial a}\left(\frac{E_{11}a\Omega J'_\mu(\Omega a)+(E_{12}+E_{13}-\frac{1}{2}E_{11})J_\mu(\Omega a)}{E_{11}a\Omega Y'_\mu(\Omega a)+(E_{12}+E_{13}-\frac{1}{2}E_{11})Y_\mu(\Omega a)}\right)=0. \tag{14b}$$

III. ANALYSIS OF A TWO-LAYER SPHERE

The formulation of the vibration problem of a laminated sphere is shown for two concentric spheres (see Fig. 1). Denoting the inner sphere by "1" and the outer one by "2" we have now two equations of motion of type (6):

$$u^{(j)}_{,rr}+\frac{2}{r}u^{(j)}_{,r}+\frac{1}{E^{(j)}_{11}}\left(\rho^{(j)}\omega^2-\frac{K^{(j)}}{r^2}\right)u^{(j)}=0, \quad j=1,2 \tag{15}$$

with

$$K^{(j)}=E^{(j)}_{22}+E^{(j)}_{33}+2E^{(j)}_{23}-E^{(j)}_{12}-E^{(j)}_{13}, \quad j=1,2 \tag{16}$$

with two boundary conditions (13) and two matching conditions at the interface $(r=c)$:

$$u^{(1)}(c)=u^{(2)}(c), \quad \sigma^{(1)}_r(c)=\sigma^{(2)}_r(c). \tag{17}$$

Note that the layered structure is assumed to have common natural frequency of vibration, ω.

The solution for each layer is

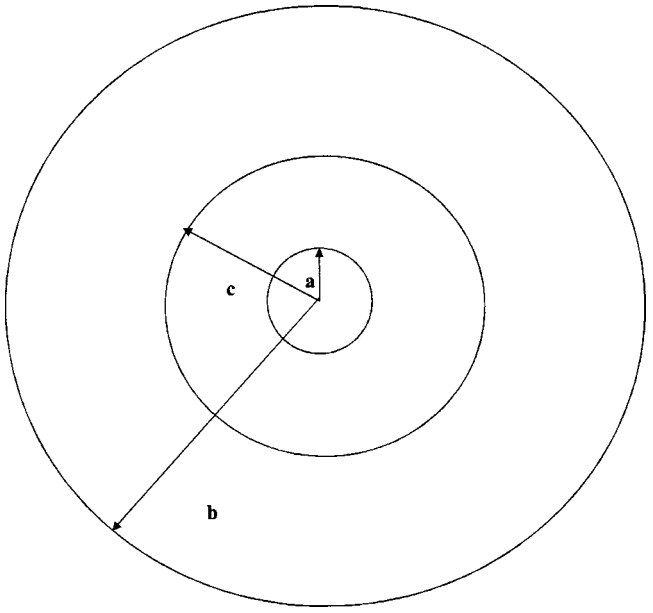


FIG. 1. Schematic of orthotropic two-layered hollow sphere.

TABLE I. Comparison between Love's results and predictions of current work.

Frequency, $\omega\times10^6$, according to Love's theory	Frequency, $\omega\times10^6$, according to current theory
1.404 108	1.399 978
3.318 411	3.325 261
5.051 865	5.091 359
6.824 036	6.833 440
8.556 802	8.566 541

$$u^{(i)}(r)=r^{-1/2}[C^{(i)}J_{\mu_i}(\Omega^{(i)}r)+D^{(i)}Y_{\mu_i}(\Omega^{(i)}r)],$$
$$i=1,2. \tag{18}$$

Application of the boundary and matching conditions gives the following four homogeneous equations for the constants $C^{(i)}$ and $D^{(i)}$:

$$\begin{bmatrix} F_{11} & F_{12} & 0 & 0 \\ 0 & 0 & F_{23} & F_{24} \\ F_{31} & F_{32} & F_{33} & F_{34} \\ F_{41} & F_{42} & F_{43} & F_{44} \end{bmatrix} \begin{bmatrix} C^{(1)} \\ D^{(1)} \\ C^{(2)} \\ D^{(2)} \end{bmatrix} = 0, \tag{19}$$

where the F_{ij} are given in the Appendix.

The natural frequencies are then extracted from the characteristic equation

$$|F_{ij}|=0. \tag{20}$$

Note that this three-dimensional elasticity theory could also be generalized and applied to the closed type solution of a multilayered sphere in the form of Eq. (18), for any linear boundary conditions. If a sphere is comprised of N layers the characteristic equation from which the natural frequencies will then be extracted will involve a $2N\times2N$ determinant.

IV. RESULTS AND DISCUSSION

Use has been made of the aforedescribed analysis to examine the frequency response of a number of single- and double-layered hollow spheres. The frequencies were computed using a routine of MATLAB that drew contours of the characteristic determinants in the frequency-inner radius plane for the single-layered sphere and the frequency-contact radius plane for the double-layered spheres. The routine also supplies numerical values of the relevant natural frequencies. As an independent check of our theory for an isotropic sphere having a pinhole inner radius of 0.001 the predictions of the determinantal equation (14a) were compared with the five lowest roots predicted by Love's classical theory[3] for a solid sphere. Excellent agreement was obtained, see Table I.

For the ensuing discussion three materials were considered. Aluminum (Al) was taken as an isotropic material with the following properties: $E_{11}=E_{22}=E_{33}=108.3\,\text{GPa}$, $E_{12}=E_{13}=E_{23}=53.3\,\text{GPa}$, $\rho=2.7\times10^3\,\text{kg/m}^3$.

Two orthotropic materials were also used for illustrative examples, glass-epoxy (GE) with properties:

$$E_{11}=40.5\ \text{GPa},\quad E_{22}=9.77\ \text{GPa},$$
$$E_{33}=10.59\ \text{GPa},\quad E_{12}=3.53\ \text{GPa},$$
$$E_{13}=3.74\ \text{GPa},\quad E_{23}=3.81\ \text{GPa},$$
$$\rho=2\times10^3\ \text{kg/m}^3$$

and boron-epoxy (BE) with properties:

$$E_{11}=211.1\ \text{GPa},\quad E_{22}=E_{33}=22.4\ \text{GPa},$$
$$E_{12}=8.16\ \text{GPa},\quad E_{13}=7.32\ \text{GPa},$$
$$E_{23}=5.8\ \text{GPa},\quad \rho=2.087\times10^3\ \text{kg/m}^3.$$

In Fig. 2 and Table II we display typical results for a single-layered GE sphere. The frequencies are shown as a function of the inner radius, a, with the outer radius, b, being 1 m. All frequencies grow monotonically as a increases except for the two lowest frequencies. The lowest frequency actually decreases monotonically with a. This decrease in frequency was about 40% as the inner radius increases from 0.1 to 0.9 m, i.e., as the sphere becomes thinner. (The same qualitative behavior was also found for single-layered aluminum and BE spheres.) The second frequency slightly de-

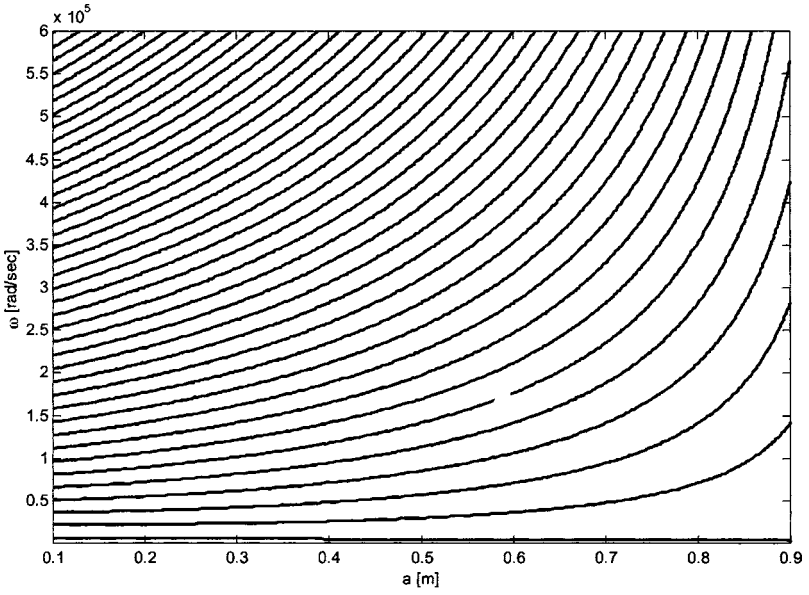


FIG. 2. Natural frequency response of a single-layered glass-epoxy hollow sphere as a function of inner radius.

TABLE II. First five frequencies for single-layered glass epoxy sphere.

a (m)	ω (rad/s) 1	ω (rad/s) 2	ω (rad/s) 3	ω (rad/s) 4	ω (rad/s) 5
0.1	6002.98	22 430.71	36 746.29	51 276.22	66 129.37
0.2	5738.54	22 279.01	38 219.33	55 017.02	72 215.19
0.3	5419.75	23 299.69	42 112.46	61 754.79	81 681.48
0.4	5106.59	25 680.33	48 258.46	71 437.06	94 812.81
0.5	4829.37	29 723.70	57 302.93	85 321.38	113 479.04
0.6	4479.95	36 337.28	71 181.19	106 364.87	141 633.96
0.7	4161.41	47 770.06	94 567.28	141 597.59	188 660.78
0.8	3887.23	71 053.51	141 569.59	212 188.48	282 848.65
0.9	3652.23	141 547.89	282 838.19	424 187.32	565 550.98

creases for small values of a before increasing monotonically for a greater than about 0.25. In general, it was found that the lowest frequencies of spheres comprised of each of the three materials obey the inequality $\omega_{\mathrm{GE}}<\omega_{\mathrm{BE}}<\omega_{\mathrm{Al}}$. This was also true for subsequent frequencies.

We next turn to bilayered spheres and plot frequency versus the radius of the interface between the two concentric spheres. The inner radius of the inner sphere and the outer radius of the outer sphere are held fixed at 0.1 and 1 m, respectively. Variation of the value of the interface radius, c, is therefore equivalent to altering the composition of the composite sphere. In Fig. 3 we illustrate the results for an Al/GE combination in which the inner sphere is made of aluminum and the outer one of glass epoxy. Numerical values of the first five frequencies are listed in Table III. For all frequencies shown it is evident that the behavior of ω as a function of the interface radius is not linear. The values of the frequency at the left/right edge of the figure tend those of a single layer GE/Al sphere. The curves undulate between these bounding values.

In Figs. 4 and 5 the frequencies are shown for both double combinations of the orthotropic materials. In Fig. 4 BE is the inner material and GE is the outer one, whereas in Fig. 5 the locations of the materials have been reversed. Similar nonlinear behavior of the frequencies as a function of c to that noted in Fig. 3 is observed in Fig. 4. In Fig. 5, in

TABLE III. First five frequencies for double-layered sphere; inner layer aluminum, outer layer glass epoxy.

c (m)	ω (rad/s) 1	ω (rad/s) 2	ω (rad/s) 3	ω (rad/s) 4	ω (rad/s) 5
0.1	6 002.98	22 430.71	36 746.29	51 276.22	66 129.37
0.2	6 750.99	25 178.07	41 435.72	56 552.20	70 070.14
0.3	7 509.92	27 598.16	44 239.51	57 242.66	71 930.99
0.4	8 536.10	30 026.34	44 528.13	59 508.80	76 852.47
0.5	9 899.88	31 385.76	45 388.40	63 678.27	76 432.78
0.6	11 668.39	31 261.79	49 286.84	63 089.34	82 327.98
0.7	13 805.72	31 463.85	50 972.54	67 288.03	82 088.99
0.8	15 845.10	34 942.32	50 362.88	69 028.88	88 706.98
0.9	17 002.47	37 688.74	55 024.29	71 210.43	88 561.76

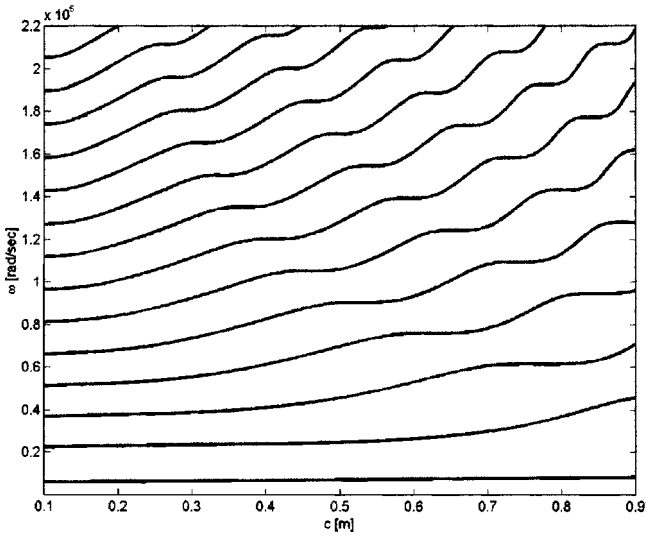


FIG. 4. Natural frequency response of a two-layered orthotropic hollow sphere as a function of interface radius; inner material BE, outer material GE.

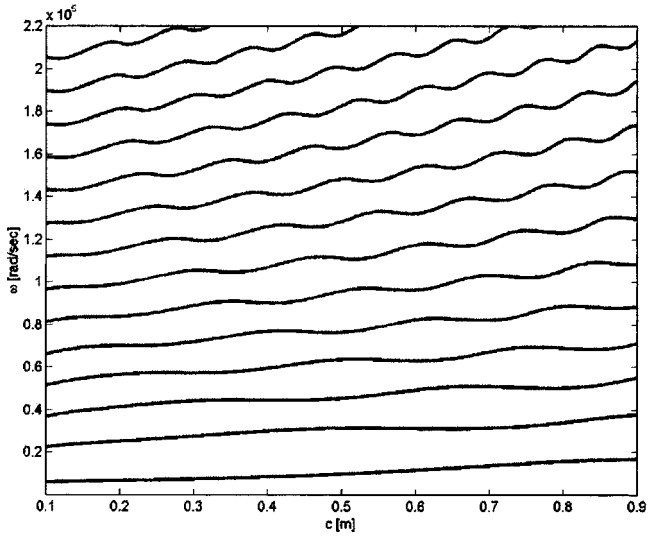


FIG. 3. Natural frequency response of a two-layered isotropic+orthotropic hollow sphere as a function of interface radius; inner material Al, outer material GE.

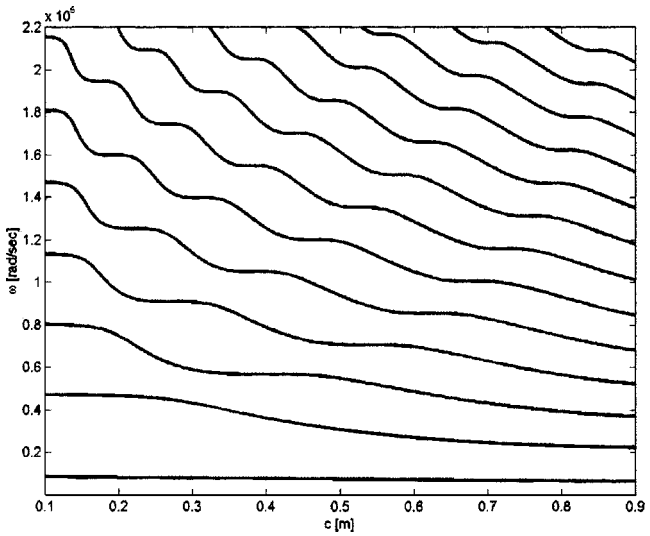


FIG. 5. Natural frequency response of a two-layered orthotropic hollow sphere as a function of interface radius; inner material GE, outer material BE.

which the inner material is GE, the effect of material reversal is illustrated. If $c=0.5005$ the mass of material in the two spheres is approximately equal. For this particular combination all the frequencies of the BE (inner)/GE (outer) sphere are lower than those of the GE (inner)/BE (outer) sphere.

These results all point to the subtle role that the material lay-up plays in determining the frequency response. In addition, the general applicability of the theory to *any* spherical shell thickness has been demonstrated.

V. CONCLUSION

In this work an exact three-dimensional solution has been developed for the radial vibrations of multilayered laminated spheres. The theory is not restricted by the shell's overall thickness. This obviates the need for using approximate shell theories and enables their accuracy to be checked (see, for example, Liew *et al.*[11] and Wilkinson[13]). The calculated results for single- and double-layered thick orthotropic shells provide a clear indication of how shell lamination (through both material and geometrical properties) could potentially be used for vibration control.

ACKNOWLEDGMENTS

J. B. G. gratefully acknowledges the partial support of the Lady Davis Chair in Aerospace Engineering and the Technion Fund for the Promotion of Research.

APPENDIX

The elements of the characteristic determinant for a two-layer shell are

$$F_{11}=E_{11}^{(1)}a\Omega^{(1)}J'_{\mu_1}(\Omega^{(1)}a)+(E_{12}^{(1)}+E_{13}^{(1)}-\tfrac{1}{2}E_{11}^{(1)})J_{\mu_1}(\Omega^{(1)}a),$$

$$F_{12}=E_{11}^{(1)}a\Omega^{(1)}Y'_{\mu_1}(\Omega^{(1)}a)+(E_{12}^{(1)}+E_{13}^{(1)}-\tfrac{1}{2}E_{11}^{(1)})Y_{\mu_1}(\Omega^{(1)}a),$$

$$F_{23}=E_{11}^{(2)}b\Omega^{(2)}J'_{\mu 2}(\Omega^{(2)}b)+(E_{12}^{(2)}+E_{13}^{(2)}-\tfrac{1}{2}E_{11}^{(2)})J_{\mu 2}(\Omega^{(2)}b),$$

$$F_{24}=E_{11}^{(2)}b\Omega^{(2)}Y'_{\mu 2}(\Omega^{(2)}b)+(E_{12}^{(2)}+E_{13}^{(2)}-\tfrac{1}{2}E_{11}^{(2)})Y_{\mu 2}(\Omega^{(2)}b),$$

$$F_{31}=J_{\mu_1}(\Omega^{(1)}c),$$

$$F_{32}=Y_{\mu_1}(\Omega^{(1)}c),$$

$$F_{33}=-J_{\mu 2}(\Omega^{(2)}c),$$

$$F_{34}=-Y_{\mu 2}(\Omega^{(2)}c),$$

$$F_{41}=E_{11}^{(1)}c\Omega^{(1)}J'_{\mu_1}(\Omega^{(1)}c)+(E_{12}^{(1)}+E_{13}^{(1)}-\tfrac{1}{2}E_{11}^{(1)})J_{\mu_1}(\Omega^{(1)}c),$$

$$F_{42}=E_{11}^{(1)}c\Omega^{(1)}Y'_{\mu_1}(\Omega^{(1)}c)+(E_{12}^{(1)}+E_{13}^{(1)}-\tfrac{1}{2}E_{11}^{(1)})Y_{\mu_1}(\Omega^{(1)}c),$$

$$F_{43}=-E_{11}^{(2)}c\Omega^{(2)}J'_{\mu 2}(\Omega^{(2)}c)-(E_{12}^{(2)}+E_{13}^{(2)}-\tfrac{1}{2}E_{11}^{(2)})J_{\mu 2}(\Omega^{(2)}c),$$

$$F_{44}=-E_{11}^{(2)}c\Omega^{(2)}Y'_{\mu 2}(\Omega^{(2)}c)-(E_{12}^{(2)}+E_{13}^{(2)}-\tfrac{1}{2}E_{11}^{(2)})Y_{\mu 2}(\Omega^{(2)}c),$$

where

$$\Omega^{(i)}=\omega\sqrt{\frac{\rho^{(i)}}{E_{(11)}^{(i)}}}\quad\text{for } i=1,2$$

and

$$\mu_i=\frac{1}{2}\sqrt{1+4\frac{K^{(i)}}{E_{11}^{(i)}}},$$

$$K^{(i)}=E_{22}^{(i)}+E_{33}^{(i)}+2E_{23}^{(i)}-E_{12}^{(i)}-E_{13}^{(i)},\quad i=1,2.$$

[1] S. D. Poisson, "Memoire sur l'equilibre et le movement des corps elastiques," Mem. De l'Acad. **8**, 357–571 (1829).
[2] H. Lamb, "On the vibrations of an elastic sphere," Proc. London Math. Soc. **13**, 189–212 (1882).
[3] A. E. H. Love, *Treatise on the Mathematical Theory of Elasticity*, 4th revised ed. (1944), Chap. XII, Sec. 195.
[4] P. J. Scafbuch, F. J. Rizzo, and R. B. Thomson, "Eigenfrequencies of an elastic sphere with fixed boundary conditions," J. Appl. Mech. **59**, 458–459 (1992).
[5] P. Debye, "Zur theorie der spezifischen warmen," Ann. Phys. (Leipzig) **39**, 789–839 (1912).
[6] H. Ding and W. Chen, "Nonaxisymmetric free vibrations of a spherically isotropic spherical shell embedded in an elastic medium," Int. J. Solids Struct. **33**, 2575–2590 (1996).
[7] W. Q. Chen, J. B. Cai, G. R. Ye, and H. J. Ding, "On eigenfrequencies of an anisotropic sphere," J. Appl. Mech. **67**, 422–424 (2000).
[8] X. Wang, C. Wang, G. Lu, and B. M. Zhou, "Thermal stresses-focussing in a transversely isotropic sphere and an isotropic sphere," J. Therm. Stresses **25**, 31–44 (2002).
[9] P. R. Heyliger and A. Jilani, "The free vibrations of inhomogeneous elastic cylinders and spheres," Int. J. Solids Struct. **29**, 2689–2708 (1992).
[10] P. R. Heyliger and Y-C. Wu, "Electrostatic fields in layered piezoelectric spheres," Int. J. Eng. Sci. **37**, 143–161 (1999).
[11] K. M. Liew, C. W. Lim, and S. Kitipornchai, "Vibration of shallow shells; A review with bibliography," Appl. Mech. Rev. **50**, 431–444 (1997).
[12] E. Kamke, *Differentialgleichungen Losungsmethoden u. Losungen* (Tuebner, 1983), Vol. 1, pp. 438–440.
[13] J. P. Wilkinson, "Natural frequencies of closed spherical shells," J. Acoust. Soc. Am. **38**, 367–368 (1965).

Active control of acoustic reflection, absorption, and transmission using thin panel speakers

H. Zhu, R. Rajamani,[a] and K. A. Stelson
Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota 55455

(Received 3 October 2001; revised 27 October 2002; accepted 6 November 2002)

This paper explores the development of thin panels that can be controlled electronically so as to provide surfaces with desired reflection coefficients. Such panels can be used as either perfect reflectors or absorbers. They can also be designed to be transmission blockers that block the propagation of sound. The development of the control system is based on the use of wave separation algorithms that separate incident sound from reflected sound. In order to obtain a desired reflection coefficient, the reflected sound is controlled to appropriate levels. The incident sound is used as an acoustic reference for feedforward control and has the important property of being isolated from the action of the control system speaker. In order to use a panel as a transmission blocker, the acoustic pressure behind the panel is driven to zero. The use of the incident signal as a reference again plays a key role in successfully reducing broadband transmission of sound. The panels themselves are constructed using poster board and small rare-earth actuators. Detailed experimental results are presented showing the efficacy of the algorithms in achieving real-time control of reflection or transmission. The panels are able to effectively block transmission of broadband sound. Practical applications for these panels include enclosures for noisy machinery, noise-absorbing wallpaper, the development of sound walls, and the development of noise-blocking glass windows. [DOI: 10.1121/1.1534834]

PACS numbers: 43.50.Ki, 43.50.Jh, 43.20.El [MRS]

I. INTRODUCTION

Active noise cancellation (ANC) is achieved by introducing a canceling ''antinoise'' wave of equal amplitude and opposite phase using a secondary source. A review of ANC developments can be found in Hansen (1991) and Kuo and Morgan (1999). Lueg (1936) first suggested the idea of active noise cancellation. Early work on ANC used analog techniques. Chaplin (1977) introduced digital techniques in his ANC patent. Since then, much work on ANC using digital-processing techniques has been published. Adaptive feedforward control is the most popular and successful approach used in ANC (Kuo and Morgan, 1996). Feedforward control involves feeding a signal related to the disturbance input (called the primary noise) into the controller which then generates a signal to drive a speaker in such a way as to cancel the disturbance. This signal related to the primary noise is called the reference signal.

Results on many successful feedforward ANC systems have been published. However, the major limitations of ANC systems must be noted. First, most ANC systems need a reference signal. In the absence of a nonacoustic reference signal (such as from a speed sensor), reference microphones can be used to pick up signals from the primary source before the noise propagates to the secondary source. However, this leads to the ''secondary source effect.'' The reference microphones will not only pick up signals from the primary source but also those from the secondary source. A second limitation is that a high coherence between the reference microphone and the primary source is needed to achieve good performance. An additional complication is that online secondary path (from secondary source to the error microphone) estimation is needed to achieve long-term performance due to the nonstationary nature of ANC systems. However, it is difficult to estimate the secondary path online since random signals need to be used to excite the system and this tends to degrade the performance. Even if all of the above limitations could be addressed, it would still only be relatively easy to cancel noise at a point, i.e., at the position of the error microphone. It is very difficult for an ANC system to achieve global noise cancellation in a 3D environment such as in an enclosure. This is especially due to the limitations in the number of speakers and microphones that can be used in practical applications. This current research aims to address some of these limitations of ANC.

In this paper, we concentrate on the development of thin panels which can be electronically controlled so as to achieve desired acoustic properties. We develop algorithms for controlling the reflection coefficients of such panels as well as for using these panels as noise transmission blockers. The advantages of this approach to active noise control are that such panels can be used to prevent the entry of noise or the creation of noise, rather than the control of noise by active cancellation after it has already entered an enclosure. For example, panels made of glass can be used as window panes to prevent the entry of sound through windows in houses close to airports. Similarly, panels can be used to develop an enclosure for noisy machinery so as to prevent propagation of noise from the machinery. Such panels can also be used as wallpaper in rooms with noisy machinery so

[a]Electronic mail: rajamani@me.umn.edu, tel: (612) 626-7961, fax: (612) 624-1398.

as to prevent acoustic reflection and the occurrence of standing waves in the room.

The control system developed in this paper utilizes the separation of sound at a point into incident and reflected waves. The earliest work related to separation of reflected and incident sound is found in Guicking and Karcher (1984). They used two microphones and analog electronics consisting of four subtractions and four delays to obtain the incident wave and the reflected wave. They then attempted control of the reflected wave by appropriately setting the phase and gain of an amplifier.

The body of work closest to our approach is the research on active impedance control conducted by several researchers. Acoustic impedance at a surface is defined by

$$Z = \frac{p}{u}. \tag{1}$$

The acoustic impedance of air is $Z_o = 1/\rho_o c$, where ρ_o is the density of air and c is the speed of sound in air. By controlling the impedance at a surface to be equal to that of air, the reflection coefficient of the surface can be made zero if normal incidence is considered. A few researchers have studied this approach to active impedance control. Mehta *et al.* (1998) designed active acoustic treatment (AAT) cells using feedback control. Each cell included a microphone, a speaker, and an absorption sheet. They then used many cells aligned with the topside of a duct to attempt to attenuate sound transmission through a duct. Their objective was to obtain a perfect sound absorber.

A research group in France has been very active in the area of impedance control. In 1994, Thenail *et al.* studied how to actively increase the absorbent properties of a porous material. They found that the absorption is maximum when the space between the porous material and a rigid boundary is maintained at odd multiples of one quarter of the wavelength, and that driving the pressure at the back of an optimized porous material to zero will give maximum absorption. The latter finding is later used by them to implement indirect impedance control (Furstoss *et al.*, 1997). As for direct impedance control, Furstoss *et al.* (1997) used an accelerometer to measure velocity directly and a microphone to measure pressure. Thus, measuring impedance directly, they then control the impedance in a duct to simulate wall impedance control (Thenail *et al.*, 1997). They also used these two methods to actively control the sound field in a cavity via wall impedance control (Lacour *et al.*, 2000). Both one-dimensional cavities and three-dimensional cavities were investigated. Other related research has been conducted by Henrioulle *et al.* (1999), who designed a 1/4-wavelength absorber. This idea is similar to the work of Thenail *et al.* (1994). However, an important difference is that they used a flat speaker as a control actuator, which saves space (Henrioulle *et al.*, 1999). All of the above noted researchers aimed to control the absorption indirectly with the help of absorbing material, while the approach presented in this paper controls the sound reflection directly without absorbing material.

As far as control of sound transmission through a wall or a panel, very few studies have been conducted. Paurobally *et al.* (1999) investigated the use of feedback control to attenuate sound transmission through a double-panel partition (DPP). In their preliminary research, they reported a single-channel system with the secondary source being placed in the partition and pure tones being used to test the control system. The research related to DPP is different from the current research presented in this paper. In this paper, the panel itself plays both the role of sound entrance and the role of secondary source.

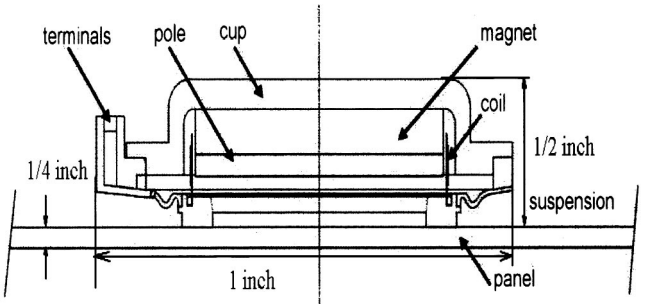


FIG. 1. NXT driver for a flat panel speaker.

The contributions of the present paper are the development of algorithms to enable control of the acoustic reflection coefficient or enable the prevention of noise transmission through the panel. The novelty of our approach is that the system developed in the paper is based purely on the use of microphones as sensors and that an algorithm to avoid the influence of the secondary source on the reference microphone is developed. The paper includes detailed experimental results documenting the performance of the both the reflection control systems and the ''transmission blocker'' systems.

II. FLAT PANEL ACTUATORS

We seek to use thin panels as speakers or actuators in our research. While a variety of exciter technologies can be considered for energizing the panel, including piezoelectric transducers, we chose a moving coil electromagnetic motor manufactured by Kodel, Inc. The use of a moving coil exciter ensures compatibility with conventional amplifiers. The exciter has a suspension which is glued to the desired panel. It also has terminals through which it can be connected to an amplifier. Allowable panels with this exciter include thin (about 6 mm) solid panels whose surface areas can range from several meters square to several centimeters square. The panels are different from conventional woofer speakers which operate under the ''pistonic'' mode of operation. Instead, they operate under a ''distributed mode'' in which the panel vibrates flexibly. The exciter is based on a technology developed by NXT (New Transducers Ltd.) under the principle of ''optimally distributed modes of vibration.'' Figure 1 shows a typical panel speaker consisting of both the panel and the exciter.

The typical response of the dynamics of a panel speaker (distributed modes loudspeaker, DML) is shown below in Fig. 2 and compared to that of a conventional cone speaker. It can be seen that the frequency response has many local valleys and peaks and does not offer the kind of flat response that would be ideal for feedback control. However, the dy-

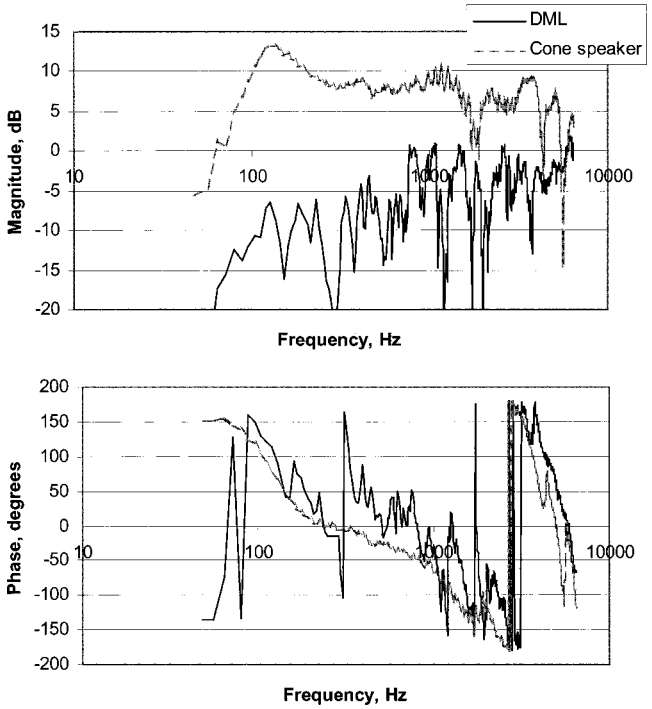


FIG. 2. Frequency response of DML and 8 in. standard cone speaker.

namics has been found to be consistent and repeatable. Feedforward control has been successfully implemented using these panels. Future research to optimize panel parameters such as size, panel material, internal damping, and exciter position for each application will be useful.

III. ACTIVE CONTROL OF REFLECTION COEFFICIENT

This section develops algorithms for separation of sound at the panel into incident and reflected signals and a control system that utilizes these signals for control of the reflection coefficient.

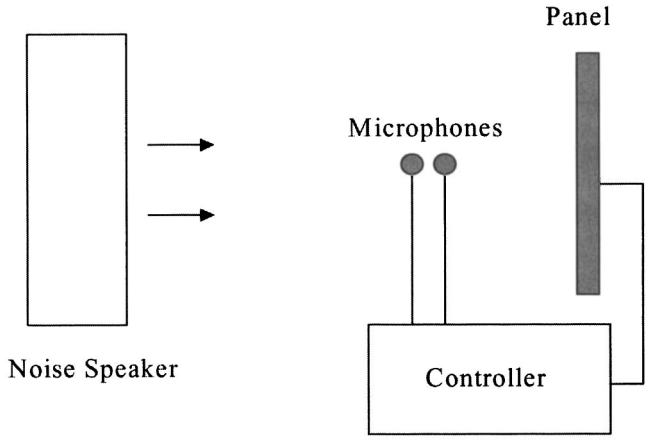


FIG. 3. Panel configuration for reflection control.

A. Wave separation using the integration method

The experimental system utilizes two microphones placed a few centimeters apart in front of the panel, as shown in Fig. 3. Normal incidence on the panel is assumed.

Let the acoustic pressure signals picked up by the two microphones be p_1 and p_2. If the distance d between the microphones is small relative to the smallest wavelength of the sound, the pressure at the midpoint is approximately

$$p=\frac{p_1+p_2}{2}. \tag{2}$$

For a plane wave, the momentum equation yields

$$\rho\frac{\partial u}{\partial t}+\frac{\partial p}{\partial x}=0. \tag{3}$$

Since the distance between the two microphones is small, the spatial derivative can be approximated by

$$\frac{\partial p}{\partial x}=\frac{p_2-p_1}{d}. \tag{4}$$

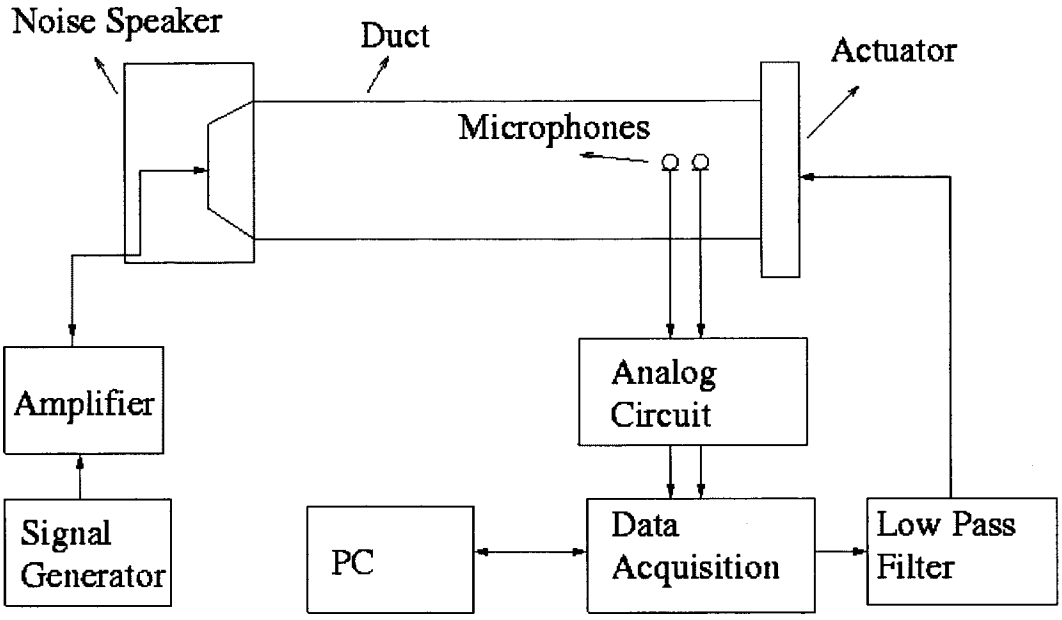


FIG. 4. Experimental setup for reflection coefficient control.

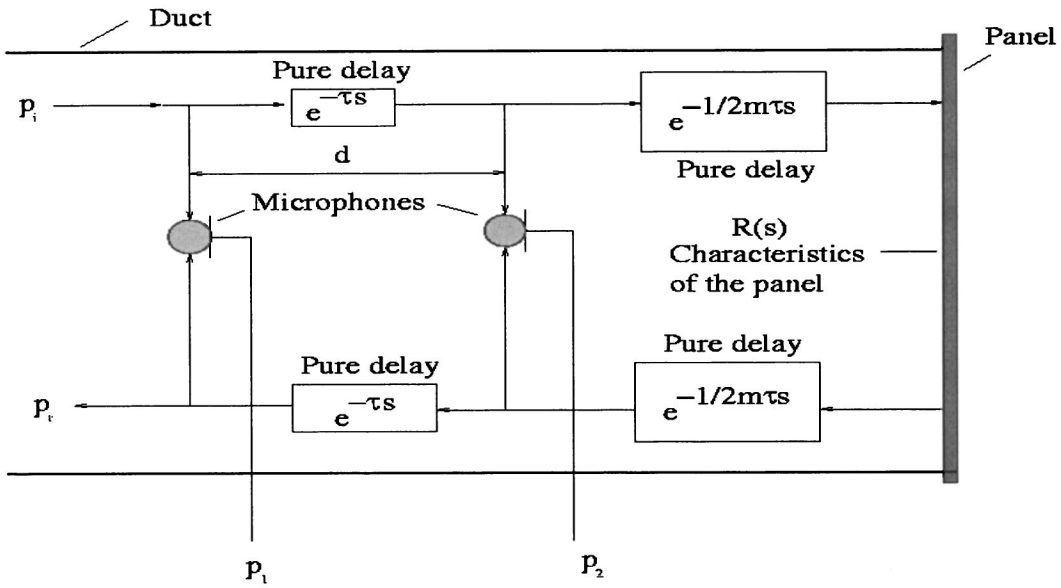


FIG. 5. The delay method for wave separation.

Substituting into Eq. (3), the particle velocity is calculated as

$$u(t)=\frac{1}{\rho d}\int_0^t (p_1-p_2)dx. \tag{5}$$

The incident wave can be expressed as (Beranek, 1954)

$$p_i=\sum_n A_n e^{j(w_n t-k_n x)}, \tag{6}$$

where the wave number k_n is related to the frequency ω_n and the speed of sound c by the relation $k_n=\omega_n/c$.

Substituting into the momentum Eq. (3), the particle velocity corresponding to the incident wave is

$$u_i=\frac{1}{\rho_0 c}p_i. \tag{7}$$

Similarly the particle velocity caused by the reflected wave can be obtained as

$$u_r=\frac{1}{\rho_0 c}p_r. \tag{8}$$

Thus, the overall particle velocity at the midpoint of the two microphones can be expressed as

$$u=u_i+u_r=\frac{1}{\rho_0 c}(p_i-p_r). \tag{9}$$

The associated pressure at the midpoint is

$$p=p_i+p_r. \tag{10}$$

Combining Eqs. (9) and (10), the incident wave and the reflected wave can be calculated by

$$p_i=\tfrac{1}{2}(p+\rho_0 c u), \tag{11}$$

and

$$p_r=\tfrac{1}{2}(p-\rho_0 c u). \tag{12}$$

To calculate p_i and p_r, p_1 and p_2 are first measured by the two closely positioned microphones. p is calculated using Eq. (2). Numerical integration is then used to update the particle velocity u, as shown in Eq. (5). Finally, Eqs. (11) and (12) are used to obtain p_i and p_r.

B. Active control of absorption or reflection

The experimental setup for control of the reflection coefficient is shown in Fig. 4. In the experiments, the primary noise is generated by a woofer speaker driven by a signal from a PC equipped with a data acquisition system. A 2-meter-long duct is used to isolate the environmental effects and ensure plane waves. The cross section of the duct is 17 by 17 cm. Thus, its cutoff frequency is about 1000 Hz. Two microphones, as described earlier, are used to measure acoustic pressure. The analog circuit provides functions of amplification and filtering. A CIO-DAS6402/12 data acquisition board is used to support data communication between PC and speakers and microphones. The control algorithm is implemented via a PC real-time toolbox with TURBO C used to develop the real-time code.

Assume that the desired reflection coefficient is a transfer function $R(s)$ that is the Laplace transform of the ratio

$$r=\frac{p_r}{p_i}. \tag{13}$$

The desired transfer function of the reflected wave is

$$P_{\text{rdes}}(s)=R(s)P_i(s), \tag{14}$$

and the corresponding desired reflected wave $p_{\text{rdes}}(t)$ can be calculated from

$$p_{\text{rdes}}(t)=\int_0^t p_i(t-\tau)r(\tau)d\tau. \tag{15}$$

The error p_r-p_{rdes} is used as the residue for feedforward control. The incident sound $p_i(t)$ is used as the reference signal. The secondary path transfer function is obtained from

$$S(s)=P_r(s)-R(s)P_i(s). \tag{16}$$

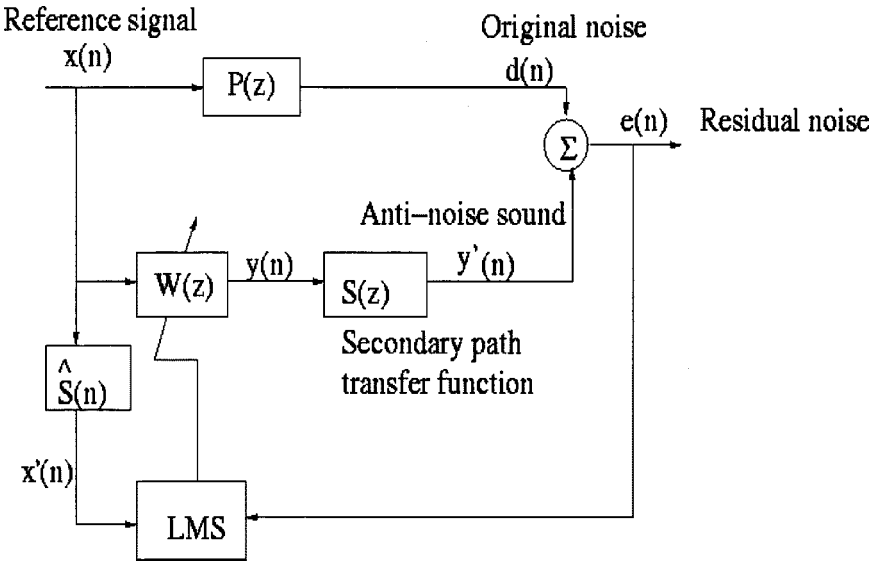


FIG. 6. FXLMS feedforward control.

Compared to the case where the error is the residual noise measured by a microphone, here the secondary path transfer function is not physically measured but is conceptual and has to be calculated from measurements. Using the incident sound as a reference signal ensures that the effect of the secondary source on the reference signal is avoided. Reflection can be controlled to obtain any desired reflection coefficient including $r=0$ (i.e., perfect absorption) and $r=1$ (perfect reflection).

C. Wave separation using the delay method

In the integration method, numerical integration must be used, slowing down the processing. In this section a different approach, the delay method, is proposed. The experimental setup is the same as shown in Fig. 4. Figure 5 is used to illustrate the idea.

As can be seen in Fig. 5, the signals p_1 and p_2 can be separated into two parts: incident wave and reflected wave. For a plane wave, the difference between the p_{1_i} and p_{2_r} is a pure time delay. Likewise, the difference between p_{2_r} and p_{1_r} is a pure time delay. The difference between p_{2_i} and p_{2_r} is another time delay multiplied by $R(s)$. If the delay between the microphones is τ, then $m=2\ell/d$, where ℓ is the distance between the second microphone and the actuator. Considering these relationships, we get

$$\frac{P_1(s)}{P_i(s)}=1+R(s)e^{-(m+2)\tau s}, \tag{17}$$

and

$$\frac{P_2(s)}{P_i(s)}=e^{-\tau s}+R(s)e^{-(m+1)\tau s}. \tag{18}$$

Let

$$x(t)=p_2(t)-p_1(t-\tau), \tag{19}$$

and

$$y(t)=p_1(t)-p_2(t-\tau). \tag{20}$$

We have

$$\frac{X(s)}{P_i(s)}=R(s)e^{-(m+1)\tau s}(1-e^{-2\tau s}), \tag{21}$$

and

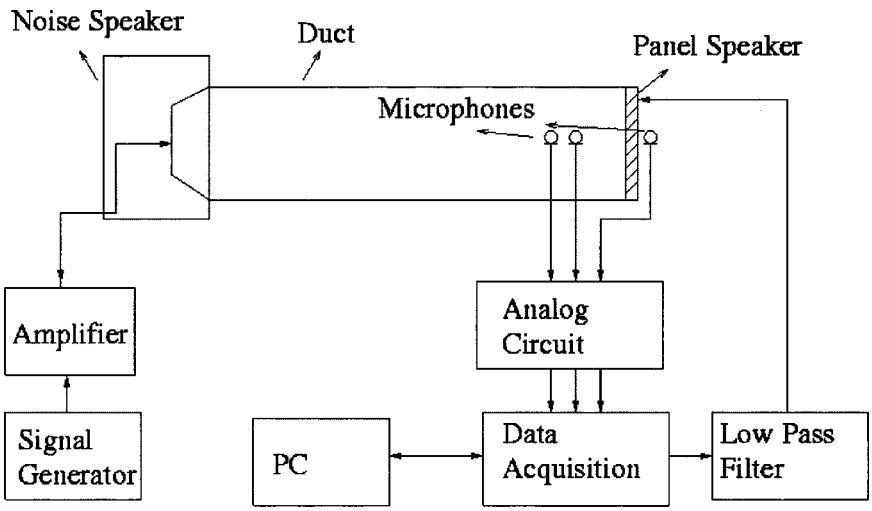


FIG. 7. Experimental setup for transmission control.

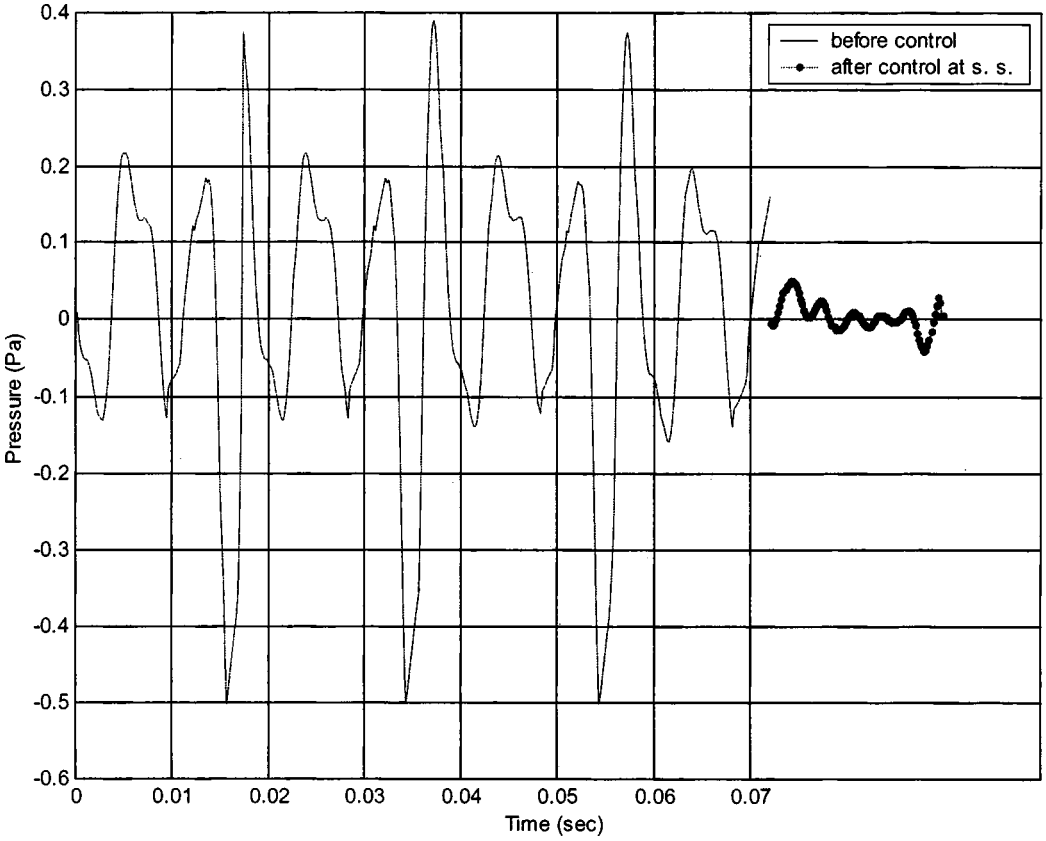


FIG. 8. Reflection control to achieve perfect absorption via integration method (primary noise consists of frequency components 150, 200, 250, and 300 Hz).

$$\frac{Y(s)}{P_i(s)} = (1 - e^{-2\tau s}). \tag{22}$$

Combining the above two equations gives

$$\frac{X(s)}{Y(s)} = R(s)e^{-(m+1)\tau s}. \tag{23}$$

Thus, in the delay method for wave separation, only delayed signals of p_1 and p_2 are used. x is controlled rather than p_r to obtain desired absorption/reflection. This method of wave separation is similar to the method presented in Guicking and Karcher (1984).

After the desired reflection coefficient transfer function, is determined, the desired transfer function of x can be found as

$$X_{\text{des}}(s) = Y(s)R_{\text{des}}(s)e^{-(m+1)\tau s}, \tag{24}$$

and the corresponding desired $x_{\text{des}}(t)$ can be calculated by the inverse Laplace transform

$$x_{\text{des}}(t) = L^{-1}(X(s)). \tag{25}$$

The error $x(t) - x_{\text{des}}(t)$ is used as the residue for feedforward control. The secondary path transfer function for feedforward control is

$$S(s) = X(s) - Y(s)R_{\text{des}}(s)e^{-(m+1)\tau s}. \tag{26}$$

Again, compared to the case where error is the residual noise measured by a microphone, here the secondary path transfer function is not physically measured but is conceptual and must be calculated from measurements. The algorithm obtained is thus quite simple with no integration required. However, an incident wave that can be used as a reference signal is not explicitly generated [instead the signals $x(t)$ and $y(t)$ are generated both of which contain information from both the incident and reflected wave]. The signal picked up by one of the microphones or $y(t)$ has to be used as the reference and the performance is therefore expected to degrade in the absence of a nonacoustic reference signal.

D. Filtered-x LMS algorithm

In the active control of reflection or absorption, the filtered-x least-mean-square (FXLMS) algorithm is used for the feedforward control. Figure 6 summarizes the popular FXLMS algorithm (Kuo and Morgan, 1996).

Here, $x(n)$ is the reference signal; $y(n)$ is the desired control (speaker) signal; $y'(n)$ is the actual sound of the secondary source; $d(n)$ is the undesired primary noise; $e(n)$ is the residual noise at downstream measured by an error microphone; $x'(n)$ is the filtered version of $x(n)$; $P(z)$ is the unknown transfer function between the reference microphone and the secondary source; $S(z)$ is the dynamics from the secondary source to the error microphone; $\hat{S}(z)$ is the estimation of this secondary path; and $W(z)$ is the digital filter that is adapted to generate the correct control signals to the secondary source. The objective is to minimize $e(n)$ via minimizing the instantaneous squared error, $\hat{\xi}(n) = e^2(n)$. The most widely used method to achieve this is the FXLMS algorithm, which updates the coefficients of $W(z)$ in the negative gradient direction with appropriate step size μ

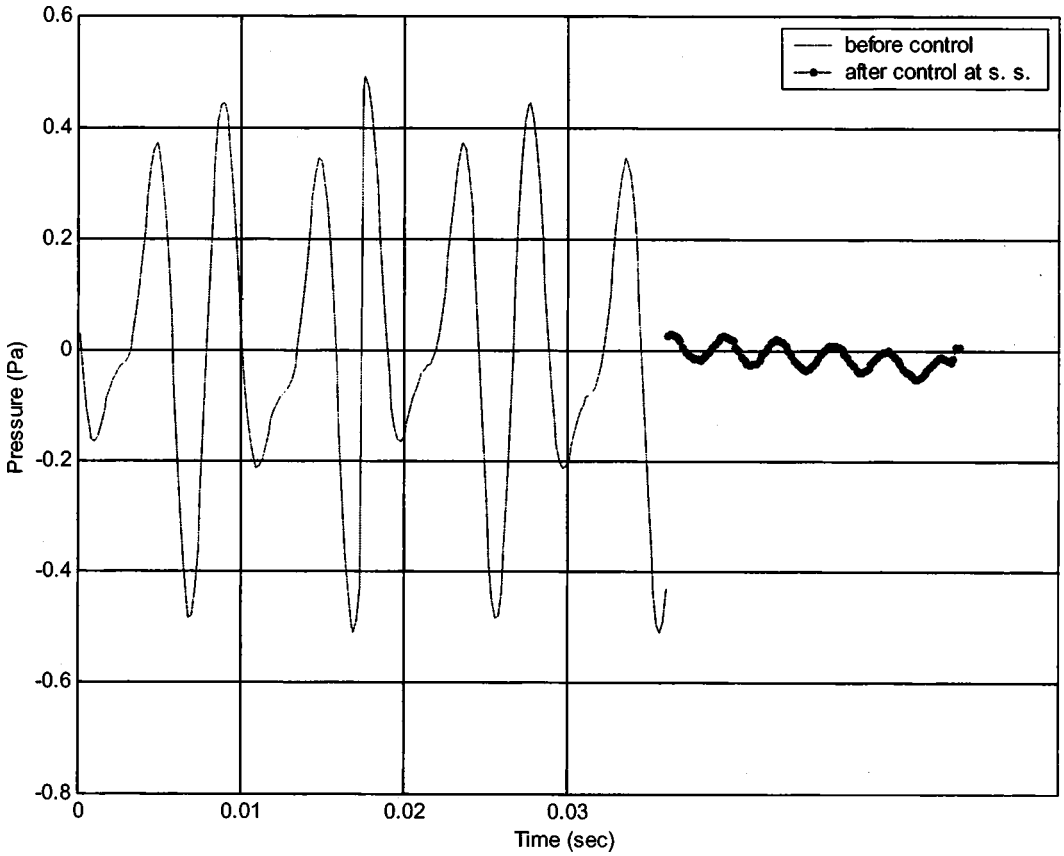


FIG. 9. Reflection control to achieve perfect absorption via integration method (primary noise consists of frequency components 100 and 300 Hz).

$$\mathbf{w}(n+1)=\mathbf{w}(n)-\frac{\mu}{2}\nabla\hat{\xi}(n), \tag{27}$$

where $\nabla\hat{\xi}(n)$ is the instantaneous estimate of the mean-square error gradient at time n, and can be expressed as

$$\begin{aligned}\nabla\hat{\xi}(n)&=2[\nabla e(n)]e(n)\\&=2[s(n)*x(n)]e(n)=2x'(n)e(n).\end{aligned} \tag{28}$$

By substituting the above equation back into (27), we have the filtered-x least-mean-square (FXLMS) algorithm

$$\mathbf{w}(n+1)=\mathbf{w}(n)+\mu x'(n)e(n), \tag{29}$$

where $x'(n)$ is estimated as $\hat{s}(n)*x(n)$.

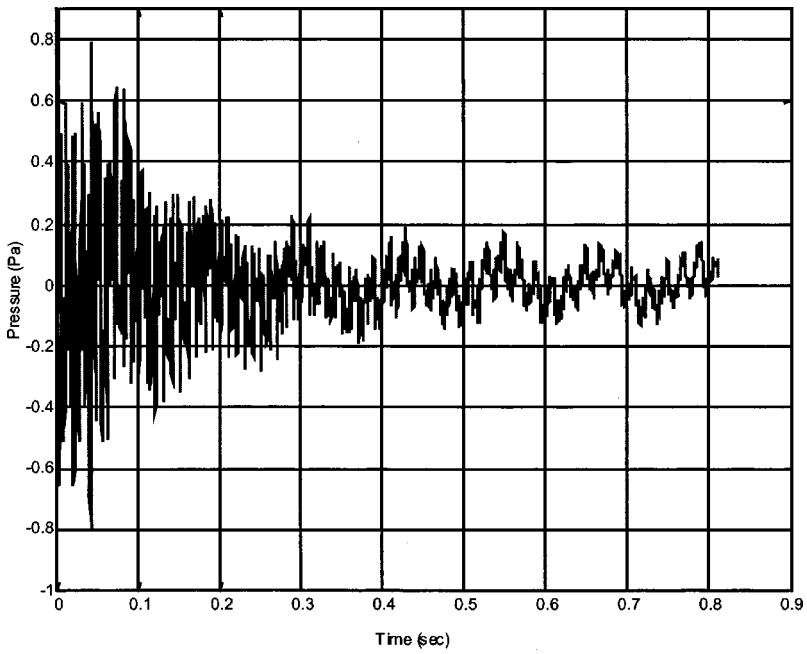
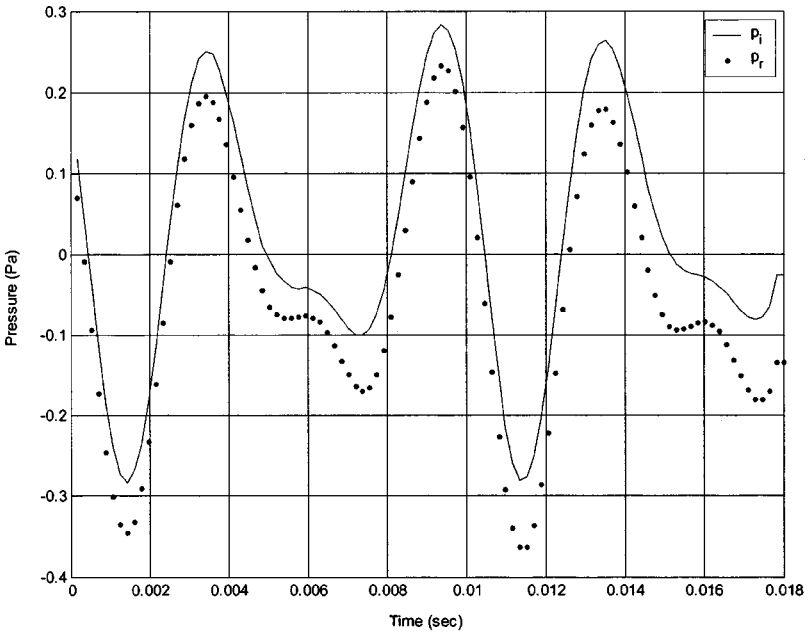


FIG. 10. Transient performance during reflection control for perfect absorption via integration method.



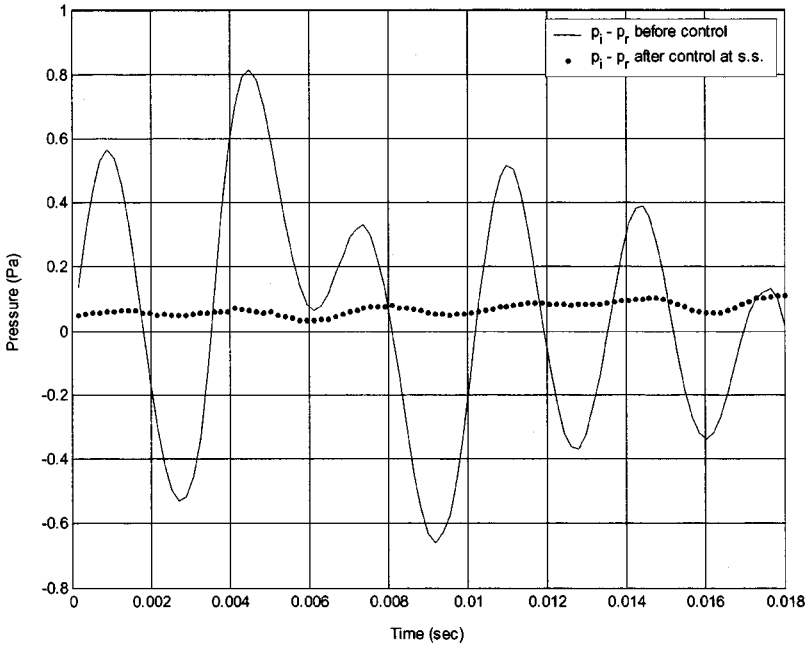


FIG. 11. Reflection control to achieve perfect reflection via integration method (primary noise consists of frequency components 200 and 300 Hz).

IV. ACTIVE CONTROL OF SOUND TRANSMISSION

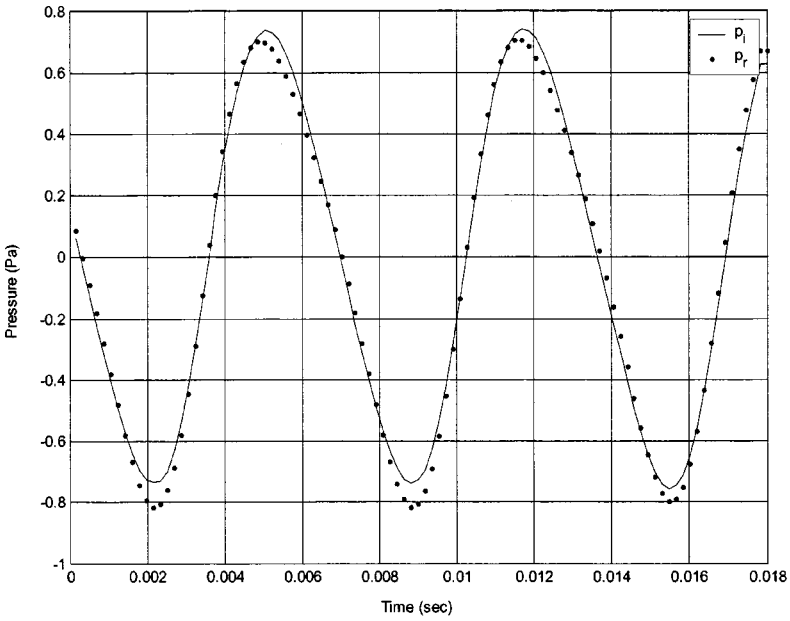
A. Experimental setup

The experimental setup for sound transmission control is shown in Fig. 7. This is very similar to that used in active reflection control (see Fig. 4). A 2-m-long duct is used to isolate the environmental effects and ensure plane waves. The cross section of the duct is 17 by 17 cm. Thus, its cutoff frequency is about 1000 Hz. An additional third microphone is placed behind the panel speaker to measure the residual sound pressure that will be used for feedforward control. The objective of the sound blocker is to drive the pressure at this residual microphone to zero.

B. Method

Several different panel materials including poster board and glass can be used as speakers once equipped with the small electromagnetic motor actuators. Thus, a glass pane with an electromagnetic motor actuator functions effectively as a panel speaker. The advantages of a panel speaker are that it is thin, space saving, and inexpensive. The disadvantages are that it has uneven frequency response and is only able to provide limited power. Since the panel will not be boxed in an enclosure, it will generate and propagate sound from both sides of its surfaces. All these factors were carefully considered in the control design.

As can be seen in the experimental setup, the two mi-



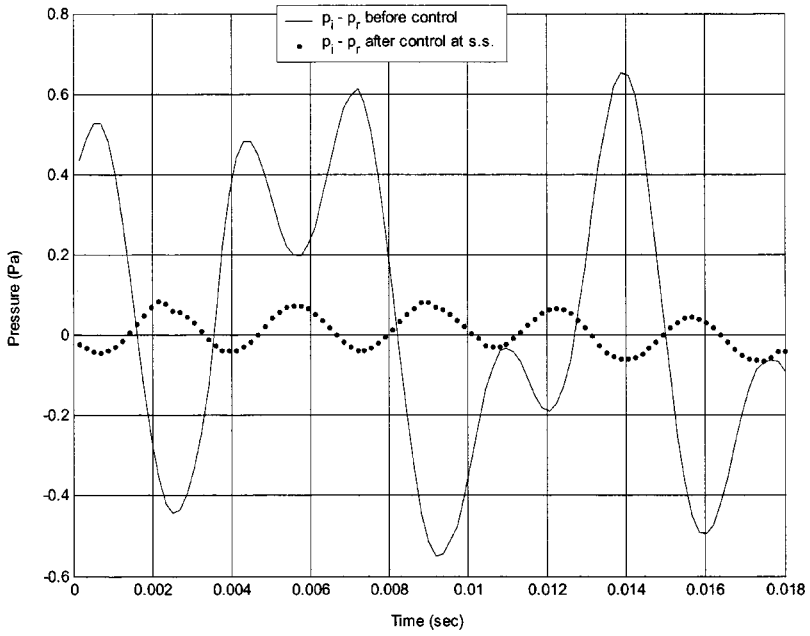


FIG. 12. Reflection control to achieve perfect reflection via integration method (primary noise consists of frequency components 150 and 300 Hz).

crophones in the path of the incident sound measure both the incident sound and the sound created from the panel itself. With the separation method, the incident wave is separated and used as a reference signal for feedforward control. The microphone at the other side of the panel measures the residual sound pressure which is then controlled to zero. A major distinguishing feature of the control system here is that a nonacoustic sensor is not needed for the reference signal. The incident sound is unaffected by the action of the speaker and hence we obtain a reference signal unaffected by the secondary source. All the signals from the microphones are sent to the PC via a data acquisition board. After wave separation, the reference signal is filtered by a FIR filter that represents the adaptive controller, the output signal is sent out to drive the panel speaker, and the signal from the error microphone is fed back to adapt the FIR filter coefficients.

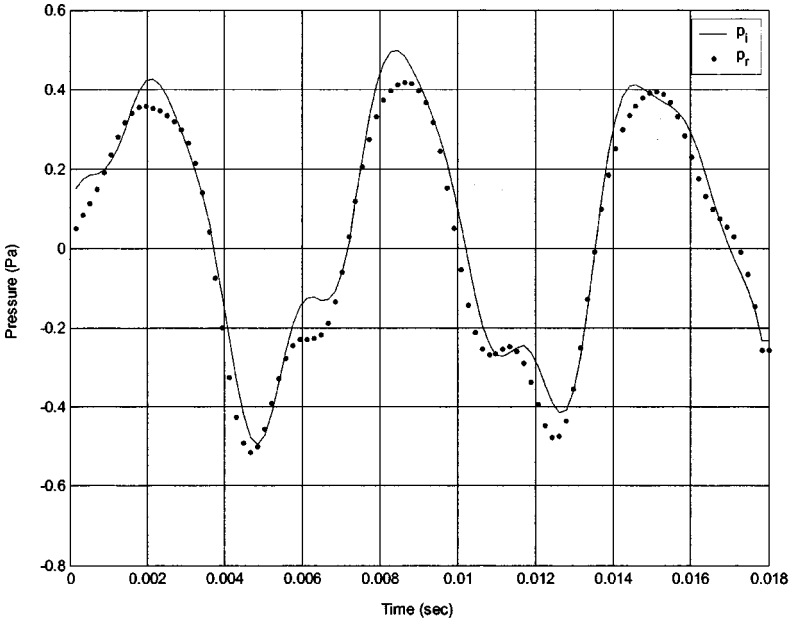
The algorithm used to drive the residual sound pressure to zero is the same FXLMS algorithm described in Sec. III D.

V. EXPERIMENTAL RESULTS ON REFLECTION CONTROL

A. Integration method

1. Perfect absorber

A panel behaves as a perfect absorber if $p_r=0$, i.e., if there is no sound reflected back. To achieve perfect absorption using active control, the value of p_r is controlled to zero using feedforward control. In the integration method of reflected sound estimation, the residual error is p_r. The secondary path is $S(s)=P_r(s)-R(s)P_i(s)$, as explained in Sec. III B.



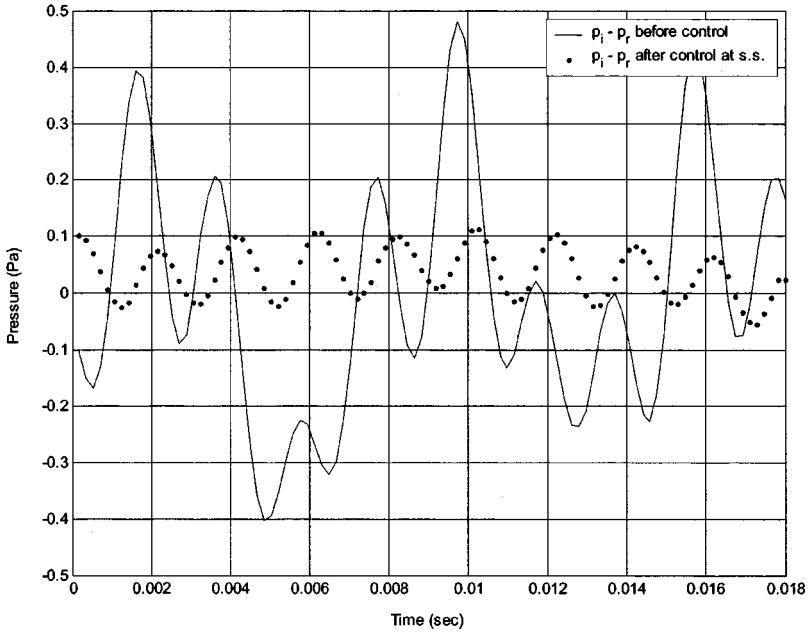


FIG. 13. Reflection control to achieve perfect reflection via integration method (primary noise consists of frequency components 150, 350, and 500 Hz).

Experimental results on reflection control to achieve perfect absorption are shown in Figs. 8, 9, and 10. The distance between the microphones in these experiments is 2.5 cm. Tonal noise is used since it is better illustrates reflection control, especially for cases where the desired reflection coefficient is nonzero. In the figures, the signals without any active control are compared with steady-state (s.s.) signals after control. The signals shown in the figures are the separated reflected waves in each case. In perfect absorption, there should be no reflection. The signals are measured indirectly via the PC. In Fig. 8 the primary noise consists of four frequency components (150, 200, 250, and 300 Hz), while in Fig. 9 the primary noise consists of two frequency components (100 and 300 Hz). As can be seen in the figures, there is a better than factor of 10 reduction in the reflected sound pressure. The transient performance of the controller is illustrated in Fig. 10. The control system has a time constant less than 0.5 s. The reflected sound pressure is seen to reach close to steady state in about 0.6 s.

2. Perfect reflector

A panel behaves as a perfect reflector if $p_r = p_i$, i.e., if all the incident sound is "reflected." To achieve perfect reflection, the reflected sound p_r is controlled to be equal to p_i using feedforward control. In the integration method, the secondary path is $S(s) = P_r(s) - R(s)P_i(s) = P_r(s) - P_i(s)$, as explained in Sec. III B. The residual error is just $p_r(t) - p_i(t)$.

The experimental performance of the reflection control system for perfect reflection is shown in Figs. 11, 12, 13, and 14. As can be seen, the control system ensures excellent

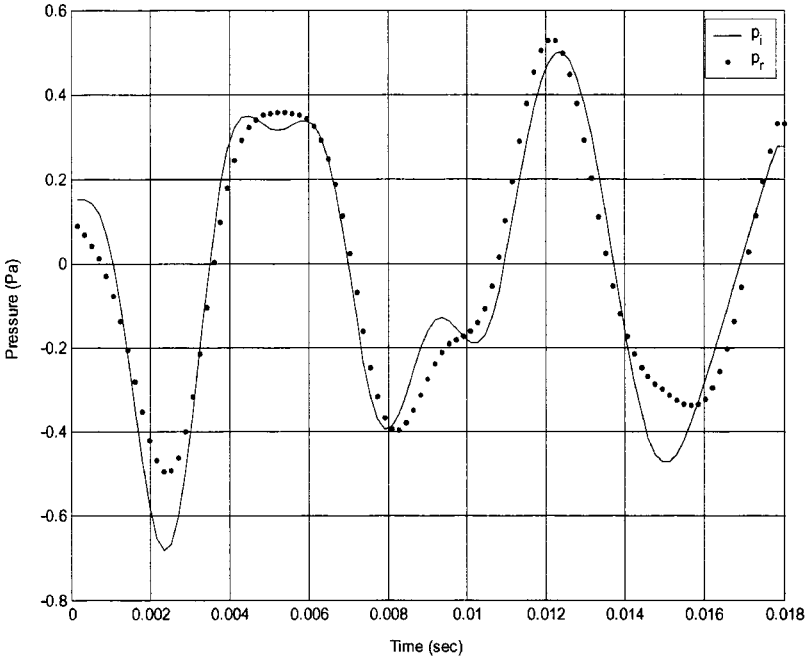
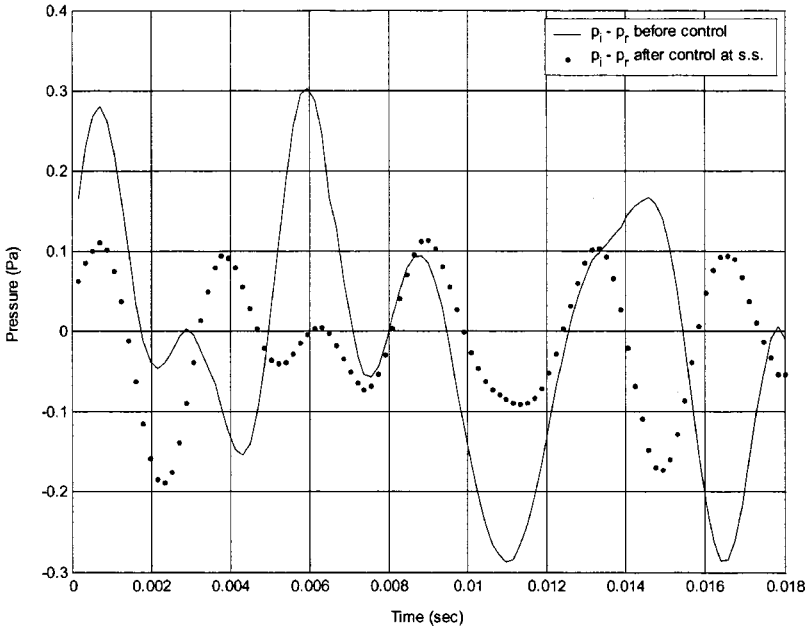


FIG. 14. Reflection control to achieve perfect reflection via integration method (primary noise consists of frequency components 150, 250, 350, and 400 Hz).



tracking between the incident sound and reflected sound waves. In every figure, there are two subfigures. The top subfigure shows the match between the incident wave and the reflected wave. The bottom subfigure shows the error between the incident and reflected waves to further illustrate the match. Multiple tones are used in each experiment. Different multiple tone combinations are shown in each figure. As can be seen in the figures, the performance tends to get worse when there are more frequencies contained in the primary noise.

3. Reflection coefficient greater than 1

In some applications such as in room acoustics, a coefficient greater than 1 may be desirable. In this section, experimental results on the case $R>1$ are shown to demonstrate the feasibility of achieving a reflection coefficient greater than 1 as long as the secondary source has adequate power. The integration method of wave separation is used.

Figure 15 shows the tonal case where the reflection coefficient is controlled to 1.2 for primary noise at a frequency of 200 Hz. As can be seen, the control system ensures good phase tracking between the incident sound and reflected sound waves with amplitude amplification. There are two subfigures in the figure—the left subfigure shows the match between the incident wave and the reflected wave and the right subfigure shows the error between the incident wave multiplied by 1.2 and the reflected wave to further illustrate the match. Reflection coefficient is controlled to 1.3 in Fig.

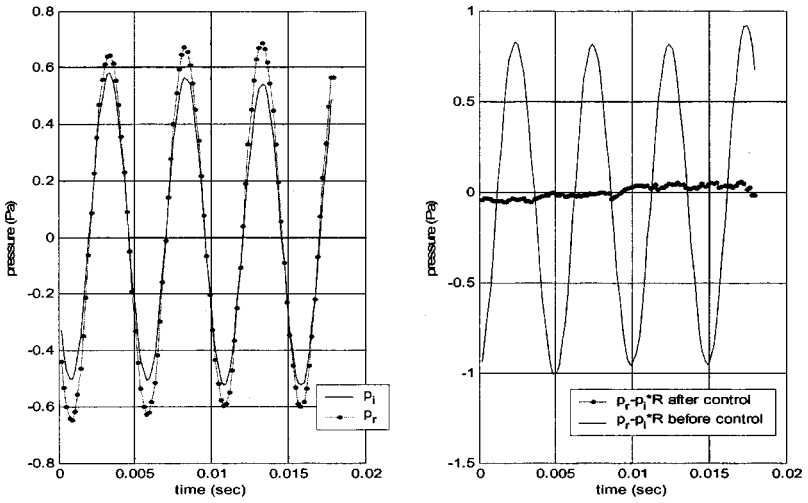


FIG. 15. Reflection control to achieve $R>1$ ($R=1.2$) via integration method (primary noise consists of frequency component 200 Hz).

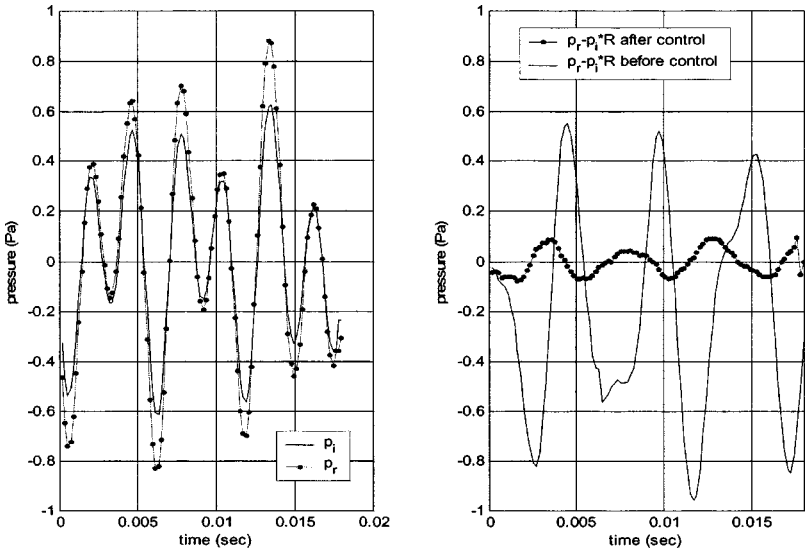


FIG. 16. Reflection control to achieve $R>1$ ($R=1.3$) via integration method (primary noise consists of frequency components 200 and 350 Hz).

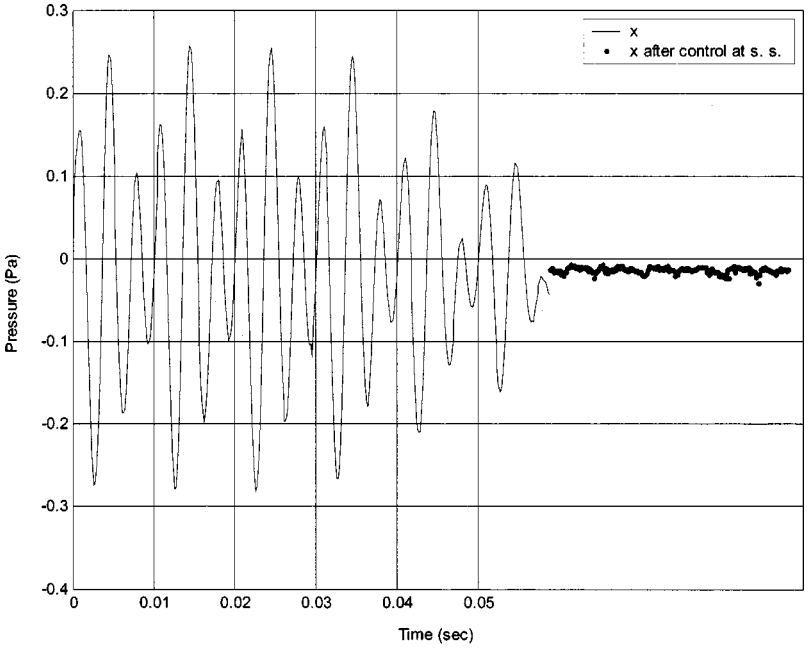


FIG. 17. Reflection control to achieve perfect absorption via delay method (primary noise consists of frequency components 200 and 300 Hz).

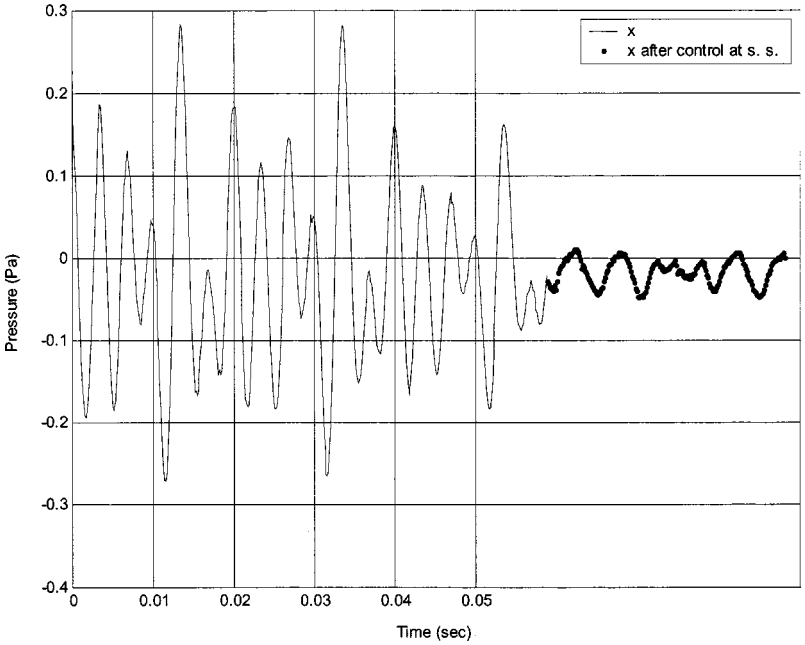


FIG. 18. Reflection control to achieve perfect absorption via delay method (primary noise consists of frequency components 150, 200, and 300 Hz).

16 with multiple tones at 200 and 350 Hz constituting the primary noise.

B. Delay method

1. Perfect absorber

In using the delay method for estimating the reflected and incident sound, the residual error signal and the secondary path change for the feedforward control system. The secondary path is $X(s)$, and the residual error is $x(t)$ as explained in Sec. III C. The objective here is to drive $x(t)$ to zero in order to get perfect absorption.

The experimental performance of the control system is shown in Figs. 17, 18 and 19. In the figures, the signals without any active control are compared with steady-state (s.s.) signals after control. The signals shown in the figures are "$x(t)$" in each case. In perfect absorption, $x(t)$ should be driven to zero. The signals are measured indirectly via the PC. In Fig. 17, the primary noise consists of two frequency components (200 and 300 Hz) while in Fig. 18, the primary noise consists of three frequency components (150, 200, and 300 Hz), and in Fig. 19, the primary noise consists of four frequency components (150, 200, 250, and 300 Hz). As can be seen in the figures, there is a better than factor of 10 reduction in the signal of "$x(t)$" in Fig. 17, and the performance gets slightly worse when the primary noise contains more frequencies.

2. Perfect reflector

For the perfect reflector case, again, the secondary path and the residual error are different compared to the integra-

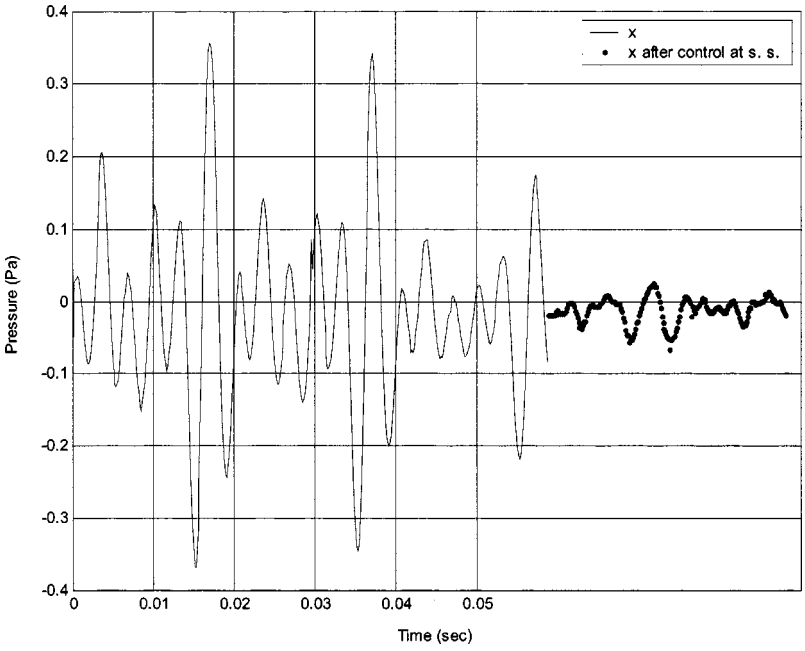


FIG. 19. Reflection control to achieve perfect absorption via delay method (primary noise consists of frequency components 150, 200, 250, and 300 Hz).

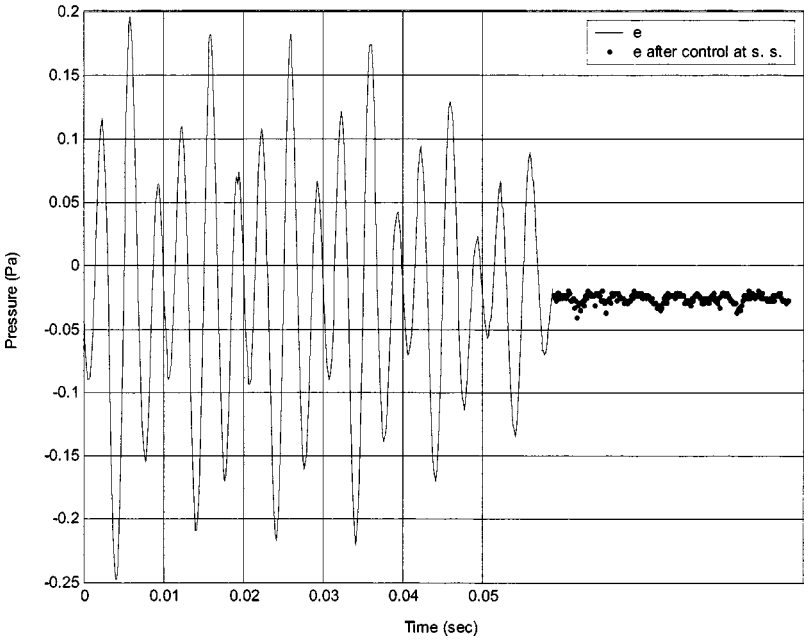


FIG. 20. Reflection control to achieve perfect reflection via delay method (primary noise consists of frequency components 200 and 300 Hz).

tion method. The secondary path is $X(s)-Y(s)e^{-(m+1)\tau s}$ and the residual error is $x(t)-y(t-(m+1)\tau)$. The objective here is to drive the conceptual residual error to zero in order to get perfect reflection.

The experimental performance of the system for achieving perfect reflection is shown in Figs. 20, 21, and 22. In the figures, the signals without any active control are compared with steady-state signals after control. The signals shown in the figures are the conceptual residual signal $x(t)-y(t-(m+1)\tau)$ in each case. In perfect reflection, the conceptual residual signal $x(t)-y(t-(m+1)\tau)$ should be driven to zero, i.e., $x(t)$ should match with a delay version of $y(t)$ in order to obtain $R=1$. The signals are measured indirectly via the PC. In Fig. 20, the primary noise consists of two frequency components (200 and 300 Hz) while in Fig. 21, the primary noise consists of three frequency components (150, 200, and 300 Hz), and in Fig. 22, the primary noise consists of four frequency components (150, 200, 250, and 300 Hz). As can be seen in the figures, there is a better than factor 10 of reduction in the signal of $x(t)-y(t-(m+1)\tau)$ in Fig. 20, and the performance gets slightly worse when the primary noise contains more frequencies.

C. Comparison between the two methods

The two methods for wave separation developed in this paper have their own advantages and disadvantages. The integration method requires the use of a high-pass filter to eliminate drift due to bias errors in acoustic pressure measurement. Also, the reflection coefficient calculated by the integration method has been found to have a dc error. However, this error remains constant and it is less than 1%. The

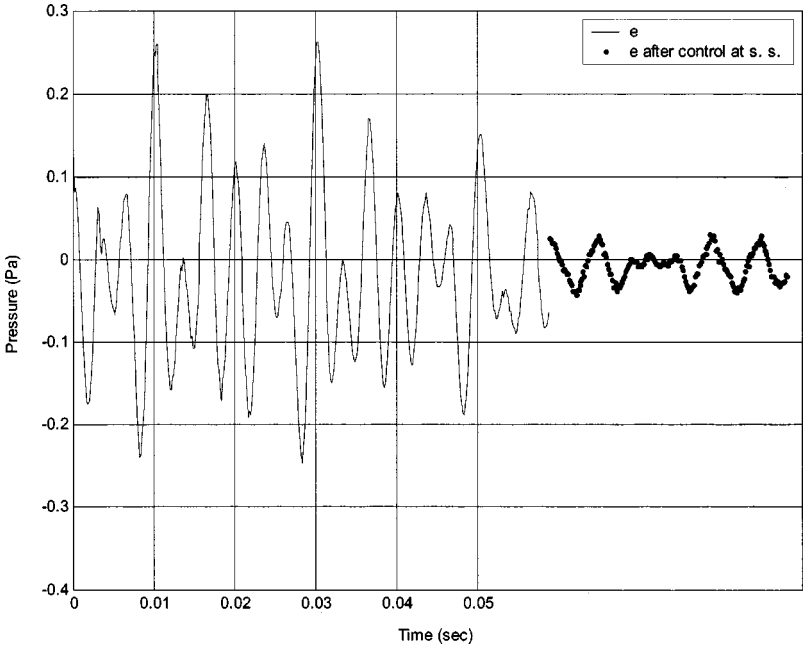


FIG. 21. Reflection control to achieve perfect reflection via delay method (primary noise consists of frequency components 150, 200, and 300 Hz).

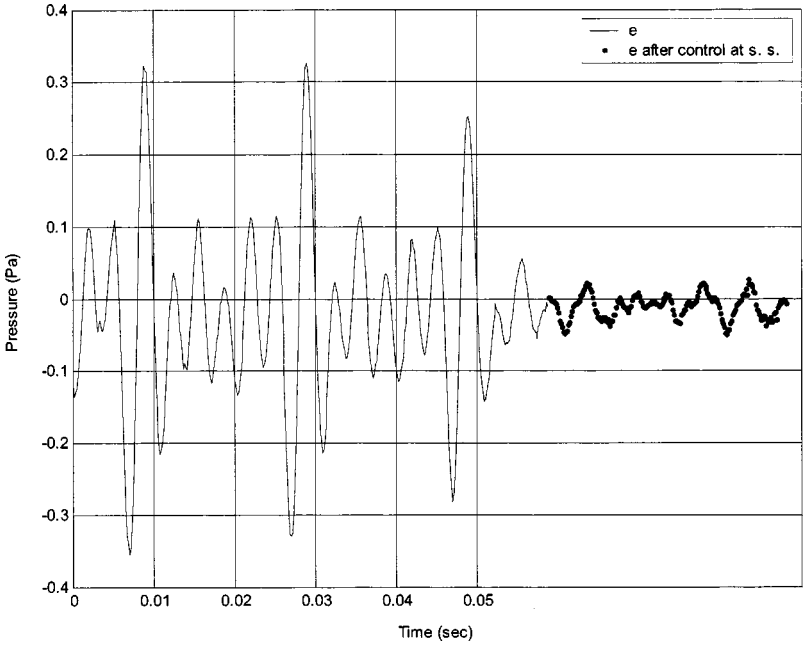


FIG. 22. Reflection control to achieve perfect reflection via delay method (primary noise consists of frequency components 150, 250, 200, and 300 Hz)

major disadvantage of the integration method is that it requires longer processing time due to the extra time needed for the integration as well as the high-pass filter operation. A big advantage of the integration method is that the incident wave is completely separated from the reflected wave. The incident wave can then serve as a reference signal for the FXLMS algorithm, and feedback effects from the secondary source can be successfully avoided. This advantage enables the method to also be used effectively in sound transmission control to obtain the reference signal.

From a faster processing time point of view, the delay method has an advantage. Only signals and their one-step delay versions are needed in this method. This speeds up the processing and a one-step delay is easy to implement in any DSP processor. The potential issues for the delay method that must be considered are that the magnitudes of x and y are small; thus, signal resolution is lost not only by the data-acquisition board but also during the DSP processing. Special care such as scaling may be needed in order to increase the signal-to-noise ratio. The largest disadvantage of the delay method is that both x and y contain information from both the incident wave and the reflected wave. The signal x cannot be used as a reference signal because it is not isolated from the actions of the secondary speaker. Unless a nonacoustic reference signal is available, the delay method is not suitable for broadband control. In this experiment, all the secondary path transfer functions were estimated off-line. As can be seen in Sec. III, for the case of perfect reflection control, the estimation of the secondary path transfer function for the delay method are more complicated than those for the integration method.

All the results shown in this section on reflection control

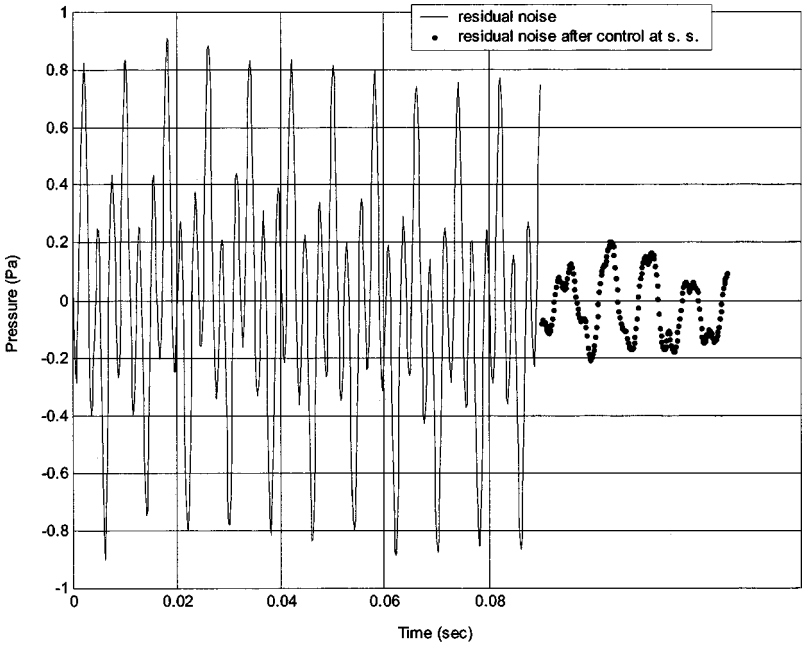


FIG. 23. Sound transmission control (primary noise consists of frequency components 125 and 375 Hz).

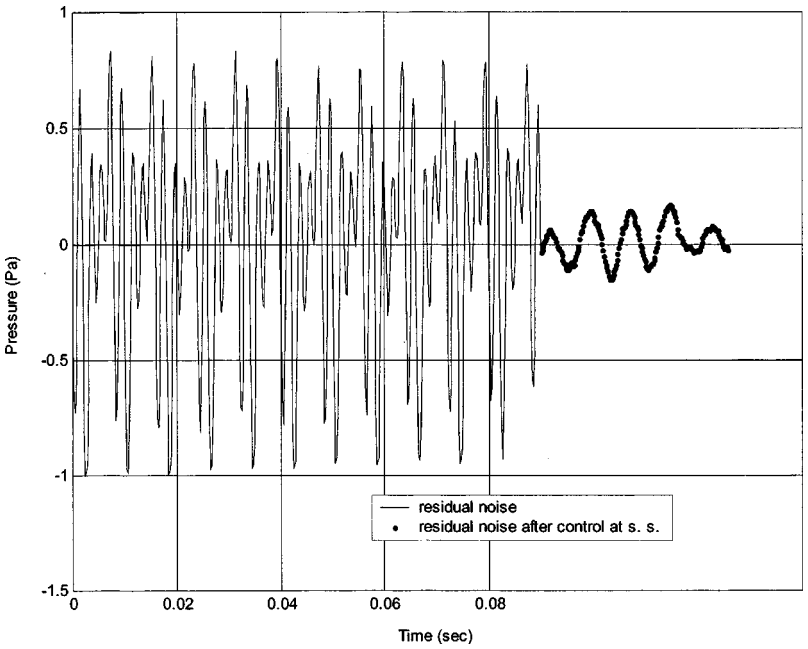


FIG. 24. Sound transmission control (primary noise consists of frequency components 125, 375, and 500 Hz).

are for harmonic noise. As can be seen in the figures, both methods work well. In the experiments, the sampling time while using the integration method was 180 microseconds while the sampling time via delay method was 140 microseconds. The time difference is expected to be larger if a microprocessor or DSP is used instead of a PC.

VI. EXPERIMENTAL RESULTS ON SOUND TRANSMISSION CONTROL

The experimental performance of the sound transmission controller described in Sec. III is shown in the figures below. In the experiment, a poster board is used as the panel. The secondary path transfer function is measured off-line. An order of 32 FIR filter is used to estimate this transfer function. Another order of 32 adaptive FIR filter is used for the FXLMS algorithm. The sampling time used is 180 microseconds.

Figures 23 and 24 show the performance of the transmission control system when the primary noise consists of discrete frequency components. Figures 25 and 26 show the performance when the primary noise consists of random noise bandlimited to frequencies below 800 Hz. In the figures, the signals without any active control are compared with steady-state signals after control. The signals shown in the figures are residual noise picked up by the third microphone positioned behind the panel in each case which should be driven to zero in order to block the sound transmission through the panel. In Fig. 23, the primary noise consists of two frequency components (125 and 375 Hz) while in Fig.

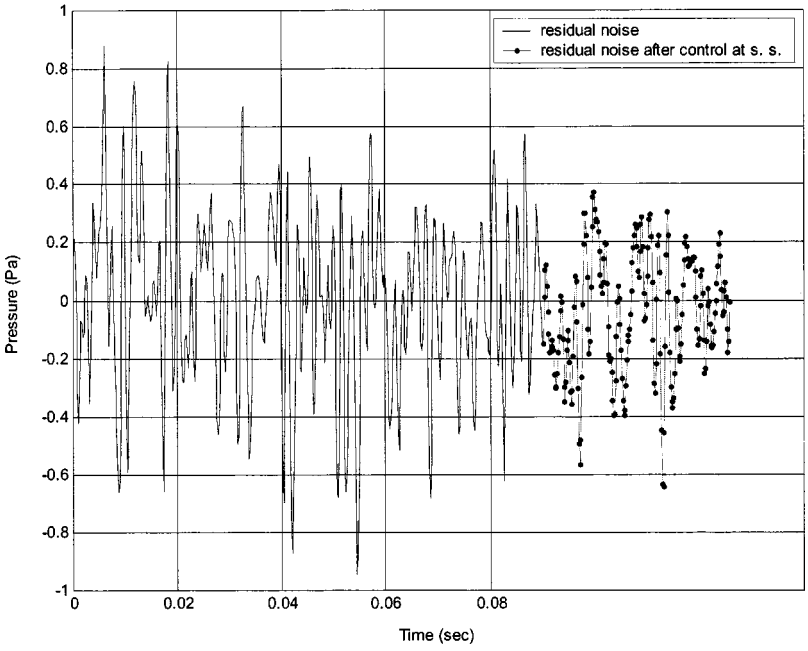


FIG. 25. Sound transmission control using a standard acoustic reference (primary noise is random containing frequency components up to 800 Hz).

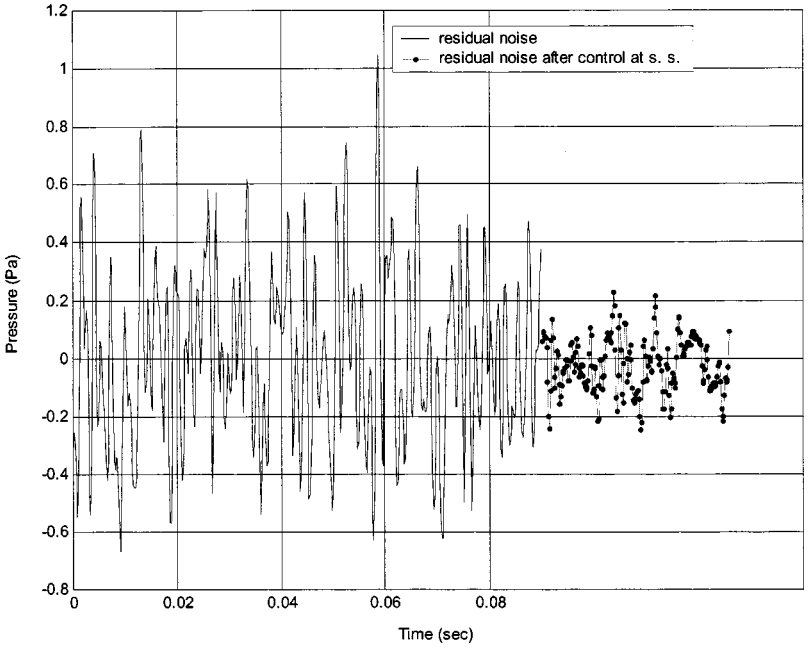


FIG. 26. Sound transmission using the separated incident sound as a reference (primary noise is random containing frequency components up to 800 Hz).

24, the primary noise consists of three frequency components (125, 375, and 500 Hz). As can be seen in the figures, there is a better than factor of 5 reduction in the residue noise for the case when the primary noise contains discrete frequency components, and the performance is worse when the primary noise contains random noise bandlimited to frequencies below 800 Hz.

In Fig. 25, a single standard microphone is used to pick up a signal which is then used as a reference. In Fig. 26, the wave separation method is used to separate the incident sound which is then used as the reference signal. As can be seen, the use of the incident sound as a reference provides significantly superior performance. This is also illustrated through a frequency response plot in Fig. 27. Figure 27 shows how the transmission control system provides significant noise attenuation over a broad range of frequencies. A comparison of performance obtained using just a single acoustic microphone to pick up the reference signal with the performance obtained when the incident sound is used as a reference is also shown. Clearly, the use of the incident sound as a reference provides superior performance. Overall, a performance of 10–15 dB is obtained over most of the frequency range via wave separation in Fig. 27.

A. Performance in terms of global sound attenuation

Previous results show that active sound transmission control at a point is possible. To make it practically useful, a global performance check is carried out too. The experimental setup is the same as in Fig. 7 except that an enclosure is connected to the duct. The enclosure is used to check the global noise reduction performance. The primary noise comes from the duct. It goes into the enclosure via the thin

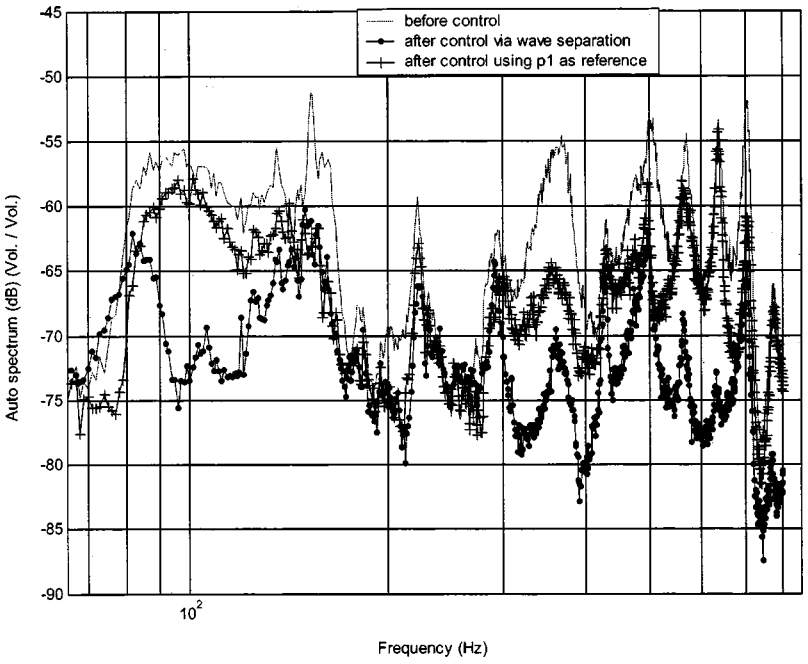


FIG. 27. Sound transmission control—frequency response (primary noise is random, containing frequency components up to 800 Hz).

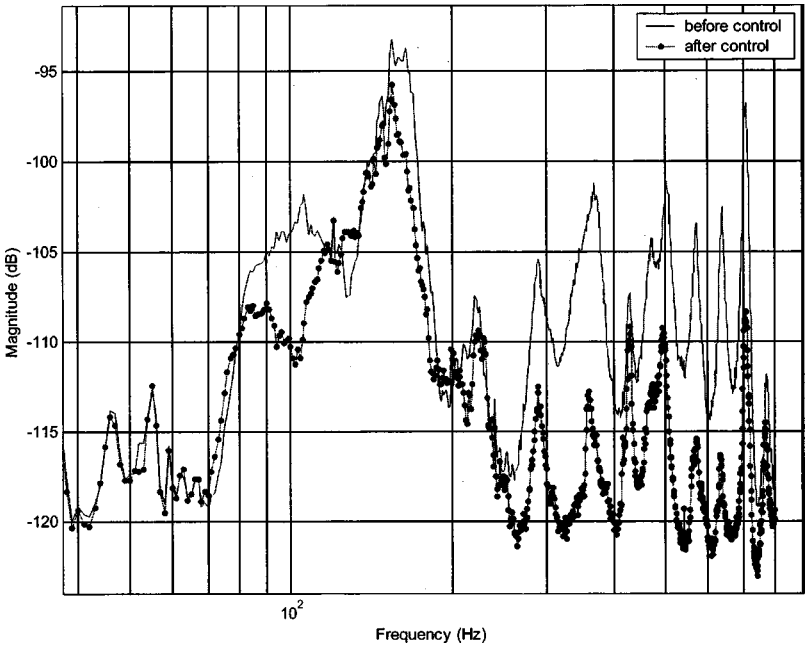


FIG. 28. Effect of the transmission controller on global sound in an enclosure.

panel. A cubic enclosure of size $0.4\times0.4\times0.4\ \mathrm{m}^3$ is considered. On the side of the enclosure that will be connected to the duct, a rectangular opening of the size of the duct is cut and a thin panel with an electromagnetic motor is mounted on this side of the enclosure. The primary noise is incident onto this thin panel. The panel is controlled so as to reduce noise transmission through the panel. Sound levels at different points inside the enclosure are then measured to evaluate if sound inside the entire enclosure is reduced by the use of this control system. Experimental results showed that the sound inside the enclosure was reduced everywhere. Figure 28 shows the sound pressure (averaged over 20 points) as a function of frequency before and after control. It shows that sound transmission is reduced at all frequencies by the control system. Figure 29 shows the performance at the error microphone. As noticed in the figures, there is coupling between the plate and the cavity. As shown in Fig. 29, the noise is dominated by an acoustic resonance at around 150 Hz. Altogether, measurements at 20 points were taken inside the enclosure. For every measurement, it was repeated (averaged) 50 times via the signal analyzer. No sound amplification at any frequency was found for every measurement, although the attenuation is not uniform throughout the enclosure.

VII. CONCLUSIONS

This paper explored the development of thin panels that could be controlled electronically so as to be either perfect reflectors or absorbers or acoustic transmission blockers. The

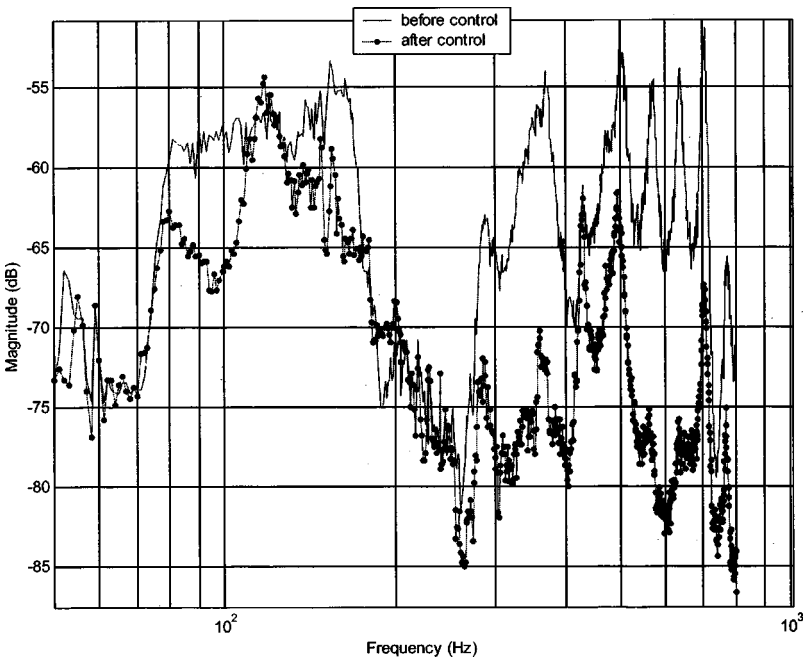


FIG. 29. Residual sound at the error microphone for the enclosure application.

panels were constructed using poster board and small rare-earth actuators. The development of the system was based on the use of a wave separation algorithm that separated incident sound from reflected sound. The reflected sound was then controlled to desired levels. The incident sound served an important purpose of providing an acoustic reference that is unaffected by the action of the control system speaker. The use of this incident signal reference also played a key role in the use of the panels as transmission blockers where the acoustic pressure behind the panel was driven to zero.

Detailed experimental results were presented showing the efficacy of the developed algorithms in achieving real-time control of reflection or transmission. The panels were able to effectively block transmission of broadband sound with the use of the incident wave playing a crucial role in allowing control without the use of a nonacoustic reference. The development of the panels is of practical importance with potential applications that include enclosures for noisy machinery, noise-absorbing wallpaper, the development of sound walls, and the development of noise-blocking glass windows. While the actuators will block light and are not transparent, they utilize rare-earth magnets and can be made with a diameter as small as 20 mm. These can be mounted near the corners of the window or hidden by patterned designs on the pane so as to be unobtrusive and therefore valuable for the glass windows application.

In the current research, only the normal incidence case has been investigated. When the incident angle is not normal, it is expected that the panel will continue to function well as long as the incident angle is not too oblique. The reflection coefficient at oblique incidence is given by

$$R = \frac{Z\cos(\theta) - Z_o}{Z\cos(\theta) + Z_o}, \tag{30}$$

where Z_o is the acoustic impedance of air, Z is the acoustic impedance at the panel surface, and θ is the angle of incidence. Thus, we can see that for small θ, the reflection coefficient is approximately equal to that at normal incidence.

In order to account for oblique incidence explicitly (if the angle of incidence were expected to be large), the relevant problem to be addressed is the development of the wave separation algorithm to reflect the change in incident angle. If the panel position is fixed and the incident angle is known, an extra microphone can be used to estimate the incident angle. If the incident angle is not known, an array of microphones around the panel would be needed for wave separation.

ACKNOWLEDGMENTS

This project was partially funded by Thermoking, Inc, Minneapolis, MN. The actuators for the panel speakers were provided by Kodel, Inc, Schaumburg, IL. The authors would like to express their appreciation to Steve Gleason and Jeff Berge from Thermoking and Mike Delazzer from Kodel.

Beranek, L. L. (**1954**). *Acoustics* (McGraw-Hill, New York), Chap. 2, pp. 16–46.

Chaplin, G. B. B. (**1977**). ''Active attenuation of recurring sound,'' U.S. Patent 4153815.

Furstoss, M., Thenail, D., and Galland, M. A. (**1997**). ''Surface impedance control for sound absorption: Direct and hybrid passive/active strategies,'' J. Sound Vib. **203**(2), 219–236.

Guicking, D., and Karcher, K. (**1984**). ''Active impedance control for one-dimensional sound,'' J. Vibr. Acoust. Stress Reliab. Des. **106**, 393–396.

Hansen, C. H. (**1991**). *Active Control of Noise and Vibration* (Department of Mechanical Engineering, University of Adelaide, South Australia), Chap. 1, Vol. 2.

Henrioulle, K., Desplentere, F., Hemschoote, D., and Sas, P. (**1999**). ''Design of an Active 1/4 Wavelength Resonance Absorber Using a Flat Loudspeaker,'' Active 99, pp. 1147–1158.

Kuo, S. M., and Morgan, D. R. (**1996**). *Active Noise Control Systems—Algorithms and DSP Implementations* (Wiley, New York), Chap. 1, pp. 1–16, Chap. 6, pp. 187–191, Chap. 3, 53–100.

Kuo, S. M., and Morgan, D. R. (**1999**). ''Active noise control: A tutorial review,'' Proc. IEEE **87**(6), 943–972.

Lacour, O., Galland, M. A., and Thenail, D. (**2000**). ''Preliminary experiments on noise reduction in cavities using active impedance changes,'' J. Sound Vib. **230**(1), 69–99.

Lueg, P. (**1936**). ''Processing of silencing sound oscillation,'' U.S. Patent 2043416, 9, June 1936.

Mehta, P. G., Zander, A. C., Patrick, W. P., and Zhang, Y. (**1998**). ''Active acoustic treatment (AAT)—a step toward a perfect sound absorber,'' in Proceedings of the American Control Conference, Philadelphia, PA, pp. 2611–2615.

Paurobally, R., Pan, J., and Bao, C. (**1999**). ''Feedback Control of Noise Transmission Through a Double-Panel Partition,'' Active 99, pp. 375–386.

Thenail, D., Galland, M., and Sunyach, M. (**1994**). ''Active enhancement of the absorbent properties of a porous material,'' Smart Mater. Struct. **3**, 18–25.

Thenail, D., Lacour, O., Galland, M. A., and Furstoss, M. (**1997**). ''The active control of wall impedance,'' Acustica **83**, 1039–1044.

Evaluation of the risk of noise-induced hearing loss among unscreened male industrial workers

Mary M. Prince[a)]
Industrywide Studies Branch, Division of Surveillance, Hazard Evaluations, and Field Studies, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226

Stephen J. Gilbert, Randall J. Smith, and Leslie T. Stayner
Risk Evaluation Branch, Education and Information Division, National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226

(Received 4 April 2001; revised 27 August 2002; accepted 8 November 2002)

Variability in background risk and distribution of various risk factors for hearing loss may explain some of the diversity in excess risk of noise-induced hearing loss (NIHL). This paper examines the impact of various risk factors on excess risk estimates of NIHL using data from the 1968–1972 NIOSH Occupational Noise and Hearing Survey (ONHS). Previous analyses of a subset of these data focused on 1172 highly ''screened'' workers. In the current analysis, an additional 894 white males (609 noise-exposed and 285 controls), who were excluded for various reasons (i.e., nonoccupational noise exposure, otologic or medical conditions affecting hearing, prior occupational noise exposure) have been added ($n=2066$) to assess excess risk of noise-induced material impairment in an unscreened population. Data are analyzed by age, duration of exposure, and sound level (8-h TWA) for four different definitions of noise-induced hearing impairment, defined as the binaural pure-tone average (PTA) hearing threshold level greater than 25 dB for the following frequencies: (a) 1–4 kHz (PTA_{1234}), (b) 1–3 kHz (PTA_{123}), (c) 0.5, 1, and 2 kHz (PTA_{512}), and (d) 3, 4, and 6 kHz (PTA_{346}). Results indicate that populations with higher background risks of hearing loss may show lower excess risks attributable to noise relative to highly screened populations. Estimates of lifetime excess risk of hearing impairment were found to be significantly different between screened and unscreened population for noise levels greater than 90 dBA. Predicted age-related risk of material hearing impairment in the ONHS unscreened population was similar to that predicted from Annex B and C of ANSI S3.44 for ages less than 60 years. Results underscore the importance of understanding differential risk patterns for hearing loss and the use of appropriate reference (control) populations when evaluating risk of noise-induced hearing impairment among contemporary industrial populations. [DOI: 10.1121/1.1536635]

PACS numbers: 43.50.Qp, 43.64.Wn [MRS]

I. INTRODUCTION

A. Background

Similar to many chronic diseases, occupational noise-induced hearing loss (NIHL) has a multi-factorial etiology. Risk factors found to explain most of the variability in hearing loss risk are increasing age and long-term exposure to continuous noise in occupational settings. There is also a large body of scientific literature suggesting that hearing loss risk among human populations may also depend on other endogenous and exogenous factors (Nakanishi *et al.*, 2000; Fechter, 1999) other than age and occupational noise exposure. Intrinsic factors, acting within the body to affect risk, include race, gender, and certain medical conditions (high blood pressure, diabetes, etc.) and family history (Jerger *et al.*, 1993; Gates *et al.*, 1993; Duck *et al.*, 1997; Brant *et al.*, 1996; Klein *et al.*, 2001; Melamed *et al.*, 2001). Conversely, exogenous factors include nonoccupational noise exposure (hunting, loud musical bands, and other loud hobbies), ototoxic chemicals, smoking, social class, education, and certain health-related behavioral factors [use of hearing protection devices (HPDs), work environmental factors, access to medical care].

One review of the controlled research concluded that the influence of many intrinsic variables is relatively small and cannot explain the wide range of hearing loss observed in epidemiologic studies (Henderson *et al.*, 1993). In another review of the literature, Ward (1995) concluded that susceptibility has not been clearly shown to be dependent on gender, skin color, any known diseases, mental attitude toward the noise, exposure history, or preexposure hearing loss. Ward further noted that it was possible that uncontrolled variables or unrecognized drug or chemical and noise interaction may obscure the relation between noise exposure and hearing loss. These intrinsic and exogenous risk factors can be difficult to control for in epidemiologic studies because of the inability to statistically separate the effects of highly correlated risk factors over time and/or the mediating effects of noise (either by decreasing or increasing susceptibility) on these factors through an intermediate causal pathway.

For example, consider the literature on hypertension and noise, an area with numerous research studies, which, on the

a)Electronic mail: mmp3@cdc.gov

whole, show contradictory and inconclusive results (Babisch, 1998; Thompson, 1983; Dijk, 1990; DeJoy, 1984; Passchier-Vermeer, 1993). Much of the literature in this area fails to adequately consider use of HPDs among industrial workers in both earlier and contemporary studies when HPD availability became more prevalent due to government and state regulations (Davis and Sieber, 1998; Royster and Royster, 1984). Proper use of HPDs by individuals exposed to noise is a confounding variable in assessing risk of hearing loss or hypertension due to noise exposure because it is associated with both the exposure of interest (noise) and the outcomes (hearing loss and possibly hypertension). It is also likely to be an effect modifier on the study results (Talbott *et al.*, 1996). In an experimental study (Ising *et al.*, 1980), blood pressure readings and urinary secretion of catecholmaines was lower on days workers wore hearing protection as compared to days they were unprotected during exposures of average noise levels of 95 dBA. The effective reduction in sound pressure level was measured to be between 10 and 16 dBA in this study. Hence, when HPD use is not accounted for in the analysis, misclassification of workers by noise exposure level may occur, thereby affecting inferences regarding exposure-response relationships.

B. Study relevance and purpose

Current standards (ANSI S3.44, 1996; ISO-1999, 1990; OSHA, 1983) examining the effects of noise on hearing loss are based on surveys conducted during the 1970s when hearing protection was not extensively used in general industry and workers were exposed to steady-state continuous noise environments (NIOSH, 1972; Baughn, 1966; Passchier-Vermeer, 1968; Burns and Robinson, 1970). The underlying data and models used to develop risk-damage criteria for these standards were based on highly ''screened'' noise-exposed and control (exposed to less than 80 dB daily 8-h average) populations. The ''screened'' populations excluded individuals with various risk factors (medical, otologic) associated with hearing loss and nonoccupational noise exposure (hunters, military service, prior job-related exposure, other sources of recreational noise). The exclusion of these workers was deemed necessary to mimic controlled laboratory experiments. However, with the advent of epidemiologic methods, sampling methods, and statistical models for analysis of population-based health data, risk factors associated with hearing loss and nonoccupational noise exposure are routinely adjusted for in analyses.

The results of analyses of 1172 (380 controls and 792 noise-exposed) white male workers, which represented a population screened to exclude otologic, medical and other sources of noise exposure, has been previously published (NIOSH, 1972; Lempert and Henderson, 1973; Prince *et al.*, 1997). However, the unscreened data that include the previously excluded white males have not been available for analysis until recently. Generalizing prior work (based on a highly screened population) to an unscreened population of workers would allow inferences to be drawn for working populations that are more representative of workers enrolled in industrial hearing conservation programs (HCPs).

The main objectives of this analysis are to examine (1) baseline risk of age-related hearing loss among unscreened low noise-exposed industrial workers; (2) the impact of common risk factors for hearing loss on excess risk estimates of NIHL; and (3) variability in excess risk among unscreened noise-exposed workers relative to the screened subpopulation. Excess risk estimates from this population are calculated to examine variability due to factors other than noise exposure on hearing threshold level. The risk profile of the unscreened population has been compared to those of the screened population in a previous paper (Prince, 2002).

In this paper, the ''screened'' population will refer to the original 1172 workers analyzed previously (NIOSH, 1972; Prince *et al.*, 1997), while those excluded from the original analysis are referred to as the ''excluded'' population ($N = 894$). The total population of workers examined in this analysis, formed by pooling the ''screened'' and ''excluded'' ONHS subpopulations, is referred to as the ''unscreened'' population.

II. METHODS

A comprehensive description of the study methods, population characteristics, and descriptive analysis of hearing levels and impairment rates by age and other risk factors are found in Prince (2002).

A. Data analysis

1. Outcome definition

NIOSH (1972) used the term ''material impairment'' to define its criteria for maximum acceptable hearing loss, and OSHA later used a slightly modified term, ''material impairment of hearing'' to define the same criteria (OSHA, 1983). In this context, a worker was considered to have a material impairment of hearing when his or her binaural pure-tone average at the audiometric frequencies 1, 2, and 3 kHz exceed 25 dB. NIOSH recently changed its definition of material impairment to include the frequencies 1, 2, 3, and 4 kHz in the binaural pure-tone average (NIOSH, 1998).

In this analysis, four definitions of noise-induced material impairment, defined as binaural pure tone averages (PTA) across the following frequencies, were examined: (a) PTA averaged over both ears for 0.5, 1, and 2 kHz (herein referred to as PTA_{512}); (b) PTA averaged over both ears 1, 2, and 3 kHz (PTA_{123}); (c) PTA averaged over both ears for 1, 2, 3, and 4 kHz (PTA_{1234}); and (d) PTA averaged over both ears for 3, 4, and 6 kHz (PTA_{346}).

2. Covariates

Variables such as age and 8-h time-weighted average (TWA) sound levels were examined as continuous factors while categories of duration of exposure (2–4, 5–10, >10 years) were defined in as in previous publications (Prince *et al.*, 1997; NIOSH, 1972) to facilitate comparison of results. Exclusion conditions were coded as originally described by Lempert and Henderson (1973) and later by Prince (2002) to include medical and otologic conditions, previous job noise exposure, and nonoccupational sources of noise [hunting, loud music, military noise exposure (weapon

TABLE I. Description of models examined in analysis.

Model no.	Parameters included	Description of logistic regression model
1	Intercept +Age (continuous) +(Duration (2–4 yrs.)×noise level)$^{\phi}$ +(Duration (5–10 yrs.)×noise level)$^{\phi}$ +(Duration (>10 yrs.)×noise level)$^{\phi}$ where ϕ=parameter describing shape of dose-response	Parameters describe effects due to age, duration, and noise exposure in the population and assume that risk is independent of screening status. This is the original model developed for 1172 screened ONHS population (Prince *et al.*, 1997).
2	Model 1 parameters +second intercept term for subgroup that failed screening procedure	Second intercept term represents an adjustment to baseline risk due to other risk factors. Model allows for an adjustment in the *intercept* (baseline risk) for subgroup that failed the screen (i.e., screened and unscreened ONHS have different intercepts). This model best described the relationship of risk of material hearing impairment for three definitions (PTA_{1234}, PTA_{123} and PTA_{346}).
3	Model 2 parameters +second age term for subgroup that failed screening procedure	Second age term represents an adjustment to the effect of age due to other risk factors. Model allows for an adjustment in the intercept (baseline risk) and *age* for subgroup that failed the screen [i.e., screened and unscreened ONHS have different intercepts (baseline risk) and age effects]. This model best described the relationship of risk of material hearing impairment for the PTA_{512} definition.
4	Model 3 parameters +second set of terms for each duration category (multiplied by noise level to denote noise dose) for subgroup that failed screening procedure	Second set of duration terms represents an adjustment to the effect of dose due to other risk factors. Model allows for an adjustment in the intercept (baseline risk), age, and *duration exposed* for subgroup that failed the screen. Screened and unscreened ONHS have different intercepts (baseline risk), age, and duration effects. The adjustments for duration were not necessary in describing the relationship of risk of material hearing impairment for any of the four definitions examined.

and nonweapon sources), and pretest noise]. For purposes of the risk evaluation, individuals with pretest noise were excluded from the analysis. To ensure adequate sample size for analysis, the other exclusion criteria were collapsed into two groups for risk evaluation: (a) medical conditions and (b) history of noise exposure (from nonoccupational sources and previously held jobs).

3. Research questions

The methods of analysis of these data are addressed within the context of these questions:

(1) Does background risk of material hearing impairment in a cross-sectional sample of workers depend on whether the population is screened for conditions not associated with occupational exposure?
(2) What is the impact of various factors on the excess risk of material hearing impairment?

To evaluate the first question, models allowing an overall adjustment for risk factors other than occupational noise exposure was applied to the unscreened (combined) sample of workers. The second question was evaluated using the model developed for the first question but with systematic inclusion of workers (with a given set of characteristics) to the screened population to examine the impact of various risk factors (i.e., medical conditions, history of noise exposure) on excess risk of noise-induced hearing impairment.

4. Statistical models

The quantitative relationship between material hearing impairment and the covariates (defined above) was modeled using logistic regression methods (Breslow and Day, 1980). These logistic regression models were fit using the nonlinear minimization (nlminb) routine in S-Plus (MathSoft, 1997). This routine was used instead of the usual logistic function to allow for nonlinearity associated with the shape of dose-response. Further details of model development have been previously published (Prince *et al.*, 1997) and statistical and technical details of the current model will be available in a future technical report. A qualitative description of the various models used in this analysis are shown in Table I. These models differs from previous analyses (Prince *et al.*, 1997; *Model 1* in Table I) in that separate parameters were added to test whether risk of material hearing impairment among the screened and excluded subpopulations differed with respect to baseline risk in the population, age, and category of duration of exposure. To examine the impact of different sets of risk factors for hearing loss [medical conditions, history of noise exposure prior to the study (past noise)] on population excess risk for a particular definition of impairment, an indicator for these risk factors was added to the model (Table I) to allow for different baseline risks, age, and duration effects between the screened and excluded subpopulations.

5. Excess risk estimation

Excess risk for a particular age is defined as the difference between the prevalence of material hearing impairment

TABLE II. Prevalence of material hearing impairment (25-dB fence) by age, exposure status, and PTA definition among the screened and unscreened ONHS subgroups.[a]

Age group in years	Material hearing impairment definition	% Prevalence (>25-dB fence) by subgroup [number/total at given age]											
		Controls						Exposed					
		Screened		Unscreened		Excluded		Screened		Unscreened		Excluded	
		%	No./Total	%	No./Total	%	No./Total	%	No./Total	%	No./Total	%	No./Total
<35	PTA_{512}	2.1	[4/188]	3.6	[10/276]	6.8	[6/88]	6.5	[19/291]	10.2	[47/459]	16.7	[28/168]
	PTA_{1234}	4.3	[8/188]	7.6	[21/276]	14.8	[13/88]	12.0	[35/291]	16.3	[75/459]	23.8	[40/168]
	PTA_{346}	9.0	[17/188]	16.3	[45/276]	31.8	[28/88]	22.7	[66/291]	29.0	[133/459]	39.9	[67/168]
35–49	PTA_{512}	5.5	[7/128]	7.9	[20/252]	10.5	[13/124]	14.3	[44/308]	20.5	[123/587]	28.3	[79/279]
	PTA_{1234}	19.5	[25/128]	25.8	[65/252]	32.3	[40/124]	36.0	[111/308]	42.0	[247/587]	48.7	[136/279]
	PTA_{346}	28.1	[36/128]	45.6	[115/252]	63.7	[79/124]	54.9	[169/308]	59.6	[350/587]	64.9	[181/279]
⩾50	PTA_{512}	17.2	[11/64]	25.2	[34/135]	32.4	[23/71]	42.5	[82/193]	40.7	[137/336]	38.5	[55/143]
	PTA_{1234}	35.9	[23/64]	51.1	[69/135]	64.8	[46/71]	60.6	[117/193]	61.1	[207/336]	62.9	[90/143]
	PTA_{346}	62.5	[40/64]	71.8	[97/135]	80.3	[57/71]	77.7	[150/193]	77.8	[260/336]	76.9	[110/143]

[a]Excluded workers with pretest noise. Subgroups refer to screening and exposure status.

among the noise-exposed population given exposure duration and the sound level, and the corresponding prevalence among controls. The excess risk associated with exposure to noise evaluated at a given age was estimated from logistic models using the following relationship:

$$\begin{aligned} \textit{Excess Risk} = \Pr[Y=1 \mid \text{age, duration,} \\ \text{and intensity of noise exposure}] \\ - \Pr[Y=1 \mid \text{age,control}], \end{aligned} \quad (1)$$

where $Y=1$ if material impairment of >25 dB is observed and $Y=0$ if material impairment of $\leqslant 25$ dB is observed.

Hence, background risk is assumed to be equivalent to the prevalence of age-related material hearing impairment. Correspondingly, excess risk is assumed to be equivalent to the increase in this background risk associated with occupational noise exposure (adjusting for other known risk factors in the population).

III. RESULTS

A. Impairment rates by age and exposure status

The distribution of impairment rates by age and exposure status (exposed or control) for different fences (>25 dB, >30 dB, >40 dB) and PTA definitions (PTA_{1234}, PTA_{123}, PTA_{346}, PTA_{512}) was examined for the screened and unscreened (combined) ONHS groups. Table II shows the prevalence using the 25 dB fence by age, exposure status, and subgroup. The prevalence of impairment was highest for the PTA_{346} definition and lowest for the PTA_{512} definition irrespective of the fence used. The differences in prevalence between the screened, excluded, and unscreened (combined) groups were most marked among controls. Across all age groups, the highest prevalence of impairment was observed among the excluded control population, followed by the unscreened population and the screened population (the lowest impairment rates). Among exposed subgroups, the difference between the populations becomes smaller with increasing age.

B. Risk evaluation

1. Choice of models

The results of model fits and statistical evaluation of nested models indicate differential risk profiles depend on the definition of material hearing impairment. For all definitions examined, separate parameters for the intercept were added to account for differing baseline risks of hearing impairment for the screened and excluded populations (Model 2, Table I). However, the addition of separate age slopes for the screened and excluded subpopulations was significant for the PTA_{512} definition, suggesting that the effect of age on risk of material hearing impairment for a screened population may differ from that of an unscreened population (Model 3, Table I). The observation that the slope for the effect of age was smaller for the excluded than the screened population can be explained by the fact that a larger percentage of the excluded population had already developed material hearing impairment, leaving a smaller fraction of the population at risk of subsequently developing hearing impairment as they age. Hence, all risk estimates and inferences made in this analysis are based on model 3 (Table I) for PTA_{512} and model 2 (Table I) for all other impairment definitions (PTA_{1234}, PTA_{123}, PTA_{346}).

2. Background risk of material hearing impairment

Fitted relationships of risk as a function of age among the controls and exposed groups for several definitions of material hearing impairment are shown in Table III. Examination of background risk is based on estimates of risk among controls, adjusting for separate age effects (for PTA_{512}) and separate intercept (for all definitions) between screened and excluded subpopulations. As shown in Table III, the excluded ONHS population had the highest background risk of material hearing impairment, followed by the unscreened group, and the screened population, which had the lowest background risk at all ages and for all definitions of material impairment. The implications of this difference in background risk of hearing impairment by age is evaluated with respect to noise exposure and duration of exposure.

TABLE III. Comparison of background risk of material impairment for screened, excluded, and unscreened (combined) ONHS population for different impairment definitions.

Definition	Population by screening status Screened ($n=1172$) Unscreened ($n=2045$)[a] Excluded ($n=873$)[a]	Background risk (%) Age 30 years	 Age 45 years	 Age 65 years
PTA_{512}	Screened (model 1)	1.9	7.46	34.6
	Unscreened (model 3)	5.2	12.6	37.4
	Excluded (model 3)	8.8	17.5	37.5
PTA_{123}	Screened (model 1)	2.9	10.0	38.7
	Unscreened (model 2)	7.3	18.6	48.3
	Excluded (model 2)	10.5	25.8	59.8
PTA_{1234}	Screened (model 1)	6.9	19.9	55.1
	Unscreened (model 2)	12.3	29.7	64.3
	Excluded (model 2)	17.4	39.5	74.6
PTA_{346}	Screened (model 1)	12.5	34.2	74.2
	Unscreened (model 2)	23.5	48.1	80.3
	Excluded (model 2)	32.8	60.8	87.8

[a]The unscreened and excluded populations omitted 21 workers with pretest noise. The estimates of background risk for the unscreened population are a weighted average of the screened and excluded population background risk estimates using weights of 1172 and 873, respectively. Estimates for the excluded population were based on fitting models 3 and 2 in Table I with 2045 workers and solving equations for parameters associated with failing the screen.

3. Excess risk of noise-induced hearing impairment by screening status

Figure 1 compares excess risk of material hearing impairment for three definitions (PTA_{346} PTA_{512}, PTA_{1234}) by sound level (dBA), age, and duration exposed (ages 30, 45, and 65 years with 2–4, 5–10, and >10 years, respectively) among the excluded and screened ONHS subpopulations. The curve of excess risk labeled ''combined'' is a weighted average of the screened and excluded curves with weights equal to the number of screened ($N=1172$) and excluded [$N=(984-21)=873$] people, respectively. Patterns of risk depended on age, duration exposed, population, and definition of impairment. Estimates of lifetime excess risk (age 65, duration >10 years) become less similar among the excluded and unscreened ''combined'' groups with the screened group showing somewhat higher excess risks for definitions that include the higher frequencies (PTA_{346}, PTA_{1234}). For PTA_{1234}, excess risks begin to differ by screening status at levels greater than 90 dBA, with differences being most marked at older ages, where the screened excess risks are the highest. A similar pattern is observed for PTA_{346} with no difference by screening status at age 30 and 2–4 years of exposure and slightly greater differences in the combined (unscreened) versus screened and excluded sub-

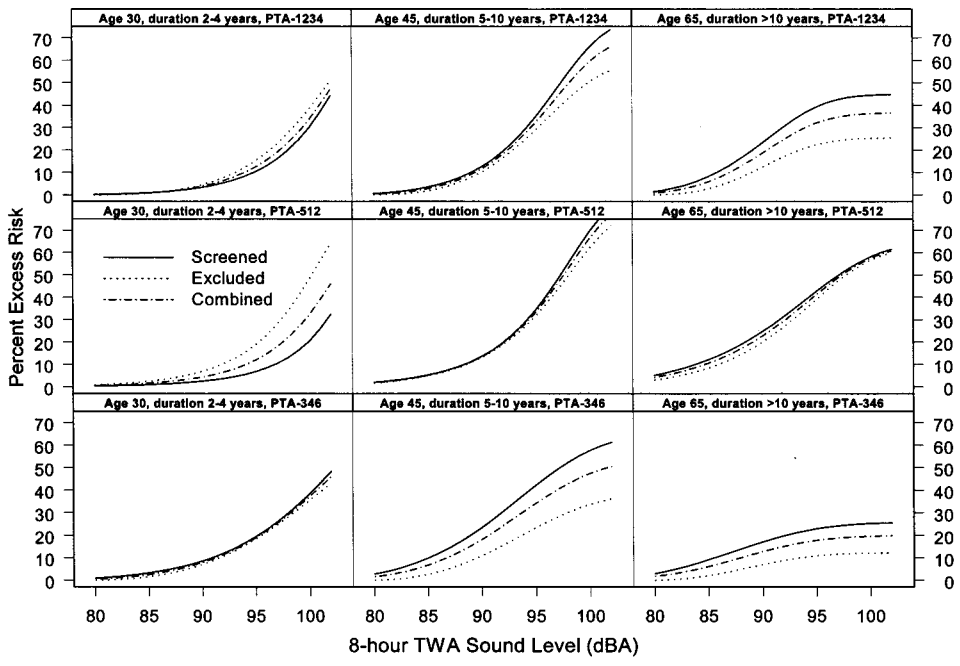


FIG. 1. Excess risk of material impairment by age, duration exposed, noise level, and population.

TABLE IV. Excess risk percent and 90% confidence limits for PTA-1234 definition of material hearing impairment among the unscreened and screened ONHS population.

8-h TWA sound level (dBA)	Unscreened population[a] Excess risk percent with 90% confidence limits: (lower, upper limits)		Screened population[b] Excess risk percent with 90% confidence limits: (upper, lower limits)	
Age 30, duration exposed 2–4 years				
80	0	(0.0, 0.4)	0.2	(0.0, 0.7)
85	0.5	(0.1, 1.5)	0.9	(0.2, 2.2)
90	3.3	(1.1, 5.8)	3.3	(0.9, 6.7)
95	11.5	(4.5, 19.2)	10.2	(2.9, 21.3)
100	32.0	(14.0, 54.8)	30.1	(8.5, 63.5)
Age 30, duration exposed 5–10 years				
80	0.0	(0.0, 0.5)	0.2	(0.0, 0.9)
85	0.8	(0.3, 2.1)	1.4	(0.4, 2.8)
90	5.0	(2.4, 7.9)	5.6	(2.6, 8.9)
95	18.7	(9.3, 29.5)	19.5	(10.2, 31.4)
100	51.1	(28.5, 78.0)	57.3	(30.3, 86.2)
Age 45, duration exposed >10 years				
80	0.0	(0.0, 2.0)	1.0	(0.0, 3.6)
85	2.9	(1.4, 7.2)	5.8	(2.2, 10.6)
90	18.2	(13.2, 23.7)	22.1	(14.5, 28.8)
95	50.9	(45.7, 57.4)	56.8	(49.4, 65.5)
100	68.8	(66.2, 72.0)	78.0	(73.1, 81.9)
Age 65, duration exposed >10 years				
80	0.0	(0.0, 2.1)	1.5	(0.0, 5.7)
85	3.1	(1.5, 7.7)	8.1	(2.9, 14.8)
90	15.5	(11.4, 20.5)	23.1	(15.1, 30.8)
95	30.5	(26.7, 35.9)	39.1	(32.0, 47.2)
100	35.4	(31.1, 41.2)	44.5	(36.2, 52.9)

[a]Excess risk estimates were based on model 2 in Table I and are a weighted average of screened and excluded subpopulation excess risk estimates using weights of 1172 and 873, respectively.
[b]Excess risk estimates were based on models described in Prince *et al.* (1997).

populations at older ages and long exposure durations. For PTA_{512}, differences are observed at age 30 with 2–4 years of exposure but not at older age groups (45 and 65 years) and longer durations of exposure (5–10 and >10 years).

Table IV shows excess risk estimates with 90% confidence limits for the PTA_{1234} definition among the screened and unscreened populations for different ages and duration categories. In general, the screened population has higher excess risks than the unscreened population, but these differences do not appear to be significant among younger workers (30 years of age) with short exposure durations 2–4 years, 5–10 years or among all workers with noise levels less than 95 dBA. The underlying model used in evaluating excess risk among the unscreened population assumed different background risk by screening status. Based on the 90% confidence limits for lifetime excess risk estimates (age 65, >10 years duration exposed), there appears to be significant differences in excess risk estimates between the screened and unscreened group for 8-h time-weighed average exposures greater than 90 dBA. Due to data sparseness for levels 85 dBA and lower, it is likely that increased variability and greater uncertainty in estimating excess risk would make it difficult to discern differences in risk at lower levels of exposure and among younger workers.

4. Effect of different risk factors on excess risk

The additional risk due to various risk factors among the ''screened'' population is shown in Fig. 2. The curves define the following populations:

(i) *Screened*—original 1172 screened workers with no known risk factors for hearing loss other than occupational noise.
(ii) *Past noise*—original 1172 screened workers plus those who failed screening because they had a history of noise exposure before the study.
(iii) *Medical*—original 1172 screened workers plus those who failed screening due to medical conditions that might affect risk of hearing loss.

Patterns of excess risk were similar for the PTA_{1234} and PTA_{123} definitions so results are only presented for three

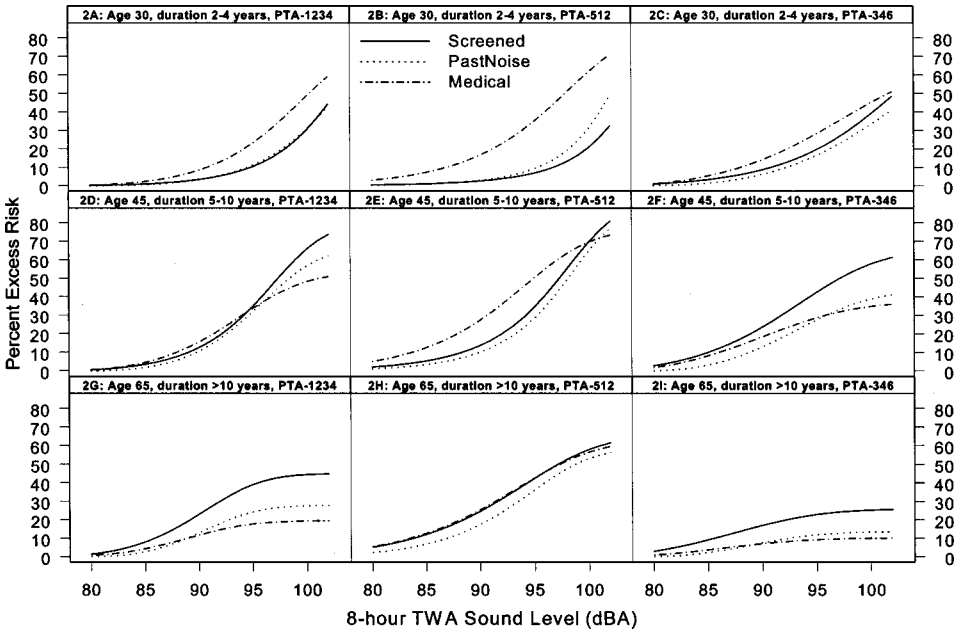


FIG. 2. Excess risk of material impairment for different definitions by age, duration, sound level, and exclusion criteria.

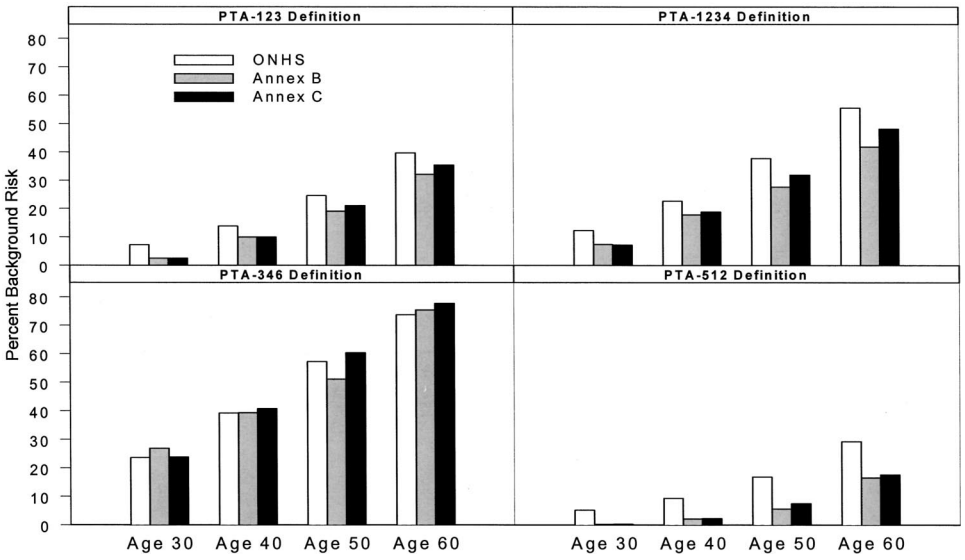


FIG. 3. Background risks by age and impairment definition for ONHS unscreened controls compared to Annex B and Annex C, ANSI S3.44 (1996).

definitions (PTA_{1234}, PTA_{512}, PTA_{346}). The screened population curve represents the dose-response associated with occupational noise. The curves associated with past noise and medical conditions among the unscreened population represent the additional risk over and above cumulative noise among the screened population for each risk factor examined. As shown in Fig. 2, the impact of certain risk factors depends on age and the definition of impairment used.

Among younger workers (age 30) with shorter exposure durations (2–4 years), background risk of hearing loss due to age is small [see Figs. 2(a)–(c)], and it is assumed that any observed excess is attributed to the extra risk associated with factors other than age. For these workers (age 30, 2–4 years exposed), the largest effect on excess risk appears to be medical conditions for all definitions. In contrast, the effect of past noise tends to lower excess risk of material impairment using the PTA_{346} definition, which includes the most sensitive hearing threshold frequencies for noise, whereas there is no effect on excess risk for the PTA_{1234} definition from this factor.

As the exposed population reaches middle age [see age 45, Figs. 2(d)–(f)], the effects on excess risk due to past noise are nominal for most exposures ranges using the PTA_{1234} and PTA_{512} definitions, while distinctly lower excess risks due to this factor are observed for the PTA_{346} definition. Compared to the screened population, medical conditions generally increased excess risk for the PTA_{512} definition (for levels below 100 dBA) and decreased excess risk associated with the PTA_{346} definition. The increased excess risk of hearing impairment for the PTA_{512} definition may be due to low-frequency hearing losses caused by medical conditions or unknown etiological factors other than noise and age. For estimates of lifetime excess risk [Figs. 2(g)–(i)], the effects of factors such as medical conditions and past noise exposure become nominal for the PTA_{512} definition. However, these risk factors tend to lower lifetime excess risks estimates when using definitions that include 3 or 4 kHz (PTA_{1234}, PTA_{346}).

Although data are not shown in this report, the effect of pretest noise on excess risk is noteworthy in that its impact on excess risk is greatest for the PTA_{346} definition. Since contamination of audiometric test results by pretest noise exposure might be assumed to occur periodically over time, its effect on risk estimates would affect the frequencies most sensitive to noise (i.e., 3, 4, and 6 kHz), resulting in artificially inflated excess risk estimates.

5. Comparison of ONHS background risk to ANSI S3.44 (Annex B and C)

The purpose of this analysis is to compare background risk of material impairment among controls from the unscreened (combined) ONHS data to Annex B and Annex C (ANSI S3.44-1996), which also represent unscreened populations. The background risk among controls from the ONHS unscreened population represent a ''weighted'' risk and is calculated as follows:

$$\text{Control (background) Risk} = [380(\text{Background Risk}_{\text{Passed}}) + 283(\text{Background Risk}_{\text{Failed}})]/665, \quad (2)$$

where

Passed = Screened population (380 is number of controls passing the screen),

Failed = Excluded population (283 is the number of controls who failed the screen).

As shown in Fig. 3, the risks generated from the ONHS control population are similar to those obtained from Annex B and C of ANSI S3.44 (1996), for the PTA_{1234}, PTA_{123} and PTA_{346} definitions for most age groups. Risks associated with the PTA_{512} definition are more variable by age with the ONHS population having higher risks than those generated from Annex B and Annex C of ANSI S3.44 (1996). The ONHS population risks are more similar to Annex C estimates of risk for most definitions, except PTA_{512} and PTA_{346} at age 50. For the PTA_{512} and PTA_{1234} definitions, there ap-

pears to be more divergence in risk estimates for ages 50 years and above. This is consistent with the predicted mean HTL generated from the ONHS data and the Annex B and C of ANSI S3.44 (Prince, 2002). The higher risks for the PTA_{512} definition among the ONHS population are likely due to artificially higher 0.5 kHz HTLs due to high background audiometric test booth levels. Other possible explanations for the difference in risk predicted from the ANSI S3.44 models and the ONHS unscreened data are discussed below.

IV. DISCUSSION

Variability in background risk patterns across populations may explain some of the diversity in risk due to noise. This analysis suggests that populations with lower background risks of material hearing impairment tend to have greater excess risks of noise-induced hearing impairment. We found that (a) background risk increases with increasing age and test frequency; (b) excess risk of noise-induced impairment was generally higher in the screened population than the unscreened population, especially at higher noise levels, longer durations of noise exposure, and older ages; and (c) the impact of medical conditions and past noise exposure affect baseline risk in the population but may not contribute greatly to lifetime excess risk of occupational noise exposure. Noise-induced damage to hearing is cumulative and increases with increasing duration of exposure and intensity. If background risk of hearing impairment is relatively high due to prior noise exposure or to other risk factors for hearing loss, the maximum excess risk due to subsequent noise should be smaller.

The modeling results support the conclusion that background risks among screened and unscreened populations are different for all definitions examined. The coefficients for the intercepts for the screened population were smaller than for the excluded population for most definitions except PTA_{512}. This suggests that the screened population background risk is lower than the unscreened population. In addition to different intercepts, the results for PTA_{512} also suggest that the excluded population has a smaller slope for age than the screened population. It is possible that the apparent diminished effect of age on excess risk may be due to a greater burden of background risk among workers in the excluded population at these lower frequencies. Similarly, the proportion of the total risk attributable to noise is less easily detected in the unscreened population. Although the magnitude of excess risk varies by whether a population is screened or not (i.e., has a smaller or larger burden of nonoccupational causes), the overall patterns of excess risk by sound level, age, duration exposed, and all definitions of material impairment are qualitatively similar for the screened and unscreened populations.

For definitions that include the frequencies most sensitive to noise damage (3, 4, and 6 kHz), a plateau in excess risk is observed after 10 years of exposure among workers older than 45 years, which is most marked at or above 95 dBA (Fig. 1). This plateau occurs because the expected proportion of the population with HTLs exceeding 25 dB (i.e., beginning material impairment) becomes relatively large as duration and intensity of sound exposure increases. Furthermore, the impact of pretest noise over time may artificially inflate excess risks with increasing age and duration of exposure.

These results underscore the importance of understanding differential risk patterns for hearing loss in working populations. The choice of reference populations for comparing risk of noise-induced hearing among industrial populations should carefully consider risk factors for hearing loss unrelated to noise exposure, other sources of noise exposure, or medical conditions associated with hearing loss. The dearth of such comparison populations for epidemiological studies of hearing loss underscores the need to utilize the data that is currently available. However, there are certain caveats to their use that should be considered.

A. Strengths and weaknesses of the underlying data and models

The current data analysis is limited to white males because raw data collected on white female workers has been lost. Summary hearing threshold level statistics for the female ONHS population are referenced in two government documents (NIOSH, 1972; OSHA, 1983). The models used in this analysis extend those used and developed for the screened ONHS population (Prince *et al.*, 1997) by evaluating differential risk patterns depending on screening status, age, duration of exposure, and baseline risk. Therefore, other functional forms for describing the relationship of noise and hearing loss risk in an unscreened population have not been considered or evaluated in this analysis. The outcome for analyses was defined as the probability of hearing loss greater than 25 dB, which limits exploration of whether the effect of noise on hearing loss becomes more or less severe as the threshold fence exceeds 25 dB. An alternative approach would be to express the outcome as a change in the distribution of hearing thresholds due to noise exposure (i.e., noise-induced permanent threshold shift or NIPTS) to explore implications for relevant etiology. For example, if a large proportion of the population has risk factors that interact synergistically with noise, thereby increasing susceptibility, the excess risk due to occupational noise among an unscreened population might increase in some cases. In addition, analysis of hearing threshold levels as a continuous variable in the model would provide more statistical power in evaluating whether the risk of crossing a 35- or 40-dB fence is as important as the risk associated with crossing a 25-dB fence. Nonetheless, hearing impairment definitions using a 25-dB fence remain valid hearing health outcomes for purposes of identifying risk of early damage to hearing when medical or public health interventions would be most likely to have the greatest impact on preventing moderate to severe hearing loss in the population.

1. Assessments of exposure and outcome

The underlying data were collected using audiometric testing procedures and noise measuring instruments available in the 1970s. Therefore, care must be taken when attempting to directly compare the distribution of noise exposure and hearing loss from the ONHS population to contemporary populations (post-1990).

An inherent assumption made in the evaluation of exposure-response relationships is that errors in outcome and exposure measures are nominal, or at best, distributed in such a manner that they do not bias estimates of effect. Exposure misclassification is therefore a valid concern in any analysis involving exposed human populations. In the ONHS survey, noise measurements were based on sound level meter readings (using state-of-the-art monitoring equipment of the 1970s) and involved taking representative samples of tasks within a job to calculate an 8-h time-weighted average (TWA) exposure. For some jobs, relatively short duration (about 10–15 min) samples were taken to estimate 8-h TWA noise levels. Such short sampling periods would be of concern if there is considerable job mobility and the number and types of tasks and noise levels vary on a daily basis. However, the ONHS study focused on sampling stable jobs with continuous, steady-state noise exposures. If there was any indication of high variability during the day, NIOSH investigators either observed the worker for longer periods or excluded the data from analysis (Lempert and Henderson, 1973). While some degree of exposure misclassification cannot be ruled out, such misclassification is expected to be small and effects on risk estimates would be limited.

Application of these models for populations exposed to intermittent or highly variable exposure conditions should be conducted with caution because the underlying data and models assume that workers stayed in the same job for the entire period of employment and that they were exposed to steady-state noise. Comparison of noise measurement data (8-h TWAs) from this study to contemporary populations using dosimetry-based 8-h TWA estimates should also be conducted with care due to differences in the precision of instrumentation over time (Earshen, 2000). Direct comparisons of risks across exposed populations from the same time period (ISO-1999, 1990; ANSI S3.44-1996, 1996) remain valid.

Audiometric testing was conducted using a mobile test booth in conformance with ANSI S3.1-1969 (ANSI, 1960) and manual audiometers that were calibrated under ANSI S3.6-1969 (ANSI, 1969). Comparison of hearing threshold data collected using automated audiometers and more recent ANSI calibration standards may require that adjustments be made to standardize measures when comparing audiometric data across study populations. Hearing threshold levels obtained by self-recording techniques are generally slightly better than those obtained with manual techniques (Burns and Hinchliffe, 1975; Knight, 1996; Harris, 1980). Nonetheless, risk estimates across populations can be validly compared if control and exposed populations are drawn from the same sample population (e.g., internal comparison group) and study procedures for measuring hearing and noise are consistent between exposure and control groups within the population.

B. Reference (control) populations: Implications for risk evaluation

There is increasing interest in assessing hearing change trends for groups of noise-exposed workers enrolled in industrial HCPs who are followed longitudinally over time in comparison to age-adjusted reference (control) populations. These populations can be compared to ANSI S3.44-1996 (ANSI, 1996) modeled predictions of hearing loss by age, to the ONHS control data, or to the Baltimore Longitudinal Study of Aging (BLSA) (Morrell *et al.*, 1991). If observed changes in hearing are no greater than that expected due to age, then it is often assumed that the program is effectively protecting workers. While this approach has the advantage of simplicity, further analysis would be necessary to identify whether observed differences (elevated HTLs above those expected from age) actually reflect variability due to factors other than a poor HCP. For example, if there are workers with intermittent exposure to noise, then multiple comparison populations may be examined, such as (a) nonindustrial control populations [Morrell *et al.*, 1991; Royster and Thomas, 1979; NCHS, 2001 (*future NHANES III adult audiometry data*)] or (b) unscreened industrial control populations (ANSI S3.44, 1996) as a means of defining a range of acceptable values from which HCP data may be compared.

V. CONCLUSIONS

Quantitative assessments of the relationship and magnitude of NIHL that use screened populations have some advantages including improved identification of susceptible worker populations, better estimation of their risks associated with noise, and a reduction of variability in non-noise related factors for hearing loss. However, unscreened populations can provide valuable information on the burden of hearing loss in the general population, particularly among non-noise exposed industrial populations. Moreover, use of unscreened low noise-exposed data are useful in comparing HCP databases, particularly if they can be drawn from populations likely to have similar risk factor profiles for other causes of hearing loss. For example, many chronic disease studies in human populations draw comparison populations from individuals living in the same geographical area as the exposed workers (i.e., community controls) or from workers in the same or similar plants having low or no exposure as a means for partially controlling for unknown risk factors that might increase the prevalence of disease in the population. Nevertheless, the present analysis suggests that differences in the distribution of background risk factors between screened and unscreened industrial populations should be controlled for in analyses of exposure-response relationships to avoid bias in risk estimates.

ACKNOWLEDGMENTS

The authors wish to thank the journal peer reviewers for their insightful review and comments. We also acknowledge the technical support of Bing Xue, Xiangdong Zhou, and Ruishan Wu in producing graphical displays of the data and to Ryan Elmore for his assistance in developing the statistical models for this paper.

ANSI (**1960**). ANSI S3.1-1960, ''American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms'' (American National Standards Institute, New York).

ANSI (**1969**). ANSI S3.44-1969, ''American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-induced Hearing Impairment'' (American National Standards Institute, New York).

ANSI (**1996**). ANSI S3.44-1996, ''American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-induced Hearing Impairment'' (American National Standards Institute, New York).
Babisch, W. (**1998**). ''Epidemiological studies of the cardiovascular effects of occupational noise—A critical appraisal,'' Noise Health **1**, 24–39.
Baughn, W. L. (**1966**). ''Noise control-percent of population protected,'' Int. Audiol. **5**, 331–338.
Brant, L. J., Gordon-Salant, S., Pearson, J. D., Klein, L. L., Morrell, C. H., Metter, E. J., and Fozard, J. L. (**1996**). ''Risk factors related to age-associated hearing loss in the speech frequencies,'' J. Am. Acad. Audiol. **7**, 152–60.
Breslow, N. E., and Day, N. E. (**1980**). ''Classical Methods of Analysis of Grouped Data,'' in *Statistical Methods in Cancer Research: Vol. I—The Analysis of Case-control Studies* (International Agency for Research on Cancer, Lyon, France), IARC Publication No. 32.
Burns, W., and Hinchcliffe, R. (**1975**). ''Comparison of auditory threshold as measured by individual pure-tone and Bekesy audiometry,'' J. Acoust. Soc. Am. **29**, 1274–1277.
Burns, W., and Robinson, D. W. (**1970**). *Hearing and Noise in Industry* (Her Majesty's Stationary Office, London).
Davis, R. R., and Sieber, W. K. (**1998**). ''Trends in Hearing Protector Usage in American Manufacturing from 1972–1989,'' Am. Ind. Hyg. Assoc. J. **59**, 715–722.
DeJoy, D. M. (**1984**). ''A report on the status of research on the cardiovascular effects of noise,'' Noise Control Eng. J. **23**, 32–39.
Dijk, F. J. H. v. (**1990**). ''Epidemiological research on non-audiory effects of occupational noise exposure since 1983,'' in *New Advances in Noise Research, part 1, Proceedings of the Fifth International Congress on Noise as a Public Health Problem*, Stockholm 1988, edited by B. Berglund and T. Lindvall (Swedish Council for Building Research, Stockholm), Vol. 4, pp. 285–292.
Duck, S. W., Prazma, J., Bennett, P. S., and Pillsbury, H. C. (**1997**). ''Interaction between hypertension and diabetes mellitus in the pathogenesis of sensorineural hearing loss,'' Laryngoscope **107**(12, Pt. 1), 1596–1605.
Earshen, J. J. (**2000**). ''Sound Measurement: Instrumentation and Noise Descriptors,'' in *Fifth Edition. The Noise Manual*, edited by E. H. Berger, L. H. Royster, J. D. Royster, D. P. Driscoll, and M. Layne (American Industrial Hygiene Association, Fairfax, VA), pp. 41–100.
Fechter, L. D. (**1999**). ''Mechanisms of Ototoxicity by Chemical Contaminants: Prospects for Intervention,'' Noise Health **2**, 10–24.
Gates, G. A., Cobb, J. L., D'Agostino, R. B., and Wolf, P. A. (**1993**). ''The relation of hearing in the elderly to the presence of cardiovascular disease and cardiovascular risk factors.'' Arch. Otolaryngol. Head Neck Surg. **119**(2), 156–161.
Harris, J. D. (**1980**). ''A comparison of computerized audiometry by ANSI. Bekesy Fixed-Frequency, and Modified ISO Procedures in an Industrial Hearing Conservation Program,'' J. Aud. Res. **20**, 143–167.
Henderson, D., Subramaniam, M., and Boettcher, F. A. (**1993**). ''Individual susceptibility to noise-induced hearing loss: an old topic revisited,'' Ear Hear. **93**(3), 152–168.
Ising, H., Günther, T., Havestadt, C., Krause, C., Markert, B., Melchert, H. U., Schoknecht, G., Thefeld, W., and Tietze, K. W. (**1980**). *Studies to quantify the risk of heart and circulatory disease in workers exposed to noise* (German), Budesanstalt für Arbeitschutz und Unfallforschung (ed), Report No. 225. Wirtschaftsverlag NW, Verlag für neue Werbung GmbH, Bremerhaven.
ISO 1999 (**1990**). ''Acoustics—Determination of occupational noise exposure and estimation of noise-induced hearing impairment,'' International Organization for Standardization.
Jerger, J., Chmiel, R., Stach, B., and Spretnjak, M. (**1993**). ''Gender affects audiometric shape in presbyacusis,'' J. Am. Acad. Audiol. **4**(1), 42–49.
Klein, B. E., Cruickshanks, K. J., Nondahl, D. M., Klein, R., and Dalton, D. S. (**2001**). ''Cataract and hearing loss in a population-based study: the Beaver Dam studies,'' Am. J. Ophthalmol. **132**(4), 537–543.
Knight, J. J. (**1996**). ''Normal hearing threshold determined by manual and self-recording techniques,'' J. Acoust. Soc. Am. **39**, 1184–1185.
Lempert, B. L., and Henderson, T. L. (**1973**). *Occupational Noise and Hearing 1968 to 1972: A NIOSH Study*, U.S. Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, National Institute for Occupational Safety and Health, Division of Laboratories and Criteria Development, Cincinnati, OH.
Melamed, S., Fried, Y., and Froom, P. (**2001**). ''The Interactive Effect of Chronic Exposure to Noise and Job Complexity on Changes in Blood Pressure and Job Satisfaction: A Longitudinal Study of Industrial Employees,'' J. Occup. Health Psychol. **6**(3), 182–195.
MathSoft (**1997**). S-PLUS 4 Guide to Statistics, Data Analysis Products Division, MathSoft, Seattle, WA.
Morrell, C. H., and Brant, L. J. (**1991**). ''Modeling hearing threshold levels in the elderly.'' Stat. Med. **10**, 1453–1464.
Nakaniski, N., Okamoto, M., Nakamura, K., Suzuki, K., and Tatara, K. (**2000**). ''Cigarette smoking and risk for hearing impairment: A longitudinal study in Japanese Male Office Workers,'' J. Occup. Environ. Med. **42**(11), 1045–1049.
NCHS (**2001**). National Health and Nutrition Examination Survey III (NHANES III). ''Audiometry/Tympanometry Procedures Manual,'' Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, MD, January 2001.
NIOSH (**1972**). ''NIOSH criteria for a recommended standard: occupational exposure to noise,'' U.S. Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, National Institute for Occupational Safety and Health, Cincinnati, OH, DHSS(NIOSH) Publication No. HIM 73-11001.
NIOSH (**1998**). ''NIOSH criteria for a recommended standard: occupational noise exposure, revised Criteria 1998,'' U.S. Department of Health and Human Services, Public Health Service, Center for Disease Control, National Institute for Occupational Safety and Health, Cincinnati, OH, DHSS(NIOSH) Publication No. 98-126.
OSHA (**1983**). ''Occupational Noise Exposure; Hearing Conservation Amendment; Final Rule,'' Occupational Safety and Health Administration, 290 CFR 1910.95; 48 Fed. Reg., pp. 9738–9785.
Passchier-Vermeer, W. (**1968**). ''Hearing loss due to exposure to steady-state broadband noise,'' Report No. 35 and Supplement to Report No. 35, Institute for Public Health Engineering, The Netherlands.
Passchier-Vermeer, W. (**1993**). *Noise and Health* (Health Council of the Netherlands, The Hague).
Prince, M. M. (**2002**). ''Distribution of risk factors for hearing loss: Implications for evaluating risk of occupational noise-induced hearing loss,'' J. Acoust. Soc. Am. **112**, 557–567.
Prince, M. M., Stayner, L. T., Smith, R. J., and Gilbert, S. J. (**1997**). ''A reexamination of risk estimates from the NIOSH Occupational Noise and Hearing Survey (ONHS),'' J. Acoust. Soc. Am. **101**, 950–963.
Royster, L. H., and Thomas, W. G. (**1979**). ''Age effect hearing levels for a white nonindustrial noise exposed population (ninep) and their use in evaluating industrial hearing conservation programs,'' Am. Ind. Hyg. Assoc. J. **40**, 504–511.
Royster, L. H., and Royster, J. D. (**1984**). ''Hearing Protection Utilization: Survey Results Across the USA,'' J. Acoust. Soc. Am. Suppl. 1 **76**, S43.
Talbott, E. O., Brink, L. L., Burks, C., Palmer, C., Engberg, R., Cioletti, M., and Inman, C. (**1996**). ''Occupational noise exposure, use of hearing protectors over time, and the risk of high blood pressure: the results of a case/control study,'' in *Proceedings of the 25th International Congress on Noise Control Engineering*, Liverpool, 1996, edited by F. A. Hill and R. Lawrence (Institute of Acoustics, St. Albans), pp. 2131–2137.
Thompson, S. J. (**1983**). ''Effects of noise on the cardiovascular system: appraisal of epidemiologic evidence,'' in *Proceedings of the Fourth International Congress on Noise as a Public Health Problem*, Turin, 1983, edited by G. Rossi (Centro Richerche E Studi Amplifon, Milanto), Vol. 1, pp. 711–714.
Ward, W. D. (**1995**). ''Endogenous factors related to susceptibility to damage from noise,'' Occup. Med. **10**(3), 561–575.

Anechoic chamber qualification: Traverse method, inverse square law analysis method, and nature of test signal

Kenneth A. Cunefare,[a] Van B. Biesel, John Tran, Ryan Rye, Aaron Graf, Mark Holdhusen, and Anne-Marie Albanese
The George W. Woodruff School of Mechanical Engineering, The Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 7 March 2002; revised 28 September 2002; accepted 14 October 2002)

Qualification of anechoic chambers is intended to demonstrate that the chamber supports the intended free-field environment within some permissible tolerance bounds. Key qualification issues include the method used to obtain traverse data, the analysis method for the data, and the use of pure tone or broadband noise as the chamber excitation signal. This paper evaluates the relative merits of continuous versus discrete traverses, of fixed versus optimal reference analysis of the traverse data, and of the use of pure tone versus broadband signals. The current practice of using widely space discrete sampling along a traverse is shown to inadequately sample the complexity of the sound field extant with pure tone traverses, but is suitable for broadband traverses. Continuous traverses, with spatial resolution on the order of 15% of the wavelength at the frequency of interest, are shown to be necessary to fully resolve the spatial complexity of pure tone qualifications. The use of an optimal reference method for computing the deviations from inverse square law is shown to significantly improve the apparent performance of the chamber for pure tone qualifications. Finally, the use of broadband noise as the test signal, as compared to pure tone traverses over the same span, is demonstrated to be a marginal indicator of chamber performance. [DOI: 10.1121/1.1527595]

PACS numbers: 43.55.Pe, 43.58.Vb [SLE]

I. INTRODUCTION

Anechoic chambers are now a common feature within many industrial, academic, and governmental laboratories. An informal survey of U.S.-based anechoic chamber manufacturers indicates that as of the time of this writing over 500 hemi- and full anechoic chambers have been built around the world since 1980, with the majority of the chambers going into industry facilities. Of critical concern to those who have acquired these facilities is the initial qualification of their free-field performance, that is, the determination of to what extent does the chamber yield a free field, and over what volume of the interior of the chamber. While there have been a number of methods proposed for the qualification of anechoic chambers,[1–5] and work in this regard continues today,[6] there is but one method that is embodied within ANSI and ISO standards. ANSI S12.35 and ISO 3745 define procedures for the qualification of anechoic chambers. The focus of the paper at hand is a detailed investigation into aspects of the implementation of an ANSI S12.35/ISO 3745 qualification test applied to an anechoic chamber, with attention paid to comparison between alternatives. While the subject matter may seem to be time sensitive to the standards of the day, the fundamental issue of demonstrating suitable inverse square law performance within an anechoic chamber is timeless (indeed, at the time of this writing the ISO 3745 standard is under revision, with many technical comments focused on the qualification issues alone).

The objective of qualifying an anechoic chamber is to determine the maximum allowable radius between a test source and a measurement location where inverse square law spreading is observed, within some tolerance, and where the location is not in the near field of the source. The process of qualifying the free-field performance of an anechoic chamber may be summed up through a paraphrased, slightly reworded segment of the ISO 3745/ANSI S12.35 standard: ''A microphone shall be moved continuously along radial paths and the sound pressure levels recorded. These levels are to be compared with the decay predicted by the inverse square law and the differences between the measured and theoretical levels.'' [7,8] These two sentences embody the very conundrum of free-field performance qualification: how does one actually sample the sound pressure levels along a path, and, then, with the data in hand, how does one analyze it? A related issue is the nature of the signal used in the qualification effort, i.e., pure tone versus broadband noise.

Elements of the procedure described in the standards (ANSI S12.35-1990 and ISO 3745:1977 are not materially different in their qualification procedures) include specifications and recommendations for the sound sources to be used in different frequency ranges, the test frequencies, the use of pure tone or broadband noise signals, the traverse directions, the requirement for a continuously moving microphone, and the permissible deviations from the inverse square law. Pass–fail determination is based upon permissible deviations from free-field performance, listed in Table I.

The generic procedure for a qualification test is to position a sound source within a chamber, and then to make acoustic pressure measurements along radials extending from the source. We will call such measurements along a

[a]Electronic mail: ken.cunefare@me.gatech.edu

TABLE I. Maximum allowable difference in anechoic rooms between measured and theoretical free-field levels per ISO 3745 and ANSI S12.35.

One-third octave band center frequency (Hz)	Allowable difference (dB)
<630	±1.5
800 to 5000	±1.0
>6300	±1.5

radial a traverse. We classify traverses as being either continuous or discrete. A continuous traverse is obtained when one continuously moves a microphone along a radial, continuously recording data as the microphone moves. A continuous traverse generally yields pressure measurements that represent a very fine spatial resolution along the traverse (typically, continuous traverse data is averaged over some span, converting a continuous data record to a sequence of finely spaced discrete samples). In contrast, a discrete traverse is obtained when one moves a microphone in discrete steps along a radial, with the microphone motionless during data acquisition at each microphone location. Discrete traverses on a fine spacing are very time consuming, such that the usual practice is to employ a spacing that is quite coarse as compared to what is attainable through the use of a continuous traverse. As of this writing, the common practice is to employ discrete traverses. This practice notwithstanding, ANSI S12.35 and ISO 3745 both call for a continuously moving microphone, that is, a continuous traverse.

Once the pressure data have been obtained, the next issue of concern becomes how to compare the measured levels' decay with distance against a theoretical free-field decay. To compute a theoretical free-field decay, one needs to know the separation distance between the source and the observation point. However, the acoustic center for the test sound source may not be known *a priori*, which introduces significant uncertainty and potential for error into the computation of the predicted decay. The current standards lack specificity on the issues of source center and on the actual method of performing the comparison to free-field decay. The common practice is to base the theoretical free-field decay upon the sound pressure level at a fixed reference distance (typically 3 ft) from the source.

With the data acquired and analyzed, it is then compared to the values in Table I in the following manner. For each traverse and each frequency, there will generally be a location along each of the traverses where the data *first* exceeds the tolerance limits of Table I. The *minimum* distance to the locations of these exceedances establishes the maximum measurement radius. The purpose of a qualification test is to find the minimum distance along any traverse at which the tolerance limits of Table I are exceeded.

The need to qualify the free-field performance of an anechoic chamber was recognized from the inception of these facilities,[1,2,9–12] with the performance assessed by measuring the spreading loss away from a sound source. Much of the qualification data presented in the literature has been based on discrete traverses.[1,2,9–15] The discrete traverse method generally uses measurements at 0.5- to 1-ft spacing along the traverse. Some more recent discrete traverses have employed finer spacings,[6,15] but, nonetheless, they still represent a coarse sampling of the test sound field. In contrast, there have been only a limited number of publications that document implementations of continuous traverses, and, interestingly, most such are earlier work. Rivin[16] developed an electro-mechanical method of performing a continuous traverse, with the deviation from free-field performance recorded directly upon a chart recorder. Ingerslev *et al.*[17] and Bell *et al.*[3] employed Rivin's method in their own qualification efforts. In contrast to the discrete traverse methods, the continuous traverses of Rivin, Ingerslev, and Bell clearly indicate a great deal more structure and complexity in the sound field generated by the test source. Indeed, in examining their traverse data, it is not difficult to realize that discrete traverse methods will miss much of the structure of the sound field, particularly at higher frequencies, though this point was not explored in these publications. Further, it is not difficult to conclude that a discrete traverse could easily miss regions of a traverse where the chamber's free-field performance is unacceptable, an issue equally unexplored. Finally, the literature considered here does not contain comparisons of traverses at different resolutions, nor attempts to substantiate a minimum resolution.

The issue of appropriate spatial resolution for traverses has been recently reinforced through Duda's[18] numerical modeling of the wave field due to pure tone and broadband sources within an anechoic chamber undergoing a qualification test. Duda used a method-of-images approach to model the wave field within an example anechoic chamber, approximating the bounding surfaces of the room as planes. By comparing numerically modeled traverses sampled at different resolutions, he demonstrated that for a pure tone test signal, the wave field within an anechoic chamber will be more complex than can be adequately sampled using discrete traverses at 1-ft spacings. While Duda did not argue for the use of continuous traverses (reflecting a concern for the difficulty of setting up such systems), he did argue for the need to sample at much finer spacings than 1 ft, particularly above 1 kHz. Duda touched upon the impact of using broadband noise versus pure tone signals for testing chamber performance, and demonstrated that pure tones provide the most stringent test of performance. Duda's analysis paralleled the earlier work of Wang and Cai,[19] however, Wang and Cai's use of the analysis was for modeling of free-field deviation, and did not address the issue of spatial resolution.

There are two common approaches documented in the literature for comparing acquired traverse data to a theoretical decay. The most common approach, the fixed reference method, uses the measured level at a given reference position from the source and applies the inverse square law to compute levels at other distances.[2,12,20–22] In contrast, the second approach, the optimal reference method, seeks to estimate a source strength and source center offset location that yields a theoretical decay that matches the observed decay in some optimal sense. This method is motivated by recognition that the effective acoustic center of a sound source may not coincide with a visually identifiable point on or near the source. Such position uncertainty makes the measurement of separation distance between the source center and an observation

point problematic. To ameliorate this issue, investigators have used either analog compensation methods[3,16,17] or mathematical compensation[13,15] to determine an effective source center position. The analog compensation method used clever electronics to generate strip-chart recordings of the deviation from inverse square performance, with the source offset introduced by a resistance. The mathematical compensation method uses a linear least square fit of traverse data to an inverse square model, computing both an effective source position and a source strength. (Note that Ballagh[23] developed an optimal reference method using a nonlinear, logarithmic fitting technique, but it has not gained widespread acceptance.) We are unaware of published comparisons between the fixed reference method and the optimal reference method.

Finally, as to the nature of the test signal to use, it appears that pure tone signals have been used for most published qualifications. However, the current and draft ISO standards permit the use of broadband noise sources, with the draft structured to specify broadband noise as the preferred signal. Both Wang and Cai[19] and Duda[18] demonstrated numerically that broadband noise traverses should show much less spatial structure than pure tone traverses. We are unaware of published comparisons between actual traverses obtained with these signal types.

The above issues, how the microphone is moved along a traverse, how one computes the theoretical free-field decay for a traverse, and whether to use pure tone or broadband noise signals, are the fundamental issues of interest in the paper at hand. The question of spatial resolution is addressed by considering traverses sampled at alternative spacings. The corollary issue, what spatial resolution is appropriate, is addressed through consideration of the structure of traverses obtained with pure-tone and broadband noise. The effect of the data analysis method is addressed by applying each method to the same data and assessing their impact. Finally, the significance of the choice of test signal is brought out by consideration of traverses over the same physical span but acquired separately with pure tone and broadband noise signals. Other issues relevant to chamber qualification, including test source design and the number of traverses to perform, are beyond the scope of the present paper.

In the following, we present our implementation of a continuous traversing microphone system. We then present the analysis procedures that we use for analyzing the resulting data. Finally, we present and discuss the results obtained, with the focus being on addressing the issues raised above, when these tools were applied to the qualification of an anechoic chamber at the Georgia Institute of Technology, the configuration of which we describe below.

A. Chamber description

Figure 1 provides a simple depiction of the major elements and components of the anechoic chamber evaluated here. The chamber is isolated from and is free-standing within an enclosing building. The chamber uses a proprietary commercial perforated metal wedge for the anechoic treatment.[24] The outer surface of the wedges is 42% open area perforated metal sheet: the inner surface of the perforate

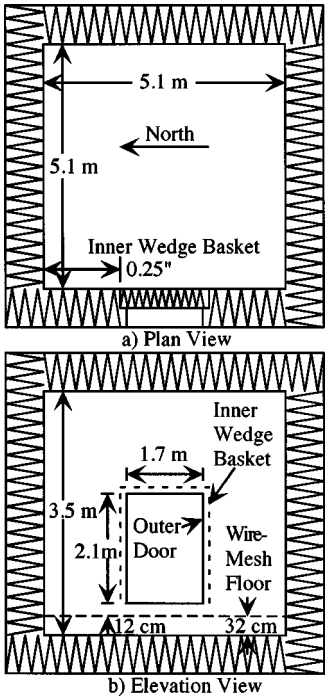


FIG. 1. Georgia Tech Anechoic Chamber, plan view (a) and elevation (b). 90-m^3 interior volume, entrance on west wall, wire-mesh floor.

is lined with a fiberglass layer. The chamber volume within the wedge tips is approximately 90 m^3. The design frequency range of the chamber is 80 to 10 000 Hz. Mounts for test samples and fixtures extend from the floor and ceiling. An edge-supported wire mesh floor is suspended 32 cm above the floor wedge tips.

II. ANALYSIS METHODS

As with previous researchers, we find that a plot of the deviation from the free-field decay versus linear distance from the source is most effective for assessing free-field performance.[3,15–17] The deviation from the inverse square law is

$$\Delta L_{pi} = L_{pi} - L_p(r_i), \tag{1}$$

where L_{pi} is the sound pressure level at each measurement position, and $L_{p(r_i)}$ is the level at distance r_i estimated by either the fixed reference or optimal reference method. With the fixed reference method, the theoretical free-field decay is computed by

$$L_p(r_i) = L_p(r_{\text{ref}}) - 20\log_{10}\left[\frac{r_i}{r_{\text{ref}}}\right], \tag{2}$$

where $L_p(r_{\text{ref}})$ is the sound pressure level at a selected reference or normalization distance r_{ref}. Typical practice in the U.S. has been to use a reference distance of 3 ft.

For the optimal reference method, the theoretical free-field decay is computed from

$$L_p(r_i) = 20\log_{10}\left[\frac{a}{r_i - r_o}\right], \tag{3}$$

where a represents the apparent strength of the source and r_o is an offset distance between the physical location of the

source and its effective acoustic center. The a and r_o parameters are computed from the measured sound pressure levels as

$$a=\frac{(\Sigma_{i=1}^{N} r_i)^2-N\Sigma_{i=1}^{N} r_i^2}{\Sigma_{i=1}^{N} r_i \Sigma_{i=1}^{N} q_i - N\Sigma_{i=1}^{N} r_i q_i} \tag{4}$$

and

$$r_o=-\left[\frac{\Sigma_{i=1}^{N} r_i \Sigma_{i=1}^{N} r_i q_i - \Sigma_{i=1}^{N} r_i^2 \Sigma_{i=1}^{N} q_i}{\Sigma_{i=1}^{N} r_i \Sigma_{i=1}^{N} q_i - N\Sigma_{i=1}^{N} r_i q_i}\right], \tag{5}$$

where

$$q_i=10^{-0.05L_{pi}}. \tag{6}$$

In Eqs. (4) and (5), N is the number of measurement points along the traverse, and the L_{pi} and r_i are as defined previously. Equations (4) and (5) follow from a linear least square fit to Eq. (3) of the traverse data, as introduced by Koidan and Huska.[13]

The traverse data used in the optimal reference method will span from some starting point close to a sound source out to some maximum distance. The maximum distance from the source is typically constrained by the geometry and configuration of the chamber under test. The closest point on the traverse, though, is subject to some selection. The closest traverse point to the test source should be in the source's far field. The standards extant at this writing do not define the minimum distance between the traverse microphone and the sound source. Some guidance on this point may be taken from the sound-power measurement portions of the ISO and ANSI standards, where the measurement surface is constrained to be no closer than 1 m to the source. In addition, the ISO 3745 draft extant at this writing may be construed to define the minimum distance as

$$r=\mathrm{MAX}(0.5,95/f)m, \tag{7}$$

where f is the frequency of interest in Hertz. The $r=95/f$ criterion in Eq. (7) is traceable to the far field requirement[25] that $(kr)^2 \gg 1$, specifically, requiring $(kr)^2>3$.

III. TRAVERSE MEASUREMENT SYSTEM

The continuous traverse system employed in the work at hand comprises a microphone traverse system, sound sources, a reference microphone, and a data acquisition system as depicted in Fig. 2. Each of these four major components of the measurement system is described below.

A. Microphone traverse system

The objective of the microphone traverse system is to move a microphone away from a sound source in a continuous, repeatable manner, while providing accurate position data. The system is comprised of a traverse wire, a microphone carriage, and a computer-controlled motion control system. The traverse wire was a 1.6-mm braided steel cable, chosen as a balance between strength and small acoustic profile. The traverse line was anchored to opposite faces of the chamber and made taut using a turnbuckle at one end. A random-incidence microphone ($\frac{1}{2}$-in. Larson-Davis, model 2560) was suspended from the traverse line by a wire carriage as depicted in Fig. 3. Plastic sleeves on the carriage allowed it to move smoothly and quietly along the traverse line. The microphone carriage was pulled up the traverse line by a take-up line. The take-up line is routed to the exterior of the chamber through a pulley system to a take-up spool mounted on a computer controlled stepper motor. The take-up line, pulley system, take-up spool, and its control system comprise the motion control system.

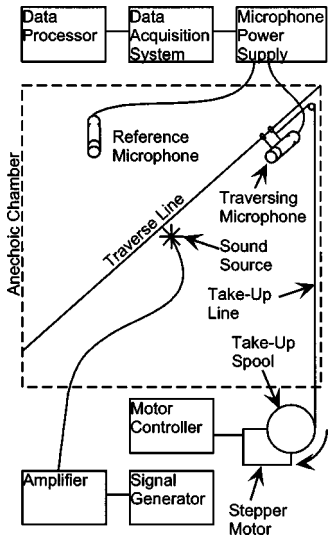


FIG. 2. Traverse system components and configuration.

To avoid slipping on the traverse line and to ensure constant traverse velocity, the microphone carriage was always pulled up the traverse line against gravity, either toward or away from the sound source, depending on the test orientation. The traverse direction was accounted for in the postprocessing stage of the procedure. With horizontal traverses, the direction of travel was toward the source.

The take-up line was fused Kevlar fishing line (''Fire Line'' 20 lb. test). This 0.13-mm-diam line was chosen to ensure that there would be negligible change in the take-up-spool's diameter as line collected on it. Also, the take-up line exhibited no measurable stretch for this application and therefore resulted in highly stable and repeatable movement and positioning of the microphone (a take-up-line that can stretch leads to jerky, erratic motion of the microphone carriage as it ''stick-slips'' along the traverse wire).

The pulley system routes the take-up line outside the chamber via a cable-pass-through pipe, which penetrates the chamber wall from outside to inside. Additional pulleys on the outside of the chamber guides the line onto the take-up-spool mounted on a stepper motor (Superior Electric Slo-Syn

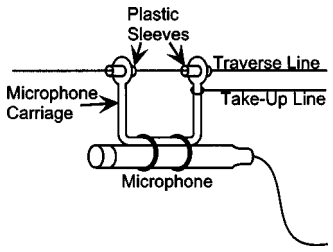


FIG. 3. Traverse microphone carriage.

TABLE II. Description of test sound sources.

Frequency (Hz)	ISO 3745 specified source	Source tested
<400	25-cm-diam speaker in a closed, damped box	15-cm diam Radio Shack 40-1331B, in a closed, damped box
400 to 2000	Two 10-cm speakers, mounted face-to-face, in phase	Two 10-cm diam Radio Shack 40-1022B
⩾2000	Baffled speaker with a cylindrical tube (<1.5-cmdiameter)	Baffled Radio Shack 40-1289A, mid-range tweeter, 1.5-cm i.d. tube (50-cm length)

Motor, Model M062-L809). Motor noise was not detectable inside the chamber. Lightweight Harken marine-grade compact ball-bearing pulleys were used throughout, as ordinary pulleys were found to produce squeaks while in motion, contaminating the test data. The take-up spool/stepper motor was controlled using a ''LabView'' virtual instrument.

Constant angular speed of the take-up spool as it rotated through a known total angle resulted in microphone motion along the traverse line at a constant velocity from one known position to another. Relating velocity and time to distance traveled provided accurate position data.

B. Sound sources and excitation system

Three sound sources were used to cover the design frequency range of the chamber, 80–10 000 Hz. The sources were constructed in accordance with the recommendations in ISO 3745, and are listed in Table II. The sources were mounted using the available floor and ceiling mounts within the chamber. Source location was near the center of the chamber, just below the traverse line and in line with the microphone traverse, with the exception of the low-frequency source. Due to its large size, the center of the low-frequency source was offset from the traverse microphone by 0.3 m; this offset distance was compensated for in postprocessing using a cosine correction in the traverse distance vector. Note that the focus of the paper at hand is upon the traverse and data processing methods, and *not* upon source design or suitability for qualification testing: such considerations are manifest in their complexity, and beyond the scope of the current paper.

A Larson-Davis function generator, Model SRC20, was used to generate both sinusoidal and broadband random noise signals. The random noise signals, to maximize the sound energy within the frequency band of interest, were filtered using a Wavetek band-pass filter, Model 432, prior to amplification by a Ling Dynamic Systems power amplifier, Model PA25E.

TABLE III. Sampling rate, traverse time, and file sizes for pure tone traverses.

Frequency (Hz)	Sampling rate (Hz)	Traverse time (s)	Data file size (MB)
80	512	120	1
125–250	2560	120	5
500	5120	120	10
1000–2000	12 500	120	24
4000	25 600	60	24
6300–10 000	51 200	60	48

TABLE IV. Sampling rate, traverse time, and file sizes for random noise traverses.

Frequency range (Hz)	Sampling rate (Hz)	Traverse time (s)	Data file size (MB)
80–250	2560	2360	92
500–2000	12 500	600	118
4000–10 000	51 200	156	122

C. Reference microphone

To provide a quality check on the traverse data, a random incidence microphone ($\frac{1}{2}$-in. Larson-Davis, model 2560) was placed in a stationary position in the chamber. The data from this microphone provided the means to assess the stability of the source over the time span of a traverse.

D. Data acquisition system

Sound pressure was recorded using a PC-based, two-channel analyzer (DSP Technology, Siglab acquisition system, Model 50-21). Analog low-pass filters were used to anti-alias the data. The sampling rate was selected to be at least 2.5 times the highest frequency of interest for each traverse. Matlab-based Siglab software (version 3.2) was used to store the time-domain, acoustic pressure data. Post-processing of the data is described in a later section.

IV. TEST PROCEDURE

Traverse data were collected in each radial direction, for each frequency (or frequency range), in the following manner: Measurement parameters of the traverse were determined and used to configure the motion control and acquisition software; the acoustic source was turned on, its level set, and dynamic range of the data acquisition system adjusted; the traverse microphone was set into motion and its data recorded. These basic steps are detailed below.

Measurement parameters were the start and stop positions defining the traverse span, the data acquisition sam-

TABLE V. The rms computation parameters for LNE and WW traverses. Pure tone excitation.

Frequency (Hz)	Data points per average		Averages per wavelength		Spatial resolution (cm)		Ratio of spatial resolution to wavelength (%)	
	LNE	WW	LNE	WW	LNE	WW	LNE	WW
80	1997	1997	37.1	64.9	11.6	6.6	2.7	1.5
125	2356	4219	100.6	100.3	2.7	2.7	1.0	1.0
250	1997	2110	59.4	100.3	2.3	1.4	1.7	1.0
500	1997	2120	60.6	100.1	1.1	0.7	1.6	1.0
1000	1997	2650	79.8	100.2	0.4	0.3	1.3	1.0
2000	1997	1997	40.0	66.5	0.4	0.3	2.5	1.5
4000	1997	1997	21.5	32.7	0.4	0.3	4.7	3.1
6300	2000	2000	27.3	41.5	0.2	0.1	3.7	2.4
8000	1997	1997	21.5	32.7	0.2	0.1	4.6	3.1
10 000	1997	1997	17.4	26.2	0.2	0.1	5.8	3.8

TABLE VI. The rms computation parameters for LNE and WW traverses. Random noise excitation.

Frequency (Hz)	Data points per average		Averages per wavelength		Spatial resolution (cm)		Ratio of spatial resolution to wavelength (%)	
	LNE	WW	LNE	WW	LNE	WW	LNE	WW
80	74 156	111 282	100.0	100.0	4.4	4.4	1.0	1.0
125	46 715	70 104	100.0	100.0	2.7	2.7	1.0	1.0
250	23 358	35 052	100.0	100.0	1.4	1.4	1.0	1.0
500	15 924	21 889	100.0	100.1	0.7	0.7	1.0	1.0
1000	9997	10 944	79.7	100.1	0.4	0.3	1.3	1.0
2000	9997	9997	39.8	54.8	0.4	0.3	2.5	1.8
4000	9997	9997	22.5	33.7	0.4	0.3	4.4	3.0
6300	9999	9999	14.2	21.2	0.4	0.3	7.0	4.7
8000	9997	9997	11.2	16.9	0.4	0.3	8.9	5.9
10 000	9997	9997	8.9	13.4	0.4	0.3	11.1	7.4

pling frequency, the traverse duration (time required to complete the traverse), and the speed of microphone movement. The start and stop positions of the traverse microphone were constrained by the physical location of the source and wedge tips. The sampling frequency was set to be at least 2.5 times the highest test frequency. Traverse duration was limited by the data storage capacity of the data acquisition system (128 MB). This limitation was significant for high frequencies ($>$1000 Hz) and for random noise excitation. Finally, traverse velocity was computed as the ratio of traverse length and duration. Traverse times, sampling frequencies, and data file size for the traverse measurements presented here are shown in Tables III and IV for pure tone and noise traverses, respectively.

The source level was set at the traverse microphone's outermost position, at a level at least 10 dB above the ambient noise floor of the frequency band of interest. This was done to ensure an adequate signal-to-noise ratio along the entire traverse. The traverse microphone was then moved to the traverse measurement position nearest the source, where the dynamic range of the microphone power supply and data acquisition hardware were adjusted to avoid data clipping during the traverse.

The traverse microphone was placed in its starting position. The motion control system was then triggered to set the traverse microphone in motion, pulling the microphone up the traverse line, and, simultaneously, triggering the data acquisition system. Output voltage data for the two microphone channels (traverse microphone and reference microphone) were recorded for the duration of the traverse.

TABLE VII. Traverse range used for optimal reference method. Minimum was constrained by near field, maximum by room dimensions.

Frequency (Hz)	Lower northeast corner traverse	West wall traverse
80	1.19–3 m	1.19–2 m
125	0.76–3 m	0.76–2 m
250	0.5–3 m	0.5–2 m
$>$250	0.5–3 m	0.5–2 m

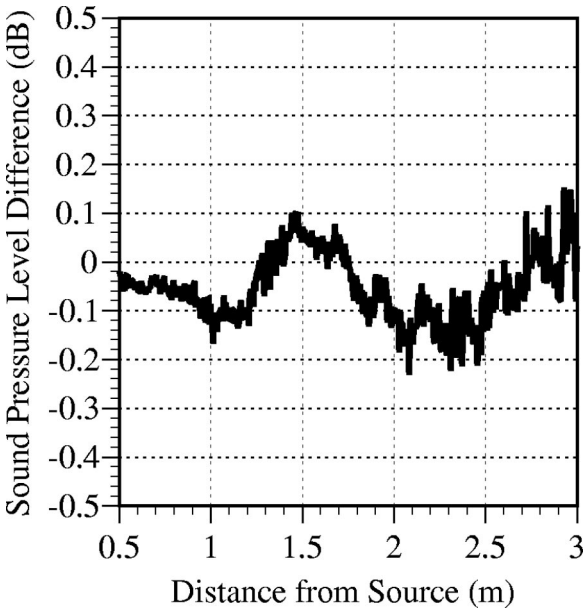


FIG. 4. Difference between two successive 1-kHz pure tone traverses into lower northeast corner.

V. POSTPROCESSING

Prior to computation of the inverse square law deviation as previously described, the traverse data was postprocessed to correct the distance vector if necessary, to band-pass filter the data, and to calculate the rms pressure along the traverse length. As stated above, the low-frequency sound source was significantly offset from the traverse line. This offset was corrected for in the distance-vector using a cosine correction. When the configuration of the traverse line required the microphone to be pulled toward the source, the distance-vector and level data were reversed, yielding data as if the microphone had moved away from the source. The time data was digitally band-pass filtered about the excitation frequency of interest. One-third octave and full octave filters were used for broadband noise excitations as appropriate to the test frequency. For pure tone excitations a narrow band-pass filter (bandwidth$=\frac{1}{20}$ of the center frequency) was used to improve the signal-to-noise ratio.

TABLE VIII. Standard deviation of reference signal for all traverse data.

Frequency (Hz)	Standard deviation of reference signal (dB)			
	LNE		WW	
	Pure tone	Random noise	Pure tone	Random noise
80	0.01	0.22	0.02	0.15
125	0.02	0.09	0.02	0.04
250	0.00	0.11	0.01	0.05
500	0.01	0.26	0.03	0.21
1000	0.01[a]	0.18	0.05	0.17
2000	0.01	0.16	0.02	0.15
4000	0.03	0.28	0.07	0.30
6300	0.03	0.32	0.04	0.31
8000	0.04	0.32	0.05	0.35
10 000	0.08	0.23	0.10	0.22

[a]Reference signal standard deviation for repeatability measurements depicted in Fig. 4 were 0.009 and 0.005 dB.

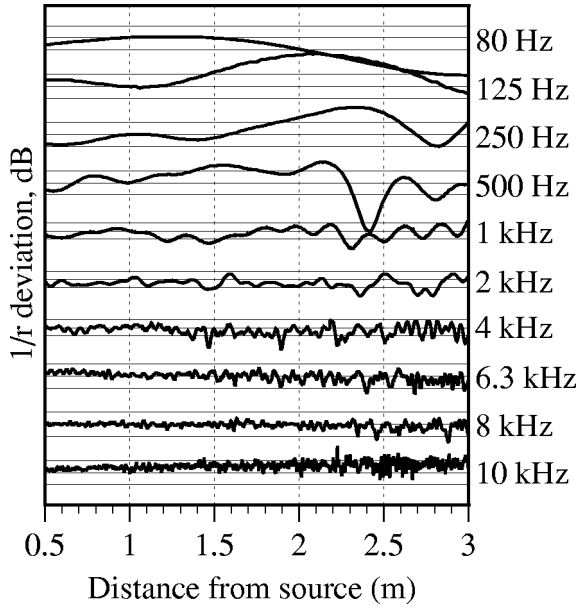


FIG. 5. Deviation from inverse square law by fixed reference method referenced to the level at 1 m for pure tone continuous traverses into LNE corner.

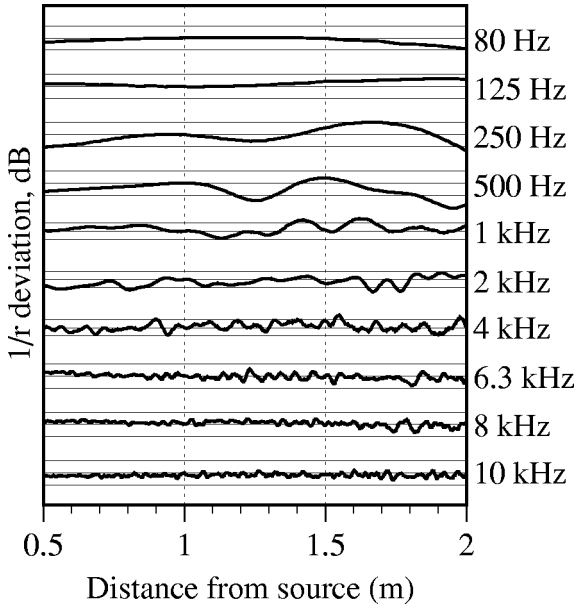


FIG. 7. Deviation from inverse square law by fixed reference method, referenced to level at 1 m, for pure tone continuous traverses into the west wall.

The filtered data was divided into blocks of equal length, and the rms pressure computed for each block. Block length was determined as a compromise between samples per wavelength at each frequency and number of available data points per average (constrained by system memory limitations). A target of 100 rms pressure values per wavelength was desired at each frequency; the spatial resolution of a traverse, expressed in terms of fractional wavelengths, is inversely proportional to the number of rms values per wavelength along the traverse. The 100 rms samples per wavelength yield a spatial resolution of 1% of a wavelength. This resolution was decreased when data acquisition memory limited the total number of data points along the traverse; these limitations manifest themselves on longer traverse and higher frequencies. The exact number of rms samples per wavelength was adjusted so that each rms average was calculated over as close to an integer number of periods as possible at the frequency of interest, thereby yielding an accurate calculation of the rms level of each block of data.

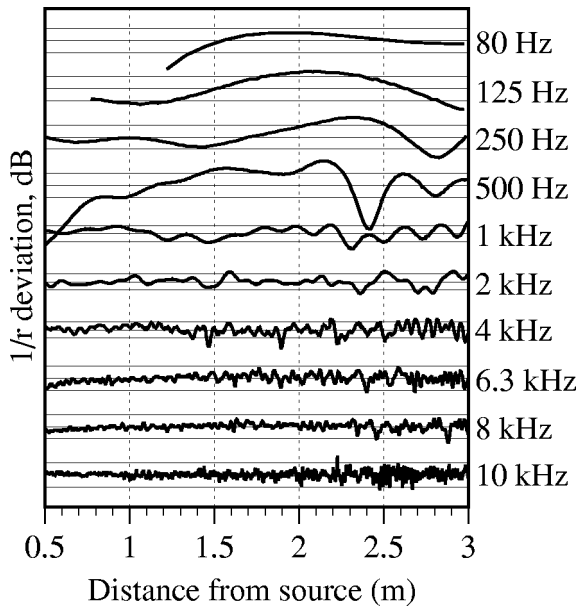


FIG. 6. Deviation from inverse square law by optimal reference method for pure tone continuous traverses into LNE corner.

The parameters used to compute rms pressure for the data presented below are displayed in Tables V and VI for pure tone and random noise traverses, respectively. Those instances in Tables V and VI where the spatial resolution is greater than 1% of a wavelength are due to memory limitations preventing acquisition of sufficient samples.

With the rms pressures and source-microphone separation distances in hand, the deviation from inverse square law performance may be calculated using the procedures detailed in Sec. II. For the optimal reference method, only that portion of the data that met the requirements of Eq. (7) was included in the analysis, yielding the traverse spans listed in Table VII. For the fixed reference method, 1 m from the source was selected as the reference distance.

VI. RESULTS

In the following we present the results obtained for two traverse directions, one into the lower northeast corner (LNE) of the anechoic chamber, and the other into the vertical center of the west wall (WW), 0.1 m to the south side of the wedge basket hinge. These directions were selected as the current standards emphasize traverses into the corners of the room under test, with other traverses selected at the user's discretion. The traverse toward the door was selected anticipating that the structure of the wedge basket support and gaps around its perimeter might generate reflections. Of course, the full qualification test for the chamber has more traverses than those documented here.

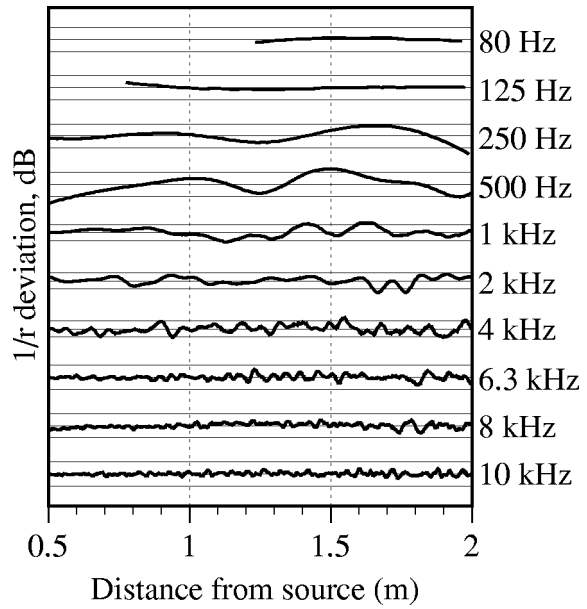


FIG. 8. Deviation from inverse square law by optimal reference method for pure tone continuous traverses into the west wall.

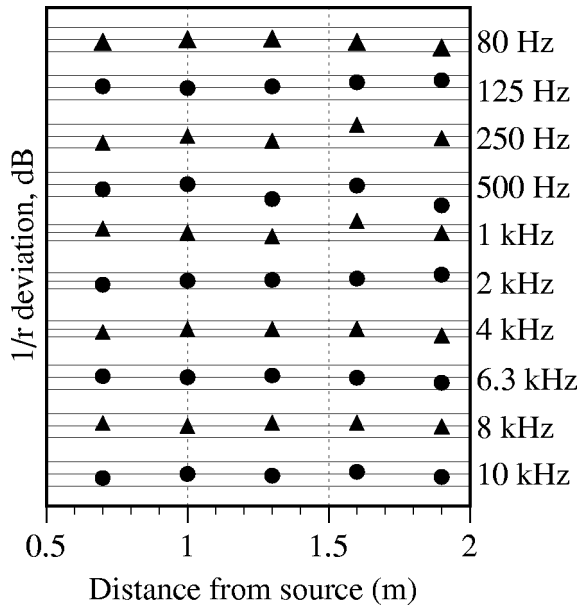


FIG. 10. Deviation from inverse square law by fixed reference method, referenced to 1 m. Data points taken from pure tone continuous traverse data at 0.3-m increments, west wall traverse.

A. Traverse repeatability and source stability

A significant value of the traversing system detailed here is its ability to produce repeatable results with high spatial resolution. The need for establishing the repeatability of the system comes down to one's confidence in comparing the derived difference data to the tolerance limits for chamber qualification, Table I. Figure 4 depicts the difference between two successive pure tone traverses at 1000 Hz into the chamber's lower northeast corner. The standard deviation of the difference between the two data sets is 0.07 dB. Similar repeatability was obtained regardless of traverse direction and frequency.

An associated concept to the spatial repeatability of the traverse is the temporal stability of the test sound source. Table VIII presents the statistics for the source stability, as measured at the reference microphone, during each of the traverses discussed below. We judge that the source was acceptably stable during the traverses. The data for the pure tone traverses clearly has a lower deviation than the broad band noise, but such is to be expected given the nature of these two signals. Indeed, the longer traverse times for the broad band noise tests were driven by the need to obtain stable source level statistics for each of the computed averages during a traverse.

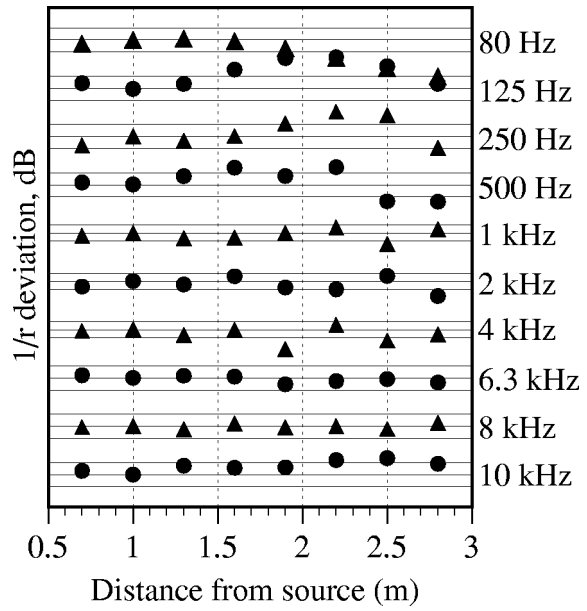


FIG. 9. Deviation from inverse square law by fixed reference method, referenced to level at 1 m. Data points taken from pure tone continuous traverse data at 0.3-m increments into LNE corner.

B. Qualification traverses with pure tones

Figure 5 presents the deviation from free-field performance obtained using the fixed reference method for pure tone traverses into the lower north east corner of the chamber. Figure 6 presents the deviation obtained using the optimal reference method. It is important to note that both figures were generated based on the same underlying raw data; the differences arise due to the method of processing the data. Figures 7 and 8 present the comparable data for traverses into the west wall. For these plots, the frequency labels along the right vertical axis define the position of the 0 dB deviation for that frequency; the dashed lines above and below the 0-dB line are the permissible variations from free-field performance, Table I. Had the measured data matched free-field performance, the deviation would have tracked identically the 0 dB line; such is clearly not the case. For the purposes of chamber qualification, though, the data in Figs. 5–8 may be used to determine the minimum distance at which the tolerance limits at any frequency exceed the tolerance limits, Table I. For example, for the 1-kHz band and fixed reference method, the first exceedance occurs at 1.25 m. Interestingly, the magnitude of the variation of the deviation is appreciably less along this traverse as compared to

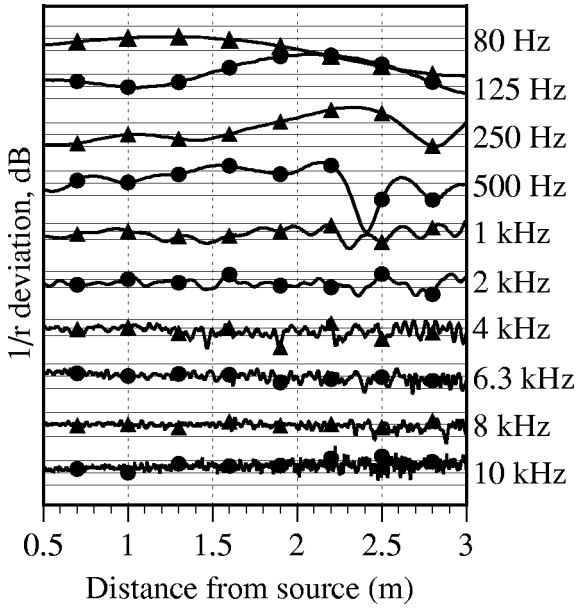


FIG. 11. Overlay of deviation from inverse square law for pure tone continuous traverse data and 0.3-m spacing data, fixed reference method referenced to 1 m, LNE corner.

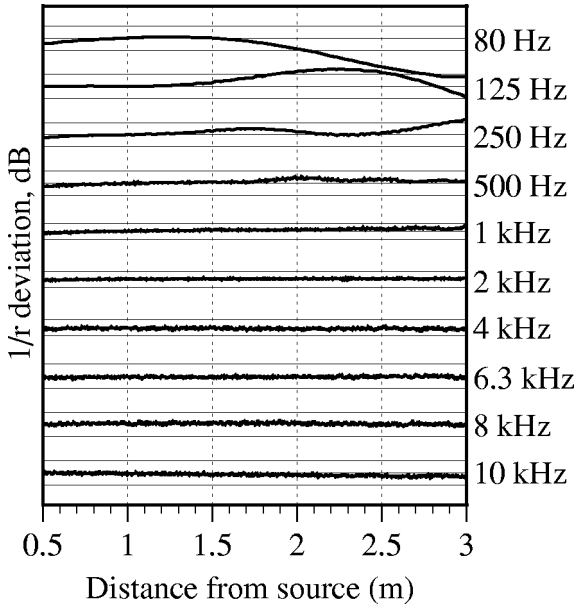


FIG. 13. Deviation from inverse square law by fixed reference method, referenced to 1 m, for random noise continuous traverses into the LNE corner.

that observed along the traverse into the LNE corner. Apparently, the structure about the wedge basket and the door on the west wall are not significant reflectors.

Discussion of the relative performance of the optimal reference method versus the fixed reference method will be deferred until after presentation of the data using broadband noise. However, qualitatively comparing Figs. 5 and 6, and 7 and 8, it is evident that the optimal reference method acts to bring more of the traverse within the tolerance limits, with the greatest observable impact in the lower frequencies (principally 80 and 125 Hz).

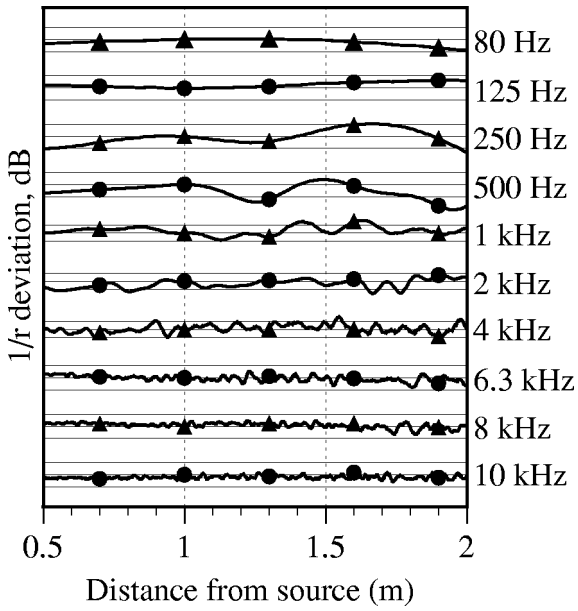


FIG. 12. Overlay of deviation from inverse square law for pure tone continuous traverses and 0.3-m spacing data, fixed reference method referenced to 1 m, west wall.

C. Discrete versus continuous traverses, pure tone

As is observable in the previous publications,[3,16,17] the deviation from inverse square performance varies significantly along a traverse. As the deviation is due to interference between the direct and reflected components of the wave field, the spatial variation is expected to occur on length scales comparable to a fraction of a wavelength, and this is, in fact, observed in the data presented in Figs. 5–8; the shorter the wavelength, the shorter the spatial scale of the deviation. The problematic significance of this observation, though, is whether or not discrete sampling methods are capable of elucidating this spatial complexity.

Figures 9 and 10 depict the results one would obtain by sampling the traverse data at discrete intervals, 0.3 m (~1 ft), analyzed using the fixed reference method. The data was extracted from the continuous traverses of Figs. 5 and 7. While this data does indicate some performance issues with the chamber, it is a poor representation of the full complexity of the sound field, as may be best seen by overlaying the discrete data on the corresponding continuous traverses, as is depicted in Figs. 11 and 12. Clearly, the inability of coarsely spaced discrete samples to capture the complexity of the deviation increases with increasing frequency. For example, for the 4-kHz traverse in Fig. 12, the 1-ft spacing discrete traverse would indicate no performance issues with the chamber over the entire traverse, while the continuous traverse identifies clear violations of the tolerance limits within 1.5 m of the source.

D. Qualification traverses with broadband noise

Next, consider the free-field performance obtained using broadband random noise for the excitation signal. As evident from Table IV, the use of broadband noise leads to much longer traverse times and larger data sets. Figures 13 and 14 are the resulting free-field deviation plots for traverses into

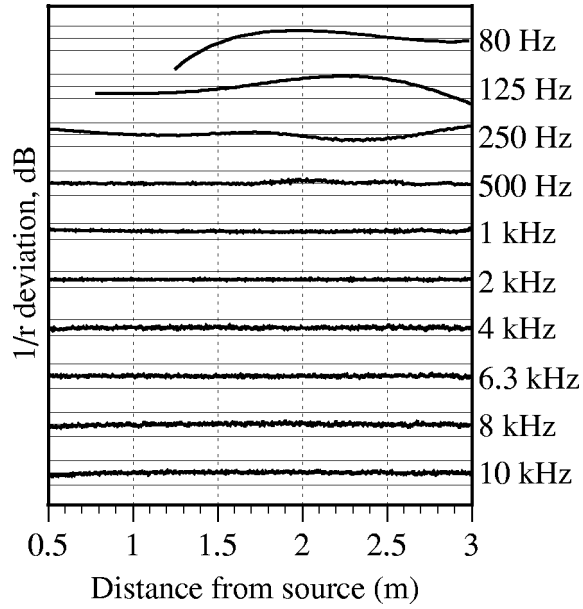


FIG. 14. Deviation from inverse square law by optimal reference method, for random noise continuous traverses into the LNE corner.

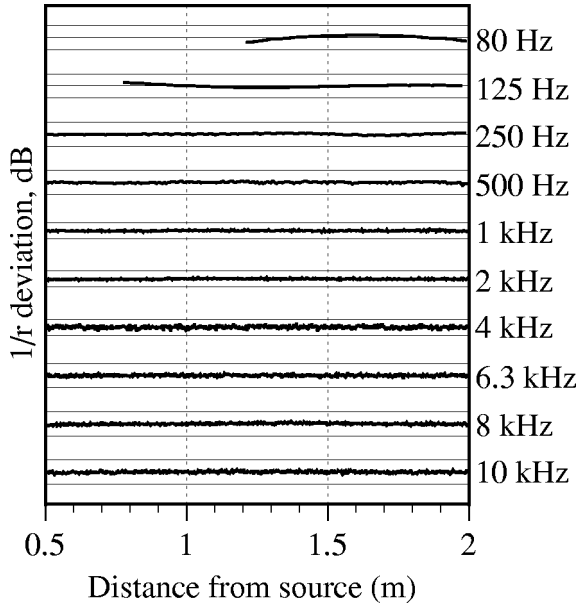


FIG. 16. Deviation from inverse square law by optimal reference method for random noise continuous traverses into west wall.

the lower northeast corner, analyzed using the fixed reference and optimal reference methods, respectively. Figures 15 and 16 are the equivalent plots for traverses into the west wall. The optimal reference method again has impact on the lowest frequencies, bringing much of the deviation to within the tolerance bounds. With the exception of the data at 80 and 125 Hz, though, there is very little deviation from inverse square law performance over these traverses, regardless of the analysis method. Indeed, for these traverses there is no question that even very coarsely spaced discrete samples would yield adequate representations of the chamber performance. In contrast to the results obtained with the pure tone excitations (Figs. 5 and 6, and Figs. 9 and 10), the deviations observed using broadband noise are unremarkable. These results support the conclusion that discrete samples at just about any spacing would seem to be suitable for broadband noise qualification.

E. Optimal versus fixed reference method; pure tone versus broadband noise

Drawing visual comparisons between Figs. 5 and 6, and between Figs. 7 and 8, it is evident that the optimal reference technique serves to bring more of the deviation data within the tolerance limits. The effect is most prevalent at the lower frequencies. The optimal reference method analyses that underlie the data in Figs. 6 and 8 used only those portions of a traverse data set that satisfied the requirements of Eq. (7) with respect to the minimum microphone to source separation distance. Figure 17 illustrates the results that may be obtained if the minimum separation distance is permitted to be closer to the source than as defined by Eq. (7). The three

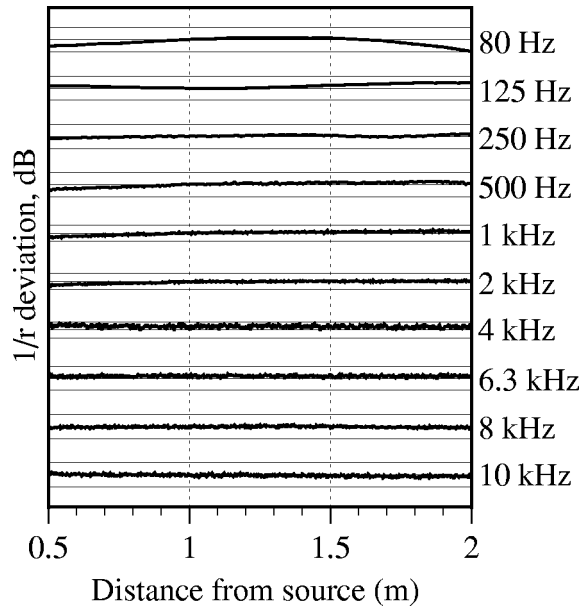


FIG. 15. Deviation from inverse square law by fixed reference method, referenced to level at 1 m, for random noise continuous traverses into west wall.

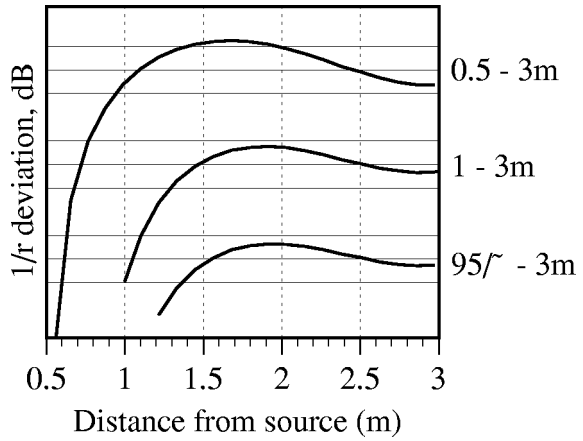


FIG. 17. Impact of minimum source to microphone separation distance on results of optimal reference method for 80-Hz pure tone continuous traverse into LNE corner. Traverse span noted along right-hand *y* axis.

TABLE IX. Percentage of traverse samples that exceed tolerance limits for optimal reference method and fixed reference method, pure tone and noise traverses. Fixed reference method normalized to level at 1 m.

Frequency (Hz)	LNE traverses				WW traverses			
	Pure tone		Noise		Pure tone		Noise	
	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal
80	56	13	54	10	0	0	0	0
125	53	52	30	5	0	0	0	0
250	29	25	4	0	8	3	0	0
500	37	51	0	0	22	18	0	0
1000	16	10	0	0	11	9	0	0
2000	14	11	0	0	14	4	0	0
4000	15	9	0	0	4	2	0	0
6300	2	1	0	0	0	0	0	0
8000	1	1	0	0	0	0	0	0
10 000	9	1	0	0	0	0	0	0

deviation lines in Fig. 17 were derived from the exact same set of physical traverse data, with the only difference being the starting point of the data that was included in the analysis of Eqs. (3)–(5). The traverse data passed to the analysis started at distances of 0.5, 1.0 and $95/f$ m. The analysis starting at 0.5 m shows a strong near-field impact, that at 1.0 m less so, and that at $95/f$ even less. The results indicate that the selection of a minimum separation distance should account for the frequency-dependent near-field characteristics of the test source.

To further quantify the impact of the optimal reference method, as well as the use of pure tone versus broadband noise, consider Table IX. Table IX contains the percentage of each traverse that exceeds the defined performance limits (Table I) over the fitting range of the traverse (Table VII). With the exception of the 500-Hz traverse into the LNE corner, the optimal reference method improves the fit of the traverse curve within the tolerance limits, bringing a greater fraction of the traverse within the bounds. The optimal reference method effectively affords the traverse data its best chance to qualify within the tolerance bounds. Further, as evidenced by the percentage of samples that violate the tolerance limits, the use of pure tones is a much more stringent test of the chamber's performance than the use of random noise.

A side benefit of the optimal reference method is that one need not know the exact source-to-microphone separation distance, as the method will compensate for that through the computation of the source offset (however, the spacing between microphone positions along the traverse must be known accurately).

The source offsets calculated in the optimal reference method, Eq. (5), are listed in Table X. Of particular interest are the two instances noted in the table where the calculated source offset exceeds twice the maximum source dimension (at the time of this writing, there was a draft ISO 3745 in circulation that proposed a maximum permissible source offset of two times the largest source dimension). Of note, these two cases of excessive source offset correspond to traverses with poor performance with respect to the tolerance bounds, as well (Figs. 6 and 14, and the 80- and 500-Hz LNE data in Table IX).

TABLE X. Source offset calculated by optimal reference method, normalized by characteristic source dimension.

Frequency (Hz)	Optimized source offset (normalized by source dimension)			
	LNE		WW	
	Pure tone	Noise	Pure tone	Noise
80	1.92	2.02[a]	0.90	0.98
125	−0.05	0.02	−0.53	−0.35
250	−0.42	−0.33	−0.33	−0.06
500	2.05[a]	−0.40	1.15	−0.55
1000	−0.38	−0.39	−0.19	−0.41
2000	0.19	−0.08	−0.74	−0.24
4000	0.21	0.11	−0.12	0.16
6300	1.26	−0.09	0.60	0.01
8000	0.65	0.21	0.71	0.07
10 000	−0.64	0.52	−0.03	0.29

[a]Optimized source offset exceeds twice the largest source dimension; exceeds the source offset constraint in the draft ISO 3745 qualification specification.

F. Spatial sampling

The deviations from inverse square law performance as depicted in any of the figures with pure tone continuous traverses exhibit a variation consistent with an interference pattern. A measure of the spatial variation within the pattern

TABLE XI. Estimated averaged spatial separation of zero crossings in deviation from inverse square law for pure tone traverse data.

Frequency (Hz)	Wavelength (m)	Estimated spatial period (m)			% Wavelength
		LNE	WW	Average	
80	4.288	1.750		1.750	40.8
125	2.744	1.100	1.000	1.050	38.3
250	1.372	0.500	0.400	0.450	32.8
500	0.686	0.250	0.250	0.250	36.4
1000	0.343	0.100	0.100	0.100	29.2
2000	0.172	0.050	0.050	0.050	29.2
4000	0.086	0.028	0.029	0.029	33.3
6300	0.054	0.022	0.019	0.020	37.0
8000	0.043	0.016	0.015	0.016	36.5
10 000	0.034	0.012	0.013	0.012	36.0

may be obtained by estimating the average separation distance between zero crossings of the deviation data. The average zero crossing separation is a measure of the spatial scale which a traverse must be capable of resolving in order to adequately sample the field. Table XI presents the results of such an analysis applied to the pure tone traverse considered here. Expressed in terms of percentage of a wavelength, the average zero crossing separation ranged from a low of 29.2% at 1000 and 2000 Hz to a high of 40.8% at 80 Hz. If one takes a Nyquist-criterion perspective on these results, then the minimum spatial sampling should be at least twice the minimum spatial scale of interest (note that higher resolution means shorter spacing between samples). For the data presented in this paper, this would imply a minimum spatial resolution (distance between samples) of 15% of a wavelength at the frequency of interest. At low frequencies such a spacing is relatively easy to achieve with discrete sampling methods; it is much less practicable at higher frequencies. This is further indication that a continuous or, at the very least, a systematically densely sampled traverse is vital to achieve meaningful qualification results with the use of pure tones.

VII. CONCLUSIONS

We have demonstrated here that it is feasible and practical to implement a system to perform continuous traverses for the purposes of qualification of anechoic chambers. We find that the optimal reference method to be merited for the analysis of the data obtained from such continuous traverses. In light of the marked differences between the pure tone and broad band noise traverses observed in our facility, it is our opinion that the use of broadband noise as the test sound signal is of dubious value for qualification purposes in chambers where sources with pure tone components are to be tested. If only sources with broadband characteristics are of interest, then broadband qualification is clearly appropriate.

It is our opinion that anechoic chambers should be tested with pure tones and a continuous traversing system. The counter-argument, that such traversing systems or that discrete sampling on a fine spacing is impractical, is specious: if the intent is to find where the performance deviates from inverse square law, then the physics demand such a spacing for pure tone testing. For chambers that will only be used to test broad band noise sources, then discrete sampling at coarse intervals would seem to be valid.

[1] E. H. Bedell, ''Some data on a room designed for free field measurements,'' J. Acoust. Soc. Am. **8**, 118–125 (1936).

[2] L. L. Beranek and H. P. Sleeper, Jr., ''The design and construction of anechoic sound chambers,'' J. Acoust. Soc. Am. **18**, 140–150 (1946).

[3] E. C. Bell, L. N. Hulley, and N. C. Mazumder, ''The steady-state evaluation of small anechoic chambers,'' Appl. Acoust. **6**, 91–109 (1973).

[4] J. L. Davy, ''Evaluating the lining of an anechoic room,'' J. Sound Vib. **132**(3), 411–422 (1989).

[5] v. H. G. Diestel, ''Messung des mittleren Reflexionsfaktors der Wandauskleidung in einem reflexionsarmen Raum'' (''Measurement of the averaged pressure reflection coefficient of the absorbing lay-out of an anechoic room''), Acustica **20**, 101–104 (1968).

[6] F. J. Babineau and B. D. Tinianov, ''Research into quality assessment methods for anechoic chambers,'' in *CD-ROM Proceedings of Noise-Con 2000* (Newport Beach, CA, 2000), paper 1pNSb6, available from Institute of Noise Control Engineering, Saddle River, NJ.

[7] ANSI, ANSI S12.35-1990, ''Precision methods for the determination of sound power levels of noise sources in anechoic and hemi-anechoic rooms'' (Standards Secretariat, Acoustical Society of America, Melville, NY, 1990).

[8] ISO, ISO 3745, ''Acoustics—Determination of sound power levels of noise sources—Precision methods for anechoic and semi-anechoic rooms'' (International Organization for Standardization, Geneva, Switzerland, 1977).

[9] H. F. Olson, ''Acoustic laboratory in the new RCA Laboratories,'' J. Acoust. Soc. Am. **15**, 96–102 (1943).

[10] H. P. Sleeper, Jr., E. E. Moots, and L. L. Beranek, ''The Harvard anechoic chamber,'' CIR-51, Electro-Acoustic Laboratory, Harvard University, 1945.

[11] P. J. Mills, ''Construction and design of Parmly Sound Laboratory and anechoic chamber,'' J. Acoust. Soc. Am. **19**, 988–992 (1947).

[12] H. C. Hardy, F. G. Tyzzer, and H. H. Hall, ''Performance of the anechoic room of the Parmly Sound Laboratory,'' J. Acoust. Soc. Am. **19**, 992–995 (1947).

[13] W. Koidan and G. R. Hruska, ''Acoustical properties of the National Bureau of Standards anechoic chamber,'' J. Acoust. Soc. Am. **64**, 508–516 (1978).

[14] M. Pancholy, A. F. Chhapgar, and V. Mohanan, ''Design and construction of an anechoic chamber at the National Physical Laboratory of India,'' Appl. Acoust. **14**, 101–111 (1981).

[15] R. R. Boullosa and A. P. Lopez, ''Some acoustical properties of the anechoic chamber at the Centro de Instrumentos, Universidad Nacional Autonoma de Mexico,'' Appl. Acoust. **56**, 199–207 (1999).

[16] A. N. Rivin, ''An anechoic chamber for acoustical measurements,'' Sov. Phys. Acoust. **7**(3), 258–268 (1962).

[17] F. Ingerslev, O. J. Pedersen, and M. P. K. Mollller, ''New rooms for acoustic measurements at the Danish Technical University,'' Acustica **19**(4), 185–199 (1968).

[18] J. Duda, ''Inverse square law measurements in anechoic rooms,'' Sound Vib. **December**, 20–25 (1998).

[19] J.-q. Wang and B. Cai, ''Calculation of free-field deviation in an anechoic room,'' J. Acoust. Soc. Am. **85**, 1206–1212 (1989).

[20] N. Olson, ''Acoustic properties of anechoic chamber,'' J. Acoust. Soc. Am. **33**, 767–770 (1961).

[21] J. Duda, M. C. Hastings, and R. D. Godfrey, ''Qualification of the sound field in a Metadyne anechoic chamber,'' in *Proceedings of ASME International Mechanical Engineering Congress and Exposition*, Dallas, TX, 1997, available from ASME, NCA Vol. 24, pp. 93–96.

[22] J. Impeduglia, ''Acoustic testing facilities raises plant capacity,'' Sound Vib. **July**, 6–8 (1999).

[23] K. O. Ballagh, ''Calibration of an anechoic room,'' J. Sound Vib. **105**(2), 233–241 (1986).

[24] P. G. Lynde and R. J. Buelow, ''Development of an alternative metal anechoic wedge,'' Sound Vib. **October**, 6–16 (1993).

[25] L. L. Beranek, *Acoustics* (Acoustical Society of America, Woodbury, NY, 1986).

Adaptation in a revised inner-hair cell model

Christian J. Sumner[a)]
Centre for the Neural Basis of Hearing at Essex, Department of Psychology, University of Essex, Colchester CO4 3SQ, United Kingdom

Enrique A. Lopez-Poveda
Centro Regional de Investigación Biomédica, Facultad de Medicina, Universidad de Castilla—La Mancha, Campus Universitario, 02071 Albacete, Spain

Lowel P. O'Mard and Ray Meddis
Centre for the Neural Basis of Hearing at Essex, Department of Psychology, University of Essex, Colchester CO4 3SQ, United Kingdom

(Received 29 March 2002; revised 24 August 2002; accepted 30 August 2002)

A revised computational model of the inner-hair cell (IHC) and auditory-nerve (AN) complex was recently presented [Sumner *et al.*, J. Acoust. Soc. Am. **111**, 2178–2188 (2002)]. One key improvement is that the model reproduces the rate-intensity functions of low- (LSR), medium- (MSR), and high-spontaneous rate (HSR) fibers in the guinea-pig. Here we describe the adaptation characteristics of the model, and how they vary with model fiber type. Adaptation of the revised model for a HSR fiber is in line with an earlier version of the model [Meddis and Hewitt, J. Acoust. Soc. Am. **90**, 904–917 (1991)]. In guinea-pig, poststimulus time histograms (PSTH) have been found to show less adaptation in LSR fibers. Evidence from chinchilla suggests that this is due to chronic adaptation resulting from short interstimulus intervals, and that fully recovered LSR fibers actually show more adaptation. However, the model is able to account for both variations of PSTH shape when fully recovered from adaptation. Interstimulus interval can also affect recovery in the model. The model is further tested against data previously used to evaluate models of AN adaptation. The tests are (i) recovery from adaptation of spontaneous rate and (ii) the recovery of response to acoustic stimuli (''forward masking''), (iii) the response to stimulus increments and (iv) decrements, and (v) the conservation of transient components. A HSR model fiber performs similarly to the earlier version of the model. However, there is considerable variation in response to increments and decrements between different model fibers.
[DOI: 10.1121/1.1515777]

PACS numbers: 43.64.Bt, 43.66.Ba [LHC]

I. INTRODUCTION

We have recently presented a revised version (Sumner *et al.*, 2002) of an earlier inner-hair cell (IHC) auditory-nerve (AN) model (Meddis, 1986, 1988; Meddis *et al.*, 1990). This model combines a modified bio-physical model of the receptor potential, a calcium-driven mechanism for neurotransmitter release, and a quantal-stochastic version of the Meddis IHC model of transmitter recycling. The model is able to reproduce the rate-intensity functions of low- (LSR), medium- (MSR), and high-spontaneous rate (HSR) fibers, when used with a suitable nonlinear model of cochlear mechanical filtering. The fiber type is largely determined by the number of calcium channels near the synapse. A striking feature of AN activity is the reliable reduction in firing rate during the response to a stimulus of more than a few milliseconds. This characteristic has been a topic of detailed study in the past (e.g., Smith, 1977; Westerman, 1985; Yates *et al.*, 1985). The original IHC model (Meddis, 1986, 1988) was conceived in large part to account for the adaptation of firing rate observed in the AN. Like other models proposed at the time (e.g., Westerman and Smith, 1988), it modeled the average characteristics of AN adaptation. Since the majority of fibers are of the HSR type, the models are representative of these. However, differences in adaptation have been reported for different fiber types. Here we report on the adaptation characteristics of the new model, and how these adaptation characteristics depend on the fiber type being modeled.

In the first part of this study we attempt to reproduce the reported differences in poststimulus time histogram (PSTH) with fiber type. In the second part, we test six model fibers against data that have been previously used to test models of AN adaptation, in order to facilitate comparisons. Third, we examine the inner workings of the model.

II. THE MODEL

The IHC model is employed as part of a complete peripheral model. This consists of four components: bandpass filtering by the middle-ear; nonlinear mechanical filtering by the cochlea; and IHC transduction and refractory effects of the AN. These components are described fully by Sumner *et al.* (2002).

The IHC model consists of three stages. The output from

[a)]Author to whom correspondence should be addressed. Kresge Hearing Research Institute, University of Michigan, Ann Arbor, MI 48109-0506. Electronic mail: cjsumner@umich.edu

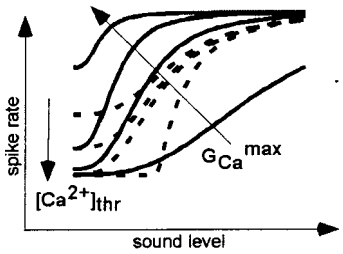


FIG. 1. Effect of varying calcium transmitter release mechanism parameters, G_{Ca}^{max} and $[Ca^{2+}]_{thr}$, on rate-intensity functions. Arrows show the direction of function change for an increase in each parameter value.

the mechanical filtering stage drives a biophysical model of the IHC, modified from a model proposed by Shamma *et al.* (1986). This reproduces realistic receptor potentials (RPs). The RP drives a simple model of the calcium ion movement in the vicinity of the synapse. The local calcium concentration determines the release probability of any neurotransmitter that is available for release. The neurotransmitter cycles around the synapse according to the scheme proposed by Meddis (1986, 1988), except that the transmitter is now contained in discrete packets within the hair-cell. These vesicles are released stochastically into the synaptic cleft, where they can cause the postsynaptic AN to fire. If the AN is not in a refractory state, a single quantum of transmitter released into the cleft will produce an action potential. The rate limited cycling of the neurotransmitter reproduces the adaptation observed in the AN.

In this study we are concerned with the calcium-based transmitter release function, and the movement of neurotransmitter around the synapse. The inward calcium current (I_{Ca}) in the vicinity of a hair-cell synapse is a function of RP (V):

$$I_{Ca}(t) = G_{Ca}^{max} m_{I_{Ca}}^3(t)(V(t) - E_{Ca}), \tag{1}$$

where G_{Ca}^{max} is the maximum conductance with all the calcium channels open, $m_{I_{Ca}}$ is the proportion of open channels (modeled as a Boltzman function of the RP), and E_{Ca} is the reversal potential of calcium. Calcium ion concentration, $[Ca^{2+}](t)$, is a low-pass function of current, and determines the probability of release of neurotransmitter vesicles:

$$k(t) = \max(([Ca^{2+}]^3(t) - [Ca^{2+}]_{thr}^3)z, 0), \tag{2}$$

where $k(t)$ is the instantaneous release probability of a vesicle, $[Ca^{2+}]_{thr}$ is the threshold calcium concentration for vesicle release, and z is a scalar.

Sumner *et al.* described how the model AN fiber response depends on the choice of two parameters in the calcium stage. The continuous lines in Fig. 1 show how rate-intensity (RI) functions vary with G_{Ca}^{max}. A large value of G_{Ca}^{max} (~8 nS) will result in a fiber with HSR type characteristics, while a small value (~2–4 nS) produces a LSR fiber. The dotted lines in Fig. 1 show how the RI functions vary with $[Ca^{2+}]_{thr}$. This parameter affects primarily the low-intensity responses, and thus affects the spontaneous rate and threshold of the unit. It allows for a largely independent control of the spontaneous rate. Together these two parameters afford excellent control over the rate responses of the model.

Adaptation in the model is due to depletion of the neurotransmitter available for release. Transmitter cycles around three reservoirs: the immediate store (q), where transmitter is ready for release; the synaptic cleft (c); and the reprocessing store (w), where transmitter is transported back in to the cell. This movement is modeled by three differential equations:

$$\frac{dq(t)}{dt} = N(w(t),x) + N([M - q(t)],y) - N(q(t),k(t)), \tag{3}$$

$$\frac{dc(t)}{dt} = N(q(t),k(t)) - lc(t) - rc(t), \tag{4}$$

$$\frac{dw(t)}{dt} = rc(t) - N(w(t),x); \tag{5}$$

x and r determine the rates of recycling. In addition, some transmitter is lost in the cleft (at a rate lc) and new transmitter is manufactured [at a rate proportional to $y(M-q)$]. Movement within the cell is a quantal stochastic process. $N(n,\rho)$ is a random process describing probabilistic transport of transmitter quanta. Each of n possible events has an equal probability, ρdt, of occurring in a single simulation epoch.

The model used here has been tuned to guinea-pig AN rate responses at a best frequency (BF) of 16.7 kHz. The complete set of parameter values is given in ''HSR set'' Table II, and ''AN set'' of Table III, of Sumner *et al.* The values of the parameters that are varied in this study are given in Table I.

III. MODELING AN ADAPTATION CHARACTERISTICS

A. Adaptation of HSR fibers

Westerman and Smith (Westerman, 1985; Westerman and Smith, 1984) have made extensive measurements of adaptation characteristics of gerbil AN fibers. These are thought to be very similar to the guinea-pig (Smith, 1977; Rhode and Smith, 1985). Adaptation can be described as a sum of two exponential functions fitted to a PSTH with a 1-ms bin-width:

$$R(t) = A_r e^{-t/t_r} + A_{st} e^{-t/t_{st}} + A_{ss}, \tag{6}$$

where A_r and A_{st} are the components of rapid and short-term adaptation, and A_{ss} is a steady-state component, while t_r and t_{st} are the respective decay time constants for the adaptation. Figure 2 compares model and gerbil data (Westerman, 1985) for all five components as a function of signal level. Since most AN fibers are of the HSR type, we compare the data with the HSR model of Sumner *et al.* (see Sec. II above). The unique parameters of this fiber are given in Table I. Figure 2(a) shows the two time constants of adaptation, t_r and t_{st}, as a function of stimulus level. The short-term time constant is stable between 10 and 40 dB SPL for both animal and the model while the rapid time constant falls over the same range for both. Figure 2(b) shows the components of rapid, short-term adaptation, and steady-state firing. The rapid component (A_r) is the largest, which has a larger dynamic range than the steady-state response (A_{ss}) for both the animal and model observations. The gray regions in the fig

TABLE I. Model parameters and summary results.

		HSR[a,b]	L3G[b]	H3E[b]	MSR[a]	L1[a]	L3C[b]
G_{Ca}^{max} (nS)		8	8	11	4.5	2.75	3.25
$[Ca^{2+}]_{thr}$ ($\times 10^{-11}$), Calcium conc.		4.48	5	5	3.2	4	4.48
M		10	10	10	10	8	10
Driven-recovery	(i) as a single exponential				√		
	(ii) faster for onset than steady state	√	√	√	√	x	x
Increments	(i) onset is additive	x	x	x	x	√	√
	(ii) short-term is additive				√		
Decrements	(i) onset is not additive	√	√	√	√	x	x
	(ii) short-term is additive	x	x	√	√	√	√
Conservation	(i) Rapid components				√		
	(ii) short-term components				√		
SR recovery	Fitted by a single exponential	√		√			

[a]Parameter sets used in Sumner *et al.* (2002).
[b]Parameter sets for fibers shown in Fig. 3. Except for HSR, first letter denotes SR, last denotes panel in Fig. 3 in which they appear.

ure represent the spread of estimates found by Westerman. The model results sit comfortably within those regions. The large steady-state response reflects the AN data to which the rate responses of the model were tuned in Sumner *et al.*

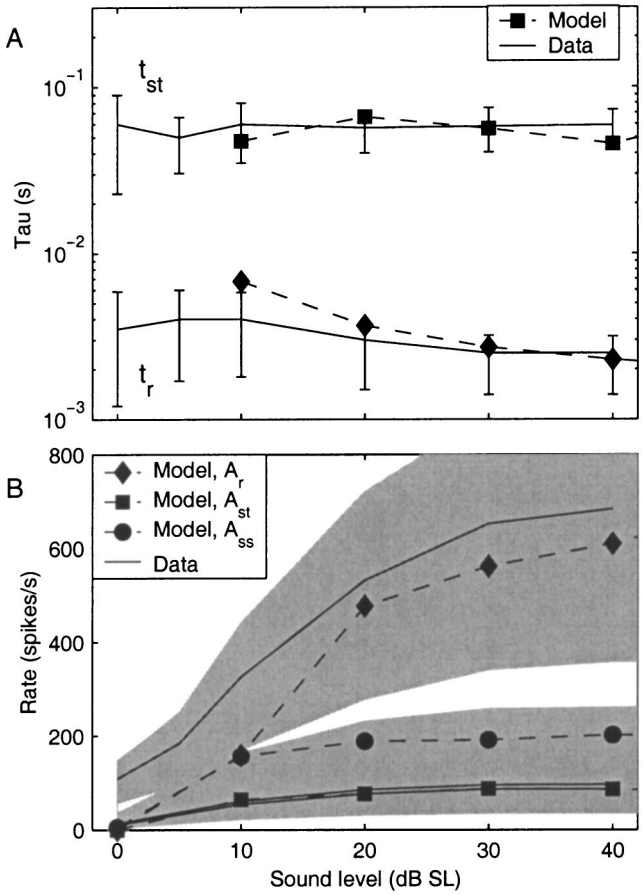


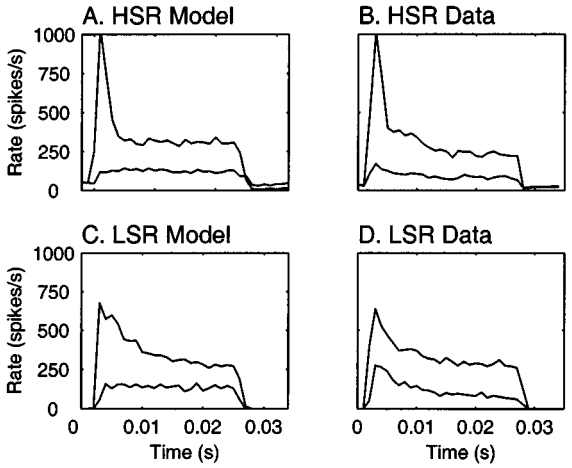
FIG. 2. AN adaptation characteristics. Continuous lines indicate the average values found by Westerman (1985). Connected symbols show the calculated coefficients of the model using the "AN set" DRNL parameters (Table III) and "HSR" synapse parameters (Table II) of Sumner *et al.*. The unique parameters of this model are also given here in Table I. (a) Rapid and short-term time constants of adaptation, t_r and t_{st}. Vertical bars show the standard deviation of the data. (b) Coefficients for rapid (A_r), short-term (A_{st}), and steady-state (A_{ss}) components of the response. Shaded areas show the range of values found by Westerman.

B. Variations in adaptation with fiber type

Rhode and Smith (1985) and Müller and Robertson (1991) investigated the variation in adaptation rates with fiber type, in cat and guinea-pig respectively. They found that HSR fibers show substantial adaptation with time while LSR fibers do not. Figures 3(a)–(d) compare the PSTH responses of two model fibers with guinea-pig data from Müller and Robertson (1991) for both HSR and LSR fibers. The difference in fiber type was modeled with only a reduction in G_{Ca}^{max}. The HSR model was the same as in Fig. 2, and the LSR model had $G_{Ca}^{max}=3.25$ nS. This model fiber will be referred to as L3C hereafter [it has a *Low* spontaneous rate and appears in Fig. 3(c)] and its unique parameters are listed in Table I. The resulting models reproduce qualitatively the observed difference in adaptation. In these simulations the synapse was fully recovered for each stimulus presentation. The lower rate of calcium influx of the LSR fiber means the synapse is driven less hard. As a consequence, transmitter depletion is not so rapid and adaptation is less marked. Both example data fibers show some adaptation at low levels. We were not able to reproduce this effect quantitatively in the model at these low firing rates by changing G_{Ca}^{max}.

Relkin and Doucet (1991) have found that LSR fibers in the chinchilla actually showed larger onsets than HSR fibers at long interstimulus intervals (>1 s). Their data suggest that LSR fibers take longer to recover from adaptation than HSR fibers and that the small amount of adaptation observed for LSR fibers in other studies was a result of a short (80 ms for Müller and Robertson) interstimulus interval. In Figs. 3(e)–(h) we consider PSTHs for fully recovered fibers, from Relkin and Doucet (1991). Figures 3(f) and (h) show two PSTHs in response to 100-ms tone bursts for a HSR and LSR fiber, respectively. Figure 3(e) illustrates the response of a HSR model fiber (H3E) that shows a response similar to the data. Figure 3(g) shows a LSR model (L3G). G_{Ca}^{max} is 8 nS for the L3G model and 11 nS for the H3E model. $[Ca^{2+}]_{thr}$ is 5×10^{-11} in both cases. This is slightly larger than the value of 4.48×10^{-11} used in the fibers HSR and L3C. The patterns of behavior seen in Fig. 3 arise from differences in the resting value of q, the number of vesicles in the immediate store.

Model compared with Müller and Robertson (1991).



Model compared with Relkin and Doucet (1991).

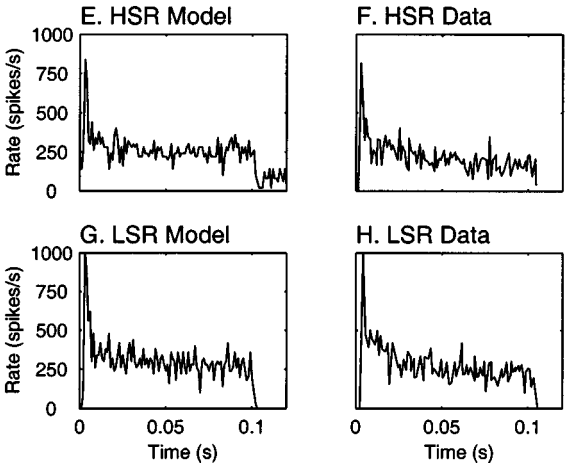


FIG. 3. The PSTH responses for tone bursts for HSR and LSR fibers, showing the variation in adaptation characteristics between fiber type and level. (a) Model response for 25-ms tone bursts at 10 kHz, 66 and 85 dB SPL, using the "AN set" DRNL (Table III) and "HSR" synapse (Table II) parameters of Sumner *et al.* (2002). Unique parameters are shown as "HSR" set in Table I here. (b) Guinea-pig HSR fiber with BF of 17 kHz for stimuli of (a). (c) Model fiber, with parameters as in (a) except $G_{Ca}^{max}=3.25$ nS, showing LSR type adaptation (L3C in Table I). Stimuli are at 12 kHz, 80 and 90 dB SPL. (d) Guinea-pig LSR fiber (from Müller and Robertson, 1991) with a BF of 20.5 kHz for the same stimuli as (c). (e) Model showing HSR response to a 100-ms tone. Parameters as in (a) except $G_{Ca}^{max}=11$ nS, and $[Ca^{2+}]_{thr}=5\times10^{-11}$ (H3E in Table I). (f) Chinchilla HSR fiber. (g) Model showing LSR response. Parameters as in (a) except $G_{Ca}^{max}=8$ nS, and $[Ca^{2+}]_{thr}=5\times10^{-11}$ (L3G in Table I). (h) Chinchilla LSR fiber (from Relkin and Doucet, 1991). (e)–(h) all use BF tone pips at 40 dB above threshold.

The unadapted vesicle release rate is qk [see Eq. (3)]. The H3E model fiber has a resting value for q of approximately 6. This is smaller than the maximum value (M) because of frequent spontaneous releases. The L3G model fiber has no spontaneous rate, and consequently $q=M=10$ in the absence of stimulation. The difference in q is larger than the opposing difference in k. Therefore, the L3G model has the stronger onset response. This mechanism is similar to the speculations of Relkin and Doucet.

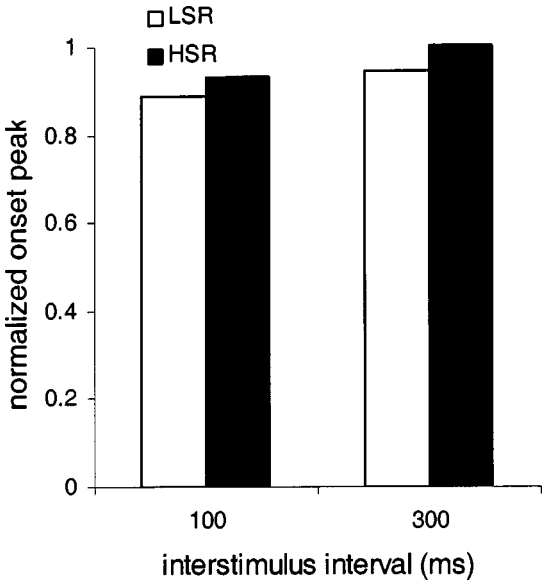


FIG. 4. Effect of interstimulus interval on normalized onset rates for two model fibers. The solid bars show the response of the HSR model fiber and the open bars show the response of the L1 fiber. The onset-rate is normalized by the onset peak measured at an interstimulus-interval of two seconds.

Relkin and Doucet (1991) found that an interspike interval (ISI) of 300 ms was adequate to demonstrate complete recovery in HSR fibers. However, LSR fibers were only 80% recovered by that time. We have simulated their experiment 2 using the model fibers HSR and L1. The tone stimuli were 100 ms in duration and were presented repeatedly with a separation of either 100 or 300 ms for 600 trials. The dependent variable was the onset spike rate computed using the number of spikes in the most populated 1-ms bin of the PSTH after the onset of the stimulus. The onset rate was normalized relative to the fully recovered rate observed when the ISI was 2 s. The stimuli were presented at 50 and 80 dB SPL for fibers HSR and L1, respectively; 40 dB above threshold in both cases. The model results, which show the same effect as Relkin and Doucet, are shown in Fig. 4. The HSR fiber is almost completely recovered when the ISI is 300 ms but the LSR fiber is less than 90% recovered despite a higher rate of response at the 100-ms ISI.

IV. MODELING OTHER MEASURES OF AN ADAPTATION

There are many other measures of AN adaptation processes. Other studies (Hewitt and Meddis, 1991; Ross, 1996) have modeled a range of these measures. The model of Meddis (1986, 1988) was among the models tested there. Here we will test the revised version of the model for the same measures, in order to compare its performance with previous models. No attempt has been made to vary the parameters to optimize the fit of the revised model to these responses. Instead, we have used three model fibers that have already been presented in this study and three from Sumner *et al.* (2002). The unique parameters of the fibers are given in

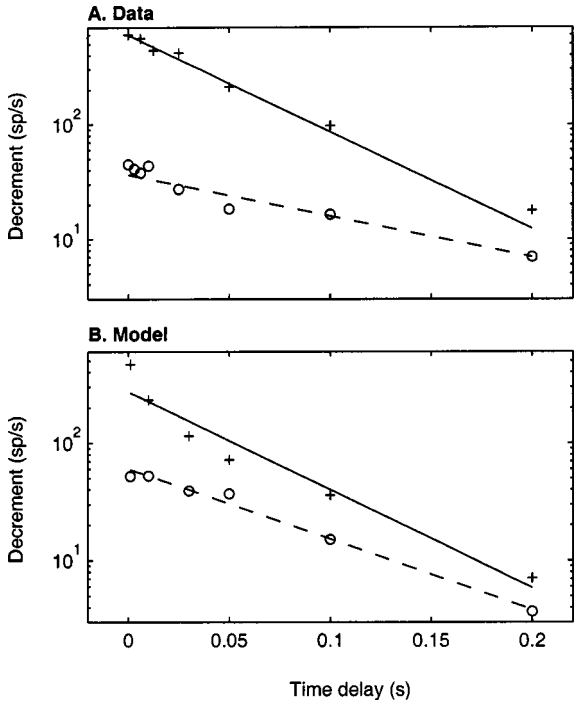


FIG. 5. Recovery of driven responses to sound. (a) Data from Westerman (1985): the decrement in firing rate of the onset (1-ms window at probe onset; crosses are data and continuous line is the fitted recovery function) and short-term (last 20 ms of probe; circles are data and dashed line is the fitted recovery function) portions of the response to a 43 dB SL 30-ms tone, when preceded by a 300-ms masker at the same level. (b) Averaged responses of model fibers HSR, MSR, H3E, and L3G for the paradigm of Westerman.

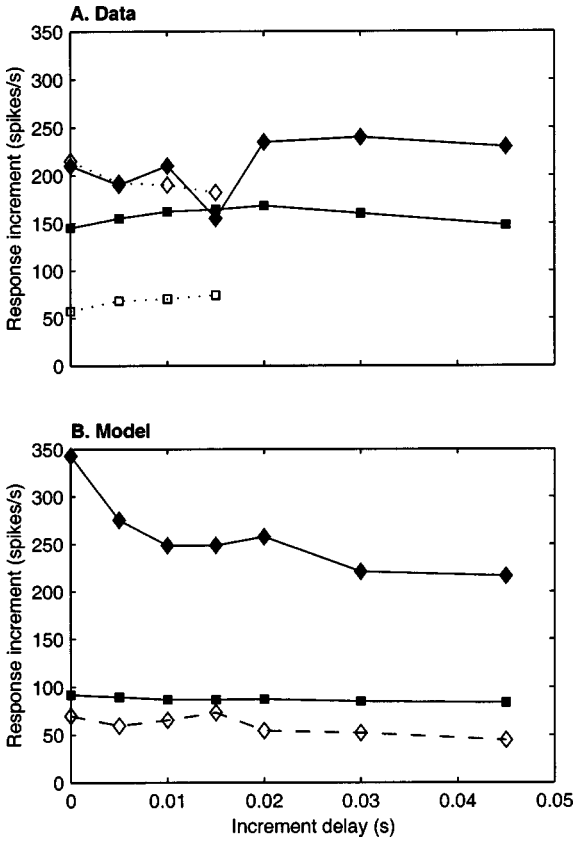


FIG. 6. Increments in firing rate in response to increments in stimulus level. (a) Data from Smith *et al.* (1985) showing the increment in firing rate with onset (continuous lines and filled diamonds) and short-term (continuous lines and filled squares) window analyses, for different increment delays. The stimulus is a 60-ms BF tone, which starts at 13 dB SL and subsequently increases in level by 6 dB. Also shown are the results of Winter *et al.* (1993) as dashed lines and open symbols (diamonds are onset window increments and squares are short-term window increments). (b) Results of applying the paradigm of Smith *et al.* to the model fibers: (diamonds and continuous line), average onset-window increments for fibers: HSR, MSR, H3E, and L3G; (open diamonds and dashed line), average of onset responses for remaining model fibers (L1 and L3C); (filled squares and continuos line), average of short-term window responses for all fibers.

Table I. All other parameters are the same as the ''AN set'' DRNL parameters (Table III) and ''HSR'' synapse parameters (Table II) of Sumner *et al.*

A. Recovery of responses to stimulation

The response of AN fibers to acoustic simulation is reduced immediately following prior stimulation. This is presumed to be a function of the adaptation observed during stimulation, and is probably in part responsible for the psycho-acoustical phenomenon of forward masking (Moore, 1997). Forward-masking paradigms have been used to study the recovery of AN responses (e.g., Smith, 1977; Harris and Dallos, 1979; Westerman, 1985). Figure 5(a) shows an example of the response of a gerbil AN fiber from Westerman. He used a 300-ms ''masker'' followed by a 30-ms probe at various delays. He analyzed the response of the probe for 1-ms windows at the onset of the probe response (onset window) and windows covering the last 20 ms of the response (short-term window) in order to reveal the recovery of rapid and short-term adaptation processes. The upper data set (crosses and continuous fitted line) is the onset window response, expressed as a decrement in firing rate relative to the corresponding portion of the masker. The lower data set (open circles and dashed fitted line) is the short-term window decrement. Westerman found, in 11 fibers, that the onset response recovered with a mean time constant of 50 ms, and the short-term window response recovered with a mean time constant of about 170 ms. Figure 5(b) shows the average of the responses of the model fibers HSR, MSR, H3E, and L3G. The responses show a good resemblance to the data, including slightly differing time constants for fast and slow recovery. The difference is attributable to the large onset-window decrements at delays of less than 5 ms that increase the gradient of the best-fit line.

B. Increments

Smith and Zwislocki (1975), in the gerbil, found that the increase in firing rate in response to an increment in stimulus level does not greatly depend on the time between the onset of the tone and the subsequent increment in level. This suggested that adaptation was additive. Smith *et al.* (1985) found that this also held if increments were analyzed with window lengths that separated rapid and short-term adaptation. Figure 6(a) shows the results of Smith *et al.* The stimulus used was a 60-ms BF pedestal tone, at 13 dB SL, with a 6-dB increase in level occurring at various delays up to 40 ms after the start of the pedestal. The diamonds joined by continuous lines indicate the response analyzed over a 1-ms window (onset window) at the start of the response to the

increment. The squares joined by continuous lines show the response over a 10-ms window (short-term window) starting at the same time as the onset window. Smith's results are additive in the sense that the effect of an increment in stimulus intensity is the same irrespective of the time at which the increment is introduced.

Winter *et al.* (1993), using guinea-pigs, subsequently found that onset increments show a statistically significant increase at delays of less than 5 ms. The effect of all other increments were additive. These results are also shown in Fig. 6(a) (open symbols and dashed lines). Westerman's average data also show slight departure from additivity for onset-window increments at short delays. The ratio of the increments at a delay of 0 ms to a delay of 30 ms was 1.2.

Figure 6(b) shows the results of simulating the increment paradigm of Smith *et al.* for the six model fibers. The squares joined by continuous lines show the short-term window increment averaged over all six model fibers. The results show additivity throughout except for increased responses at short delays for onset windows. This agrees quite well with the average measurements of Winter *et al.* (1993). The diamonds joined by a continuous-line show the average of the onset window increments for three model fibers (HSR, H3E, L3G). These fibers all displayed departure from additivity at short delays. The ratio of the increments at $t=0$ to $t=30$ ms is 1.44. This trend is the same as found by Winter *et al.*, although the departure from additivity is greater in the model. The open diamonds joined with a dashed line show the average of the onset-window increment for the three remaining model fibers (MSR, L1, L3C). They were additive for all delays in agreement with Smith's data. This line falls below that of the average short-term response, because the firing rates of fibers L1 and L3C are considerably less than the other fibers.

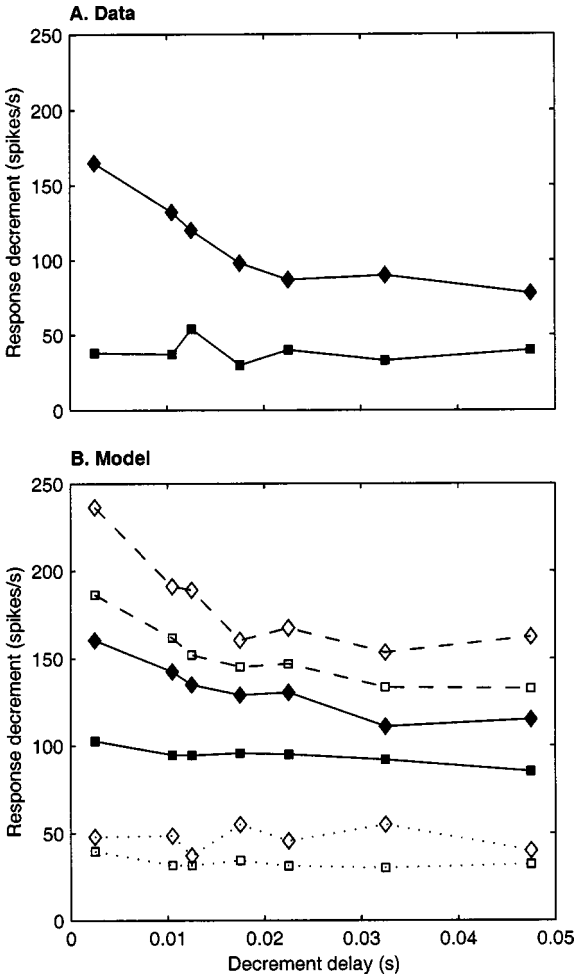


FIG. 7. Decrements in firing rate in response to decrements in stimulus level. (a) Data from Smith *et al.* (1985) showing the decrement in firing rate of onset-window (diamonds) and short-term window (squares) analyses for different decrement delays. The stimulus is a 60-ms BF tone, which starts at 13 dB SL and subsequently drops in level by 6 dB. (b) Results of applying the paradigm of Smith *et al.* to the model fibers. Diamonds show onset-window decrements and squares are short-term window decrements: continuous lines with filled symbols are fibers MSR and H3E; dashed lines with open symbols are fibers HSR and L3G; dotted lines with open symbols are L1 and L3C.

C. Decrements

The change in AN firing rates has also been studied for decrements in stimulus level. Decrement responses were found to be additive for short-term window analysis, but onset window decrements were clearly not additive. Figure 7(a) shows an example of the response of a gerbil AN fiber to decrements (Smith *et al.*, 1985). The stimulus was a 60-ms BF tone, which started at a level of 13 dB SL, and dropped by 6 dB after various intervals. The analysis windows were of the same form as used for measuring increments. The short-term window decrement (continuous line and squares) remains roughly constant with decrement delay. In contrast, the onset-window decrement (continuous line and diamonds) depends strongly on the decrement delay. Smith found this result to hold in nine units. The average ratio of the decrements at 30 to 0 ms was 0.56 for the onset-window decrement and 0.93 for the short-term-window decrement. Figure 7(b) shows the response of the six model fibers to the same paradigm. Squares indicate short-term decrements and diamonds denote onset decrements. The closest result to the data is for the MSR and H3E fiber models averaged together (continuous lines and filled symbols). Both onset and short-term components qualitatively resemble the data, although the departure from additivity for the onset response was not very marked. The model fibers HSR and L3G also gave very similar results and have been averaged together (dashed lines and open symbols). They showed realistic onsets, but did not show additivity for the short-term window analysis. The two LSR remaining fibers (L1, L3C) were also very similar and are shown averaged (dotted lines and open symbols). These showed additivity in both short-term and onset responses. The model is clearly capable of a diverse range of responses, some agreeing well with the data. Averaging the ratios of decrements at 30 to 0 ms for the model fibers HSR, L3G, MSR, and H3E yields 0.69 for the onset response and 0.79 for the short-term responses. L1 and L3C were excluded to make the sample more representative of Westerman's data.

D. Conservation

Westerman (1985; Westerman and Smith, 1987) characterized the transient AN responses for two contiguous 300-ms tone bursts, with the first tone varying in level, and

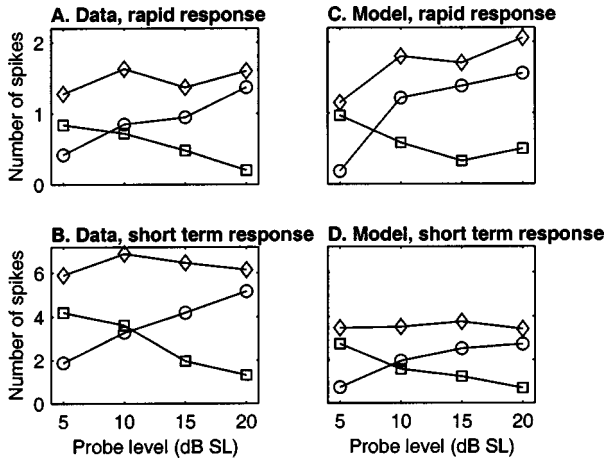


FIG. 8. Conservation of AN adaptation. (a) and (b) the rapid and short-term components averaged across seven gerbil fibers, from Westerman and Smith (1987). A 300-ms background tone of varying sensation level is followed by a contiguous 300-ms increment, fixed at 43 dB SL. Squares show the response of the increment period. Circles show the response during the background time. Diamonds show the total of both background and increment. (c) The average rapid components responses of the four model fibers HSR, MSR, H3E, and L3G. (d) The average short-term components of all the model fibers.

the second tone fixed at a higher level. He found that as the level of the first tone increased, the amount of transient response associated with it also increased. At the same time the transient activity in the second tone decreased. The combined transient response associated with the two tones remained roughly constant. He termed this property ''conservation.'' Figures 8(a) and (b) show the average results of seven gerbil fibers reported by Westerman and Smith. A 43 dB SL tone is preceded by a tone at four different levels (5, 10, 15, or 20 dB SL). The transient responses are characterized by fitting Eq. (6) to each portion of the response separately. The total amount of transient activity associated with rapid and short-term adaptation is then described by $A_r\tau_r$ and $A_{st}\tau_{st}$, respectively. Figure 8(c) shows the number of spikes associated with the rapid component, averaged across the four of the model fibers HSR, MSR, H3E, and L3G. The LSR model fibers L1 and L3C were omitted because they did not show any significant rapid component. The total rapid activity does show some increase with level. Figure 8(d) shows the spikes associated with the short-term component, averaged across all six model fibers. This shows excellent conservation of the total short-term adaptation, although the absolute number of spikes associated with the transient components is lower than the data.

E. Recovery of spontaneous activity

Immediately after the onset of a tone, spontaneous activity of AN fibers is greatly reduced for several tens of milliseconds. The recovery of spontaneous firing can be described by a single exponential function. A range of time constants have been described for different species from 20 to 100 ms, depending on stimulus level (Smith, 1977; Harris and Dallos, 1979; Westerman, 1985). We fitted exponential functions to the recovery of spontaneous activity for the model fibers HSR and H3E. The models were stimulated for 100 ms at 40 dB SPL, as in Hewitt and Meddis (1991). The time constants were 25 ms for the HSR model, and 22 ms for the H3E model. Hewitt and Meddis (1991) reported 30 ms for the original Meddis model. These values are in good agreement with the average value of 20 ms measured by Yates *et al.* (1985) for the guinea-pig, although they are on the low end of the range of values reported across other studies.

V. INNER WORKINGS OF THE MODEL

This section illustrates the operation of the model for two simple cases. The vibration of the basilar membrane gives rise to RP changes that lead to calcium flow into the cell. It is the accumulation of calcium that controls the release of vesicles of transmitter into the synaptic cleft. Both the half-wave rectified RP and the calcium accumulation are subject to low-pass filtering. As a result, constant-intensity, high-frequency tones produce a steady dc response. This is a useful simplification when studying the response at the synapse. We can think of the proximal stimulus at the synapse as a step function at the onset and offset of the stimulus.

Figure 9(a) shows the transmitter flow arrangements. Calcium accumulates in the cell at a rate determined by the RP. This accumulation causes transmitter to be released from the available pool (q) into the synaptic cleft (c). In the cleft, the transmitter has postsynaptic effects and some of it is lost but most is taken back into the cell where it is stored in a reprocessing pool (w). From there it is eventually returned to the available pool. Losses from the cleft are replaced by manufacture of new transmitter.

Figure 9(b) shows the changes in the values of q, c, and w when a high-frequency tone is presented only once against a background of silence. A high-intensity (90 dB SPL) tone and a LSR fiber (L3G) are chosen for clearest presentation. Following tone onset, the transmitter in the available store, q, is reduced and pulses of transmitter can be seen in the synaptic cleft, c, as individual vesicles are released. Very soon, the transmitter in the reprocessing store, w, has increased as the transmitter is recycled back from the cleft. Following tone offset, q recovers partly as the result of input from the reprocessing store and partly from replenishment through manufacture of new transmitter. Toward the end of the tone, there is enough transmitter in the reprocessing store to make four replacement vesicles. However, the available pool must recover from almost zero to ten vesicles. The balance comes from remanufacture at a much slower rate. In our example, the final replacement vesicle does not arrive until 100 ms after the end of the stimulus. The rate of manufacture is given by the expression, $y(M-q)$. As q increases during recovery, this rate becomes smaller.

Figure 9(c) shows a different pattern of response. The stimulating tone is now only 50 dB SPL and the fiber is HSR. Release into the cleft is continuous, even during silence. The resting level of q, the available store, is variable but typically below the maximum, M, of 10. This is an important consideration because ''recovery'' is complete for HSR model fibers when this resting level is reached, not when the maximum level is reached, as was the case for the L3G fiber. The available store does not become fully depleted at this level of

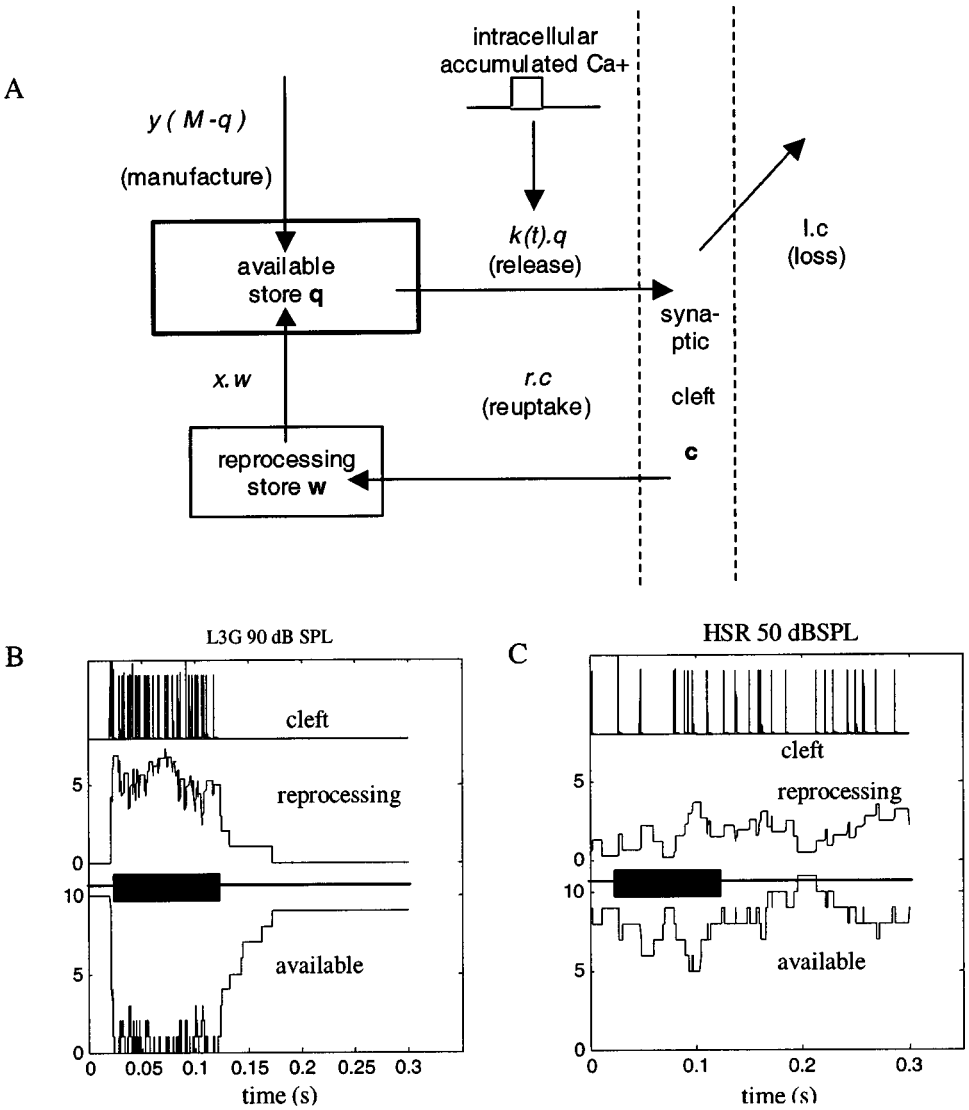


FIG. 9. The inner workings of the model. (a) Schematic of the recycling of neurotransmitter around the internal reservoirs. (b) The contents of the presynaptic reservoirs of the L3G fiber in response to a 100-ms tone burst at 90 dB SPL. (c) The contents of the presynaptic reservoirs of the HSR fiber in response to a 100-ms tone burst at 50 dB SPL. The reservoirs, arranged from top to bottom, are the cleft (c), reprocessing (w), and the immediately available store (q). The black bar shows the presence of the stimulating tone.

stimulation. It shows only a gentle depression after the tone onset. Under these conditions, recovery of the transmitter levels in the available store is mostly sourced from the reprocessing store. Thus differences in adaptation behavior arise from changes in the input to the adaptation mechanism alone.

VI. DISCUSSION

The basic adaptation characteristics of the revised model fall well within the range described by Westerman for the gerbil. The modifications required to make the new model included additional low-pass filtering, quantal stochastic vesicle release, and refractory effects. These changes did not greatly affect the characteristics of adaptation HSR model, and the parameter values for the synapse have changed little from Meddis *et al.* (1990).

Rhode and Smith (1985) and Müller and Robertson (1991) found that LSR fibers show very little adaptation. The data of Relkin and Doucet (1991) suggests that the lack of adaptation observed was due to an insufficient time for recovery between stimuli, and that fully recovered LSR fibers actually show more adaptation than HSR fibers. The model can reproduce the data of Müller and Robertson and Relkin and Doucet even when the simulations are fully recovered for each stimulus presentation. In addition the model can also be affected by interstimulus interval. Although highly speculative, the flexibility of the model raises the possibility that interstimulus intervals might not explain all the differences between observation. It is interesting to note that the three studies all used different species, and different divisions of spontaneous-rate classes. Species differences seem to exist for RI functions. Purely straight RI functions have been associated with LSR fibers in guinea-pig (Winter *et al.*, 1993). We know of no data showing such functions in other species. The different LSR model fibers have quite different RI functions. The L3C model used to model guinea-pig PSTH has a straight, sloping shape, while the L3G fiber's RI function is steep (see Fig. 1). More data may clarify how to set up the model.

The differences in adaptation with fiber type in the model extend to measures of recovery, and stimulus increments and decrements. Overall, the model is not additive enough. However, average data suggests that the issue is not completely clear-cut, and the lack of additivity in the model is not as great as presented previously (Hewitt and Meddis, 1991). The different model fibers show different patterns of

response, and the LSR model fibers especially differ from the responses reported previously. Unfortunately, the data modeled cannot inform us of variation of these characteristics with fiber class, as they are of small numbers of fibers, and naturally biased towards higher spontaneous rates.

It is interesting that the model parameters determining rate characteristics also affect the adaptation characteristics, without any change to the adaptation mechanism itself. Differences in adaptation occur when the driving force to the synapse, or the resting supply of vesicles, varies. Changes in adaptation are to be expected if the difference between fiber types is located before the adaptation mechanism. Thus differences in adaptation between fiber types may offer clues to the location of fiber difference.

ACKNOWLEDGMENTS

This research was supported by the Wellcome foundation (grant ref. 003227), and also by the Consejería de Sanidad of the Junta de Comunidades de Castilla-La Mancha (ref. 01044). The authors would like to thank the two anonymous reviewers for their efforts, and to acknowledge Evan Relkin for his valuable input on the material that was originally submitted as part of Sumner *et al.* (2002).

Harris, D. M., and Dallos, P. (**1979**). ''Forward masking of auditory nerve fiber responses,'' J. Neurophysiol. **42**, 1083–1106.

Hewitt, M., and Meddis, R. (**1991**). ''An evaluation of eight computer models of mammalian inner hair-cell function,'' J. Acoust. Soc. Am. **90**, 904–917.

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.

Meddis, R., Hewitt, M., and Shackleton, T. (**1990**). ''Implementation details of a computational model of the inner hair-cell/auditory-auditory nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1816.

Müller, M., and Robertson, D. (**1991**). ''Relationship between tone burst discharge pattern and spontaneous firing rate of auditory nerve fibers in the guinea-pig,'' Hear. Res. **57**, 63–70.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, New York), 4th ed.

Relkin, E. M., and Doucet, J. R. (**1991**). ''Recovery from prior simulation I: Relationship to spontaneous firing rates of primary auditory neurons,'' Hear. Res. **55**, 215–222.

Rhode, W. S., and Smith, P. H. (**1985**). ''Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers,'' Hear. Res. **18**, 159–168.

Ross, S. (**1996**). ''A functional model of the hair-cell primary fiber complex,'' J. Acoust. Soc. Am. **99**, 2221–2238.

Shamma, S. A., Chadwick, R. S., Wilbur, W. J., Morrish, K. A., and Rinzel, J. (**1986**). ''A biophysical model of the cochlear processing: Intensity dependence of pure tone responses,'' J. Acoust. Soc. Am. **80**, 133–145.

Smith, R. L. (**1977**). ''Short-Term Adaptation in Auditory-Nerve Fibers: Some poststimulatory effects,'' J. Neurophysiol. **40**, 1098–1112.

Smith, R. L., and Zwisloki, J. J. (**1975**). ''Short-term adaptation and incremental responses of single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.

Smith, R. L., Brachman, M. L., and Frisina, R. D. (**1985**). ''Sensitivity of auditory nerve fibers to changes in intensity: A cichotomy between decrements and increments,'' J. Acoust. Soc. Am. **78**, 1310–1316.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). ''A revised model of the inner-hair cell and auditory nerve complex,'' J. Acoust. Soc. Am. **111**, 2178–2188.

Westerman, L. A. (**1985**). ''Adaptation and recovery of auditory nerve responses,'' Special report ISR-S-24, Syracuse University.

Westerman, L. A., and Smith, R. L. (**1984**). ''Rapid and short-term adaptation in auditory nerve responses,'' Hear. Res. **15**, 249–260.

Westerman, L. A., and Smith, R. L. (**1987**). ''Concervation of adapting components in auditory nerve responses,'' J. Acoust. Soc. Am. **81**, 680–691.

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.

Winter, I. M., Palmer, A. R., and Meddis, R. (**1993**). ''The response of guinea-pig auditory nerve fibers with high-spontaneous discharge rates to increments in intensity,'' Brain Res. **618**, 167–170.

Yates, G. K., Robertson, D., and Johnstone, B. M. (**1985**). ''Very rapid adaptation in the guinea-pig auditory nerve,'' Hear. Res. **17**, 1–12.

Factors contributing to bone conduction: The outer ear

Stefan Stenfelt,[a] Timothy Wild,[b] Naohito Hato,[c] and Richard L. Goode
Division of Otolaryngology—Head and Neck Surgery, Stanford University Medical Center, Stanford, California

(Received 2 May 2002; accepted for publication 4 November 2002)

The ear canal sound pressure and the malleus umbo velocity with bone conduction (BC) stimulation were measured in nine ears from five cadaver heads in the frequency range 0.1 to 10 kHz. The measurements were conducted with both open and occluded ear canals, before and after resection of the lower jaw, in a canal with the cartilage and soft tissues removed, and with the tympanic membrane (TM) removed. The sound pressure was about 10 dB greater in an intact ear canal than when the cartilage part of the canal had been removed. The occlusion effect was close to 20 dB for the low frequencies in an intact ear canal; this effect diminished with sectioning of the canal. At higher frequencies, the resonance properties of the ear canal determined the effect of occluding the ear canal. Sectioning of the lower jaw did not significantly alter the sound pressure in the ear canal. The sound radiated from the TM into the ear canal was investigated in four temporal bone specimens; this sound is significantly lower than the sound pressure in an intact ear canal with BC stimulation. The malleus umbo velocity with air conduction stimulation was investigated in nine temporal bone specimens and compared with the umbo velocity obtained with BC stimulation in the cadaver heads. The results show that for a normal open ear canal, the sound pressure in the ear canal with BC stimulation is not significant for BC hearing. At threshold levels and for frequencies below 2 kHz, the sound in the ear canal caused by BC stimulation is about 10 dB lower than air conduction hearing thresholds; this difference increases at higher frequencies. However, with the ear canal occluded, BC hearing is dominated by the sound pressure in the outer ear canal for frequencies between 0.4 and 1.2 kHz. [DOI: 10.1121/1.1534606]

PACS numbers: 43.64.Bt, 43.64.Ha, 43.66.Ba [LHC]

I. INTRODUCTION

Hearing by bone conduction as a physical phenomenon can be divided into three general routes (Tonndorf, 1966):

(1) the sound radiated into the external ear canal, termed the osseotympanic route,
(2) compression and expansion of the petrous bone, which results in displacement of fluid into the cochlea and, consequently, basilar membrane motion, and
(3) the inertial effect of the middle ear ossicles and the inner ear fluids.

It is the first route that is the scope of this investigation, in particular the different parts of the external ear (cartilage and soft tissue part of the canal, bony part of the canal, and the tympanic membrane) which contribute to the sound radiation into the external ear canal. Moreover, the influence of the sound radiated into the external ear canal on the total BC hearing is estimated.

When the head is subjected to a vibration, the vibration is transmitted to the temporal bone of the skull and causes a hearing sensation: this is termed BC sound or hearing by BC. The BC sound makes the skull vibrate relative to the surrounding air, which causes the surrounding air to be compressed and expanded, and an air-borne sound is radiated from the skull. Similarly, with BC stimulation, an air-borne sound is set up in the external ear canal. The slight difference between the sound radiated into the air surrounding the head and into the ear canal is that the ear canal itself is compressed and expanded from the skull vibrations; this distortion of the ear canal walls is the source of the radiated sound in the external ear canal.

Sound radiation into the ear canal with BC stimulation is well known and has been reported extensively in the literature. Berthold was one of the first to objectively present this phenomenon; he used a microphonic flame to show the alternating air pressure in the ear canal when a BC sound was presented to the skull (Bárány, 1938). Later, extensive investigations of ear canal sound pressure with BC stimulation were reported (Bárány, 1938; Huizing, 1960; Békésy, 1960; Elpern and Naunton, 1963; Tonndorf, 1966; Khanna *et al.*, 1976).

Conflicting theories and results on the influence of the ear canal sound pressure (ECSP) on hearing by BC have been reported. Wever and Lawrence (1954) describe the osseo-tympanic stimulation with BC as secondary to the inertial or compressional mode of BC stimulation. Kirikae (1959) stated that inertial effects dominate low-frequency BC hearing and compressional effects dominate high-

[a]Present address: Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Göteborg, Sweden. Electronic mail: stenfelt@s2.chalmers.se
[b]Present address: Department of Head and Neck Surgery, Kaiser Permanente Medical Center, Vallejo, CA.
[c]Present address: Department of Otolaryngology, Ehime University School of Medicine, Shitsukawa, Shigenobu-cho, Onsen-gun, Ehime, Japan.

frequency BC hearing, i.e., ECSP does not have much effect on BC hearing. Allen and Fernandez (1960) believed that the compressional component is the major BC response, while Brinkman *et al.* (1965) argued that the inertial effects on the ossicular chain and cochlear fluids are the major contributor to BC sound. Huizing (1960) found that the sound pressure produced in the external ear canal following BC stimulation was greater for frequencies below 0.5 kHz and less for frequencies above 0.5 kHz, when compared with an AC stimulation giving the same sensation, which indicates that the outer ear contributes to BC hearing for frequencies below 0.5 kHz. Tonndorf (1972) found a major contribution from the external ear component for frequencies below 2 kHz in cats. Khanna *et al.* (1976) showed that when an AC tone was subjectively cancelled by a BC tone, the sound in the external ear canal was also cancelled for frequencies below 0.8 kHz; this indicates that the majority of low-frequency BC sound is transmitted through the external ear canal. However, clinical findings in patients with congenital atresia, cholesteatoma, serous otitis media, stapes otosclerosis, or ossicular discontinuity often show normal or close to normal low-frequency BC thresholds with air-bone gaps of 40 to 60 dB (Ginsberg and White, 1994). These clinical findings indicate that the influence of the ECSP on BC hearing is insignificant.

When the external ear canal is occluded at the opening, a low-frequency increase in subjective BC hearing, as well as in objectively measured ECSP, is obtained (Huizing, 1960; Goldstein and Hayes, 1971). This result indicates that, even if BC hearing is not caused by the ECSP in an open ear canal, the BC hearing is dominated by the ECSP at low frequencies (below 1 kHz) when the ear canal is occluded. The phenomenon of increased low-frequency hearing, after occlusion of the ear canal, with BC stimulation is used in the Bing test: a vibrating tuning fork is applied to the skull and, when the sound is no longer heard, the ear being tested is gently occluded. A perception of the increase of sound in the occluded ear indicates a functional middle ear.

The influence of the lower jaw on the ECSP has also been disputed in the literature. Békésy (1960) argued that the motion of the condyle of the lower jaw, part of which lies against the cartilage and soft tissues of the outer ear canal, yields a motion relative to the skull which results in a sound pressure in the ear canal. Franke (1956) found that the resonance frequency for the lower jaw was somewhere between 110 and 180 Hz, a fact later confirmed by Howell *et al.* (1988). Consequently, below this resonance frequency, the skull and lower jaw move in phase and only minor relative motion arises; in contrast, far above the resonance frequency, the jaw is almost at rest which gives large relative motion between the jaw and skull. Tonndorf (1966), who ascribed this phenomenon a minor role for the sound pressure in the ear canal with BC stimulation, argued that the major radiation was from the bony part of the ear canal. In an experiment on the influence of the jaw with BC, it was concluded that ECSP was present in the ear canal even after the jaw had been removed (Howell and Williams, 1989).

The aim of this study is to investigate the relative importance of contributions to ear canal sound pressure (ECSP) with BC stimulation, from four parts of the external ear canal: (1) the soft tissue part of the ear canal, (2) the bony part of the ear canal, (3) the tympanic membrane (TM), and (4) the condyle of the lower jaw. In addition, the effect on the ECSP with BC stimulation after occlusion of the ear canal is measured. By using measurements of the ECSP together with the umbo motion, for BC as well as air conduction (AC) stimulation, the influence of the osseo-tympanic route on BC hearing is estimated.

TABLE I. Data on the five cadaver heads used in the study. The circumference is measured in a line across the middle of the forehead, just above the ear canal openings and across the occiput.

No.	Sex	Age (years)	Circumference (cm)	Ear–ear via vertex (cm)	Mass (kg)
1	M	60–70	58	36	3.78
2	M	60–70	56	31	3.43
3	M	60–70	53	31	3.25
4	M	60–70	54	33	3.31
5	M	60–70	57	34	3.59

II. MATERIALS AND METHODS

A. Whole head experiments

The ear canal sound pressure (ECSP) was measured in nine ears from five human cadaver heads. The heads had been severed between the third and fourth vertebra and were frozen at time of autopsy. Apart from being severed, the heads were left intact, i.e., there was no cutting or drilling of the heads during the autopsy, and the soft tissues and brain were also left intact. The size and weight of the heads are presented in Table I. No history of the heads was known, except that they were male and between 60 and 70 years old. They were defrosted 24 h prior to the measurements. At the time of measurement, the ear was examined with an operation microscope; crust and hair in the ear canal was removed, and the tympanic membrane (TM) was inspected. One ear was not used in this investigation since it had a perforated TM. Small reflective targets (glass sphere, Ø 5 μm) were placed on the tip of the malleus umbo and on the bony wall of the ear canal [Fig. 1(a)]. These targets were used for the velocity measurements made with a laser Doppler vibrometer (LDV). A hole was drilled in the bone and tapped, 35 mm posterior to the ear canal opening, and a transducer was rigidly attached to the skull by a threaded connector attached to the threaded hole. The transducer, referred to as the mini-transducer, was a remodeled bone anchored hearing aid transducer with screw attachment (Tjellström *et al.*, 2001). During the measurements, the head was placed on an inflatable pillow to avoid artifacts from the measurement table.

An 8-mm-long tapered plastic speculum, with an inner-end diameter of 6 mm and outer-end diameter of 13 mm, was inserted into the ear canal [Fig. 1(a)]. The speculum intruded 3 to 5 mm into the ear canal and was surrounded by Vaseline to provide a good sound seal. A glass cover fitted into the middle of the speculum was used to occlude the ear canal; the glass cover provided a sound seal and also transmitted the laser beam. A probe tube microphone (ER-7C, Etymotic Research, Elk Grove Village, IL) was inserted into the ear

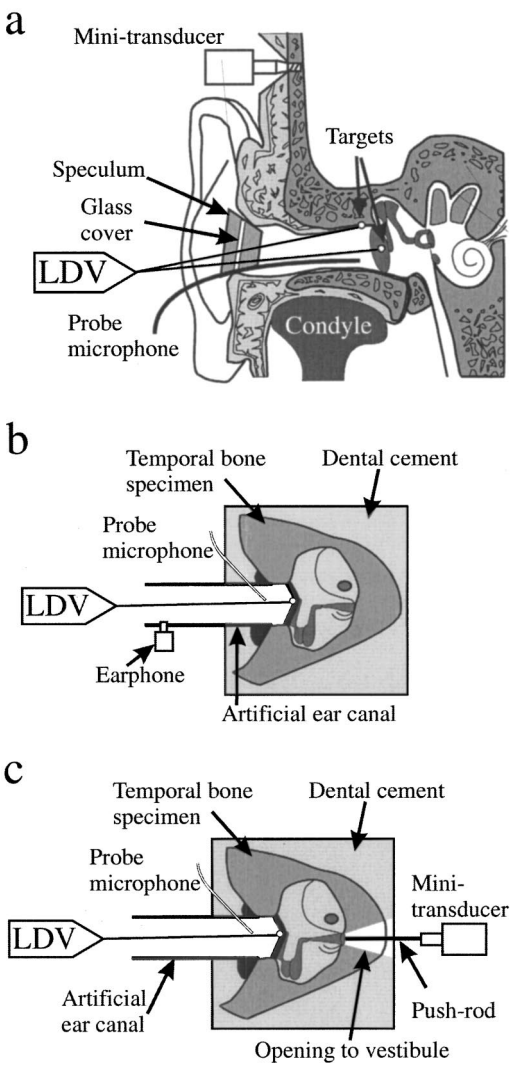


FIG. 1. (a) Measurement setup for whole head measurements with BC stimulation. The BC stimulation is transmitted to the skull by a mini-transducer rigidly coupled to the skull 35 mm behind the ear canal opening. A speculum, placed at the end of the ear canal, can put a glass cover in place as an occlusion device, and a probe tube microphone is positioned 2 mm in front of the TM. The velocity of the skull bone and the malleus umbo is measured by a LDV, and reflective targets are used to enhance the reflection of the laser beam. The condyle of the lower jaw is positioned below the ear canal to improve the visibility of ear structures. (b) Setup for measurements of the malleus umbo velocity with AC stimulation. The temporal bone specimen is sealed in a block of dental cement; an artificial ear canal is glued onto the specimen perpendicular to the plane of the TM. An earphone provides the sound stimulus at the lateral end of the artificial ear, the ECSP is measured 2 mm in front of the TM by a probe tube microphone, and the velocity of the malleus umbo is measured by a LDV. (c) Setup for measurement of the ECSP and umbo velocity when the ossicles are stimulated in reverse. The temporal bone specimen is sealed in a block of dental cement; an artificial ear canal is glued onto the specimen perpendicular to the plane of the TM. The stimulation is provided by a mini-transducer extended by a rod that is glued to the vestibular side of the stapes footplate. The ECSP is measured 2 mm in front of the TM by a probe tube microphone, and the velocity of the malleus umbo is measured by a LDV.

canal and positioned about 2 mm from the TM; this microphone was used to measure the ECSP. The vibrations of both the bony wall of the ear canal and the malleus umbo were measured with the HLV-1000 laser Doppler vibrometer (Polytec, Waldbronn, Germany). The sensor head of the HLV-1000 was mounted with a joystick-controlled mirror on an operating microscope, which enabled easy control of the laser beam. The stimuli were provided by a PC-based software, SYSid 6.5, using a DSP-16+ signal processing card (www.sysid-labs.com). The output from the computer was fed through a power amplifier (D-75, Crown, Elkhart, IN) to the mini-transducer.

B. Temporal bone experiments

The temporal bones were extracted from human cadavers, within 48 h of death, using a Schuknecht bone saw at the time of autopsy. The temporal bone specimens were wrapped in gauze, placed in a 1:10 000 merthiolate solution in normal saline and stored at 5 °C. All measurements on individual bones were conducted on the same day within 6 days after death. The TM and middle ear were inspected for each bone using an operating microscope; bones with abnormal TMs or middle ears were excluded from the investigation. Thirteen temporal bones were studied; they were from 12 males and 1 female, with an average age of 60.2 years and a range from 52 to 75 years. Connective tissue and muscle were removed and the bony external ear canal was drilled down to 2 mm from the tympanic membrane annulus. The artificial external ear canal assembly (8.5 mm internal diameter, 25 mm long) used contained an earphone adapter on the side near the lateral end and a probe tube opening 2 mm from the medial end. This assembly was attached to the bony rim of the ear canal with clay, so that the axis of the tube was perpendicular to the annulus, while the remainder of the temporal bone was embedded in Hydrock dental cement (Kerr Co., Romulus, MI), forming a solid airtight specimen block.

A reflective target (glass sphere, Ø 5 μm) placed on the malleus umbo increased the reflection for the laser measurement. An ER-7C probe tube microphone was inserted into the artificial ear canal with the probe tube opening 2 mm from the TM. The same SYSid measurement system was used for the temporal bone measurements; however, when measuring AC umbo motion, the output from the power amplifier was fed to an earphone (83-13A/024, Tibbets Industries, Camden, ME) that was inserted in the adapter in the artificial ear canal. The ECSP in the artificial ear canal was measured with the probe tube microphone and the umbo motion was measured with the HLV-1000 vibrometer [Fig. 1(b)].

When the sound radiation from the TM was measured, the mini-transducer was extended by a 10-mm-long and 0.5-mm-diam rod [Fig. 1(c)]. The temporal bone was opened from the internal ear canal, and a hole into the vestibule was made so the stapes footplate was clearly visible. After the cochlea was drained, the mini-transducer with the extended rod was inserted with a micro-manipulator into the hole in the vestibule. A drop of cyanoacrylate glue (Loctite 430, Loctite, Rocky Hill, CT) was put on the tip of the rod; when the rod made contact with the footplate, they were bonded without adding any stress on the annular ligament. With this assembly, the ECSP due to TM vibration alone was measured. The output from the computer was fed to the mini-transducer, the ECSP was measured with the probe micro-

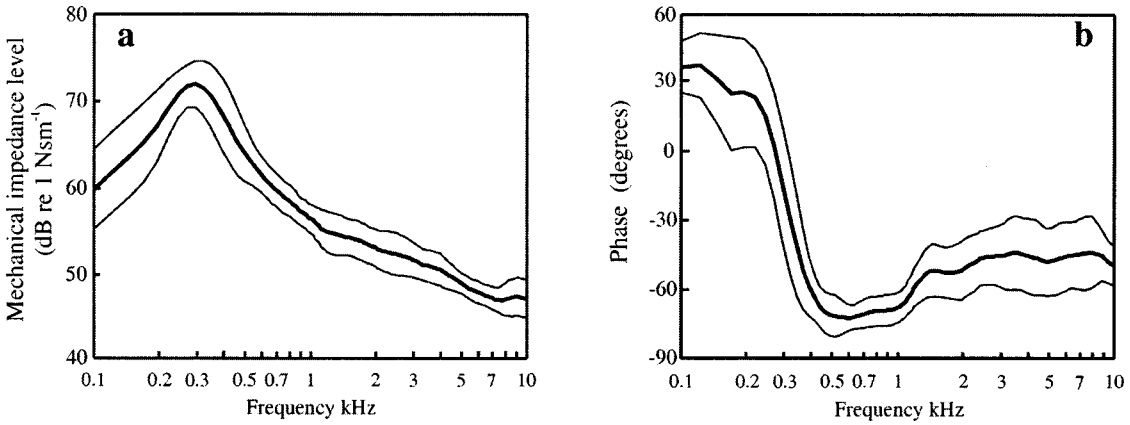


FIG. 2. Level (a) and phase (b) of the mechanical point impedance of the bone screw attached to the mastoid of the skull. The impedance level is defined as $20 \cdot \log_{10}$ of the magnitude data. The position was approximately 35 mm behind the ear canal opening. The thick line represents the average values from nine positions (nine ears from five heads) and the thin lines are ±SD. Frequency resolution is 50 points/decade.

phone 2 mm in front of the TM, and the malleus umbo motion was measured with the HLV-1000 vibrometer.

C. Calibration

The calibration of the HLV-1000 vibrometer and the ER-7C probe microphone is described in Stenfelt *et al.* (2002). The mini-transducer was calibrated on a Skull Simulator TU-1000 to obtain the correct force output (Håkansson and Carlsson, 1989). Provided the point impedance of the connection point on the head is much larger than the output impedance of the mini-transducer, the Skull Simulator TU-1000 gives correct force data. To verify that the point impedance at the bone screw used to attach the mini-transducer to the skull was similar to that of a normal living head, the mechanical point impedance of the bone screw *in situ* was measured. The mechanical impedance data were obtained by a B&K type 8001 impedance head connected to a threaded adapter for rigid attachment between the impedance head and the bone screw. The apparent mass (force/acceleration) of a known mass, approximately 30 times the mass above the force gauge, was measured in order to calibrate the impedance head. The apparent mass of the impedance head with the adapter unloaded was then measured; these data were used to compensate for the mass above the force gauge in the postprocessing of the impedance data. Figure 2 gives the mechanical impedance for the nine positions (five heads). The results are similar to impedance measurements conducted on living human heads with permanent titanium fixtures at a similar position (Håkansson *et al.*, 1986).

III. RESULTS

A. A Normal ear canal

The average ear canal sound pressure produced by BC stimulation in nine ears with intact ear canals is presented in Fig. 3. The results are shown for both open and occluded ear canals. With the ear canal open, the level of the ECSP is approximately flat for frequencies below 500 Hz; above this frequency the sound pressure rises by approximately 15 dB per octave until the frequency reaches the resonance frequency of the ear canal at 2.7 kHz (quarter wavelength resonance). With the ear canal occluded, the level of the ECSP is some 10 to 15 dB greater at low frequencies. The difference between the ECSP for occluded and open ear canals starts to diminish at 1 kHz and is zero at 2 kHz. The quarter wavelength resonance at 2.7 kHz becomes an antiresonance when

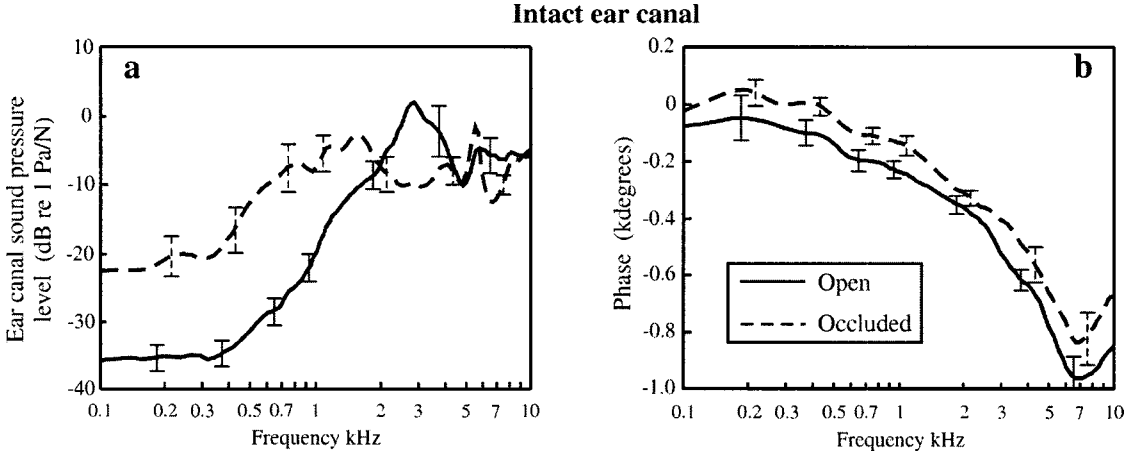


FIG. 3. Average level (a) and phase (b) of the sound pressure in an intact ear canal when the BC stimulation was at the mastoid. The solid line shows the result in open ear canals, while the dashed line represents occluded ear canals. The vertical bars indicate ±1 standard error of the mean. Frequency resolution is 50 points/decade.

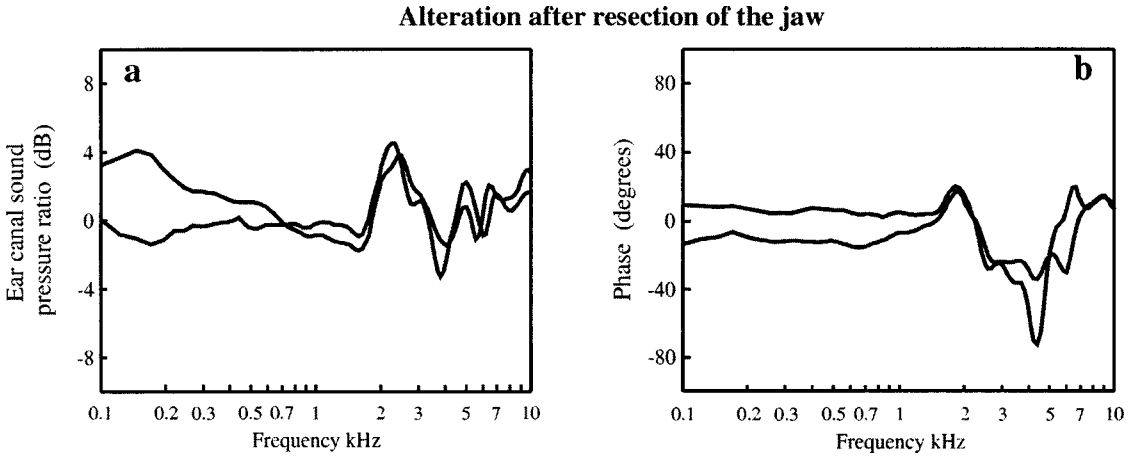


FIG. 4. Alteration of the level (a) and phase (b) of the ECSP in the intact open ear canal after resection of the lower jaw, when the BC stimulation was at the mastoid. The data are obtained for two ears from one head. Frequency resolution is 50 points/decade.

the ear canal is occluded; a half wavelength resonance around 5.5 kHz is visible in about half of the ears tested. An average sound pressure increase can therefore be found around 5.5 kHz with the ear canal occluded. The bars in the figure show the standard error of the mean, which is 2 to 5 dB in the frequency range 0.1 to 10 kHz.

The phase of the ECSP relative to the input force, for an open ear canal, shows a decaying response, indicating a time delay [Fig. 3(b)]. This time delay for the ECSP consists of two components: one is the traveling time for BC vibration in the skull, and the other is the time delay of the sound pressure wave traveling from the radiation point in the ear canal to the probe tube microphone. These two components cannot be separated by the current measurement. The ECSP phase shows similar behavior for both open and occluded ear canals; however, the phase with the ear canal occluded leads that of an open canal by about 180 degrees.

B. Alteration after resection of the jaw

For one head, after the measurement of the ECSP in intact ear canals, the condyles of the lower jaw were resected. The jaw was pulled so there was about 1 cm of air space between the condyle and the cartilage and soft tissue of the ear canal at both ears. Thereafter, the ECSP in both ears was remeasured with the ear canals open and occluded. The level and phase alterations of the ECSP for the two ears with open ear canals after resection of the jaw are shown in Fig. 4. In one ear, the ECSP difference is within 1 dB for frequencies below 2 kHz, while the other ear shows an increased ECSP of 4 dB, at the low frequencies, which falls off with frequency and vanishes at 1 kHz. Above 2 kHz, the ECSP after resection of the jaw increases and peaks at a level of 4 dB at 2.3–2.5 kHz. Both traces show a decrease in ECSP levels of 2–3 dB at 4.0 kHz and have other peaks of 1 and 2 dB at 5 kHz. The resonance pattern continues at the higher frequencies with an ECSP decrease around 6.0 kHz and another increase around 7.0 kHz. The increases and decreases seen at the higher frequencies are within 2 dB.

The phase alteration of the ECSP after resection of the jaw is around +10 to −10 degrees at the low frequencies. At 2.0 kHz, a peak of 20 degrees in phase alteration is seen in both ears. Above this frequency the phase drops to a minimum at 4.3 kHz: −35 and −70 degrees for the two ears. At the higher frequencies, the phase difference decreases and becomes approximately 10 degrees at 10 kHz. The results for the alteration in level and phase were similar whether the ear canal was open or occluded.

C. Alteration of the ear canal

About one-third of the ear canal consists of cartilage; when this was removed, an approximately 15–20-mm-long bony ear canal remained. The ECSP results with BC stimulation at the mastoid after the pinna, cartilage, and soft tissue parts of the ear canal were removed are given in Fig. 5. Compared with the intact ear canal (Fig. 3), the ECSP level is 5 to 10 dB lower for frequencies below 1.0 kHz and 10 to 15 dB lower for frequencies between 1.0 and 4.0 kHz, with the ear canal open. With the ear canal occluded, the sound pressure in the bony ear canal remnant is about 15 dB lower at frequencies below 3 kHz than in an intact occluded ear canal. The occlusion effect with the cartilage part of the ear canal removed is 5 to 10 dB for frequencies below 2 kHz, i.e., about 5 dB less than in an intact ear canal. Figure 5(b) shows the phase of the ECSP relative to the input force at the bone screw in the mastoid. The phases are similar for both open and occluded ear canals. Compared with the phase for the intact ear canal in Fig. 3(b), the phase for the ear canal with the cartilage removed decreased more rapidly.

The data in Fig. 6 were obtained after the cartilage, soft tissue, and TM of the ear canal had been removed. The levels in Fig. 6(a) are somewhat higher than the results obtained with only the cartilage removed in Fig. 5(a): with an open ear canal, the levels are 5 to 10 dB greater for frequencies between 1.5 and 4 kHz, with a tendency to resonance around 3 kHz. This resonance is a result of the longer effective ear canal length after the TM is removed; the middle ear space is a part of the total canal length. Occluding the ear canal at the lateral end of the bony part increases the ECSP by up to 5 dB between 0.1 and 1.0 kHz and 5 to 10 dB for frequencies between 1.0 and 2.0 kHz. Between 2.0 and 5.0 kHz, the occlusion lowers the ECSP by up to 10 dB.

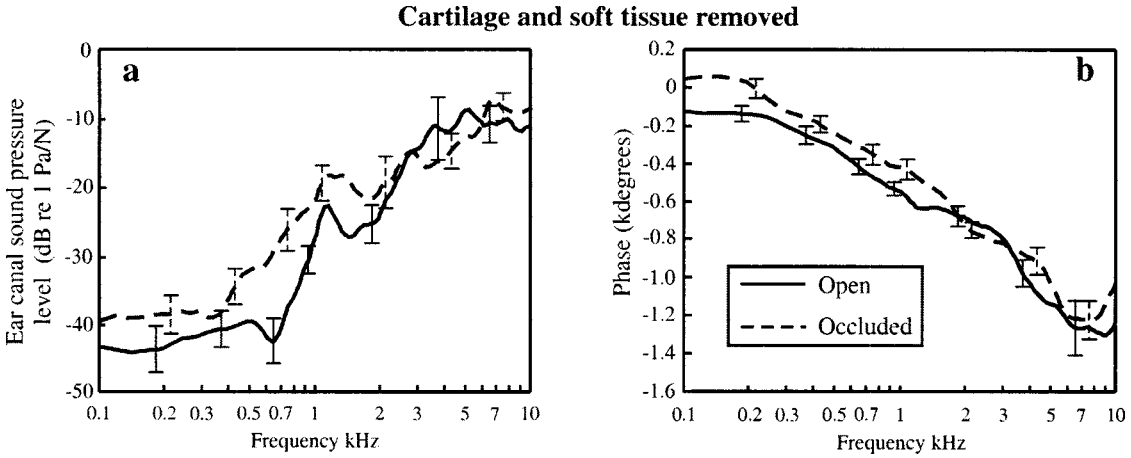


FIG. 5. Average level (a) and phase (b) of the sound pressure in the bony ear canal after removal of the ear canal cartilage and soft tissue. The BC stimulation is applied at the mastoid. The solid line shows the result for an open ear canal, while the dashed line is for an occluded ear canal. The vertical bars indicate ±1 standard error of the mean. Frequency resolution is 50 points/decade.

Figure 6(b) displays the phase of the results after the cartilage, soft tissues, and TM were removed from the ear canal. The phases obtained are similar to the ones obtained with the intact ear canal shown in Fig. 3(b). For the three conditions measured in Figs. 3, 5, and 6, the variance of the measurements is similar: the standard error of the mean is within 5 dB for the frequency range 0.1 to 10 kHz. The standard error of the mean for the phase in the same figures is also similar: 10 to 100 degrees in the frequency range 0.1 to 10 kHz.

D. Ossicles driven in reverse

The sound radiated from the TM alone was assessed by driving the ossicular chain in reverse, with a small vibrator (mini-transducer) attached to the vestibular side of the stapes footplate by a thin rod, while measuring the ECSP and the malleus umbo motion. First, the motion of the temporal bone specimen itself was measured by the LDV with a target on the bony annulus close to the TM. The measurement verified that the motion of the temporal bone specimen itself is 40 dB lower than the umbo motion at 100 Hz; the difference between the velocities of the umbo and temporal bone increased with frequency. Above 1.0 kHz the velocity of temporal bone became immeasurable.

The measurements with the ossicles driven in reverse were conducted in temporal bone specimens at rest, whereas the ECSPs with BC stimulation were measured in vibrating skulls. Therefore, to facilitate comparing the results, the ECSP is referenced to the relative velocity of the malleus umbo. The relative umbo velocity is defined as the umbo velocity minus the bone velocity measured in the bony ear canal, i.e., $V_{\text{rel,umbo}} = V_{\text{umbo}} - V_{\text{bone}}$. For BC stimulation, the velocity difference between the umbo and the bony part of the ear canal was used to compute the relative umbo velocity; whereas, with the reverse stimulation of the middle ear ossicles, the relative umbo velocity was the same as the umbo velocity since the temporal bone was considered to be at rest.

Figure 7 shows the ECSP level and phase relative to the umbo velocity with an open ear canal for three conditions: (1) an intact ear canal with BC stimulation of the head, (2) the cartilage part of the ear canal removed, with BC stimu-

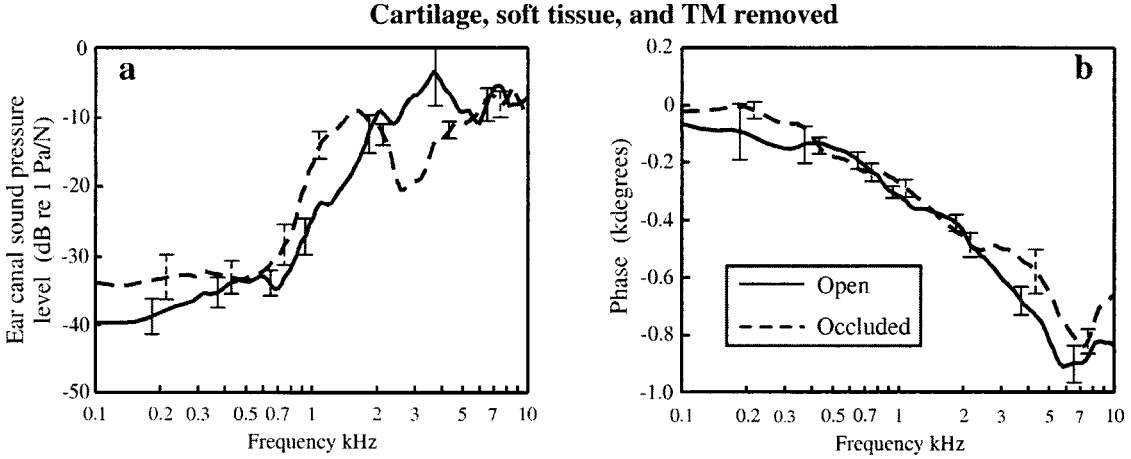


FIG. 6. Average level (a) and phase (b) of the sound pressure in the bony ear canal after removal of the ear canal cartilage, soft tissue, and the TM. The BC stimulation is applied to the mastoid. The solid line is the result with the ear canal open, while the dashed line is with the ear canal occluded. The vertical bars indicate ±1 standard error of the mean. Frequency resolution is 50 points/decade.

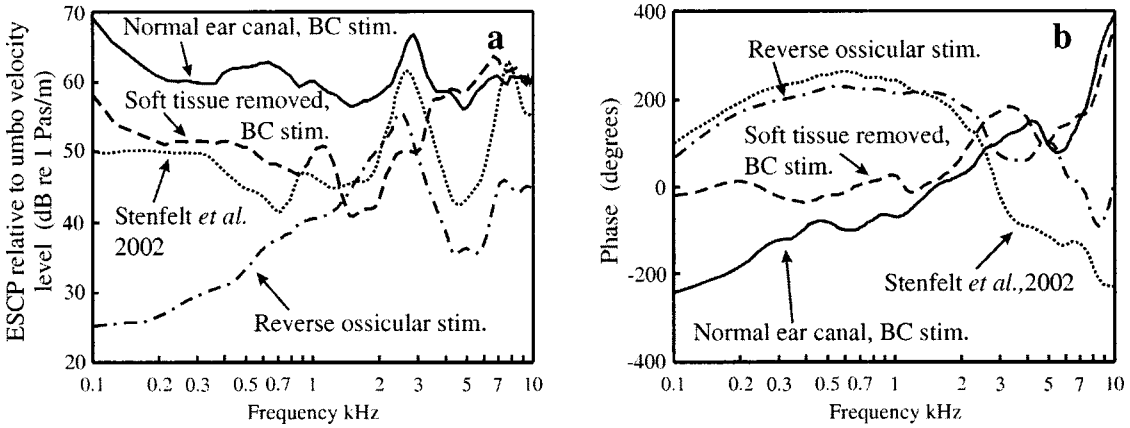


FIG. 7. Average level (a) and phase (b) of ECSP relative to umbo velocity in an open ear canal. The straight line shows the result with a normal intact ear canal and BC stimulation (mean of nine ears), the dashed line shows the result when the cartilage and soft tissue of the ear canal have been removed and stimulated by BC (mean of nine ears), and the dash-dotted line shows the result from the TM radiation into an artificial ear canal when the middle ear ossicles are stimulated in reverse (mean of four temporal bone specimens). Also included are the ECSP relative to umbo velocity data from Stenfelt *et al.* (2002) in temporal bone specimens stimulated by shaking the entire specimen (dotted line). Frequency resolution is 50 points/decade.

lation of the head, and (3) TM radiation into an artificial ear canal in a temporal bone specimen when the ossicles are driven in reverse. For comparison, data from Stenfelt *et al.* (2002) for the ECSP in temporal bone specimens when the whole specimen is shaken is also included in Fig. 7.

The greatest ECSP relative to umbo velocity is obtained with BC stimulation in an intact ear canal. The ECSP level relative to umbo velocity, due to TM radiation alone, increases by approximately 6 dB/oct for frequencies between 0.2 and 2.0 kHz and resonates around 2.5 kHz; this resonance frequency corresponds to the quarter wavelength resonance of the total length of the artificial ear canal, including the remnant of the bony canal. Although the ECSP drops above the resonance frequency, it recovers around 7 kHz, which is a result from the three-quarter wavelength resonance. Figure 7 shows that the radiation from the TM into the ear canal is significantly lower than the ECSP with BC stimulation in an intact ear canal, except near the ear canal resonance, where the TM radiation is only 5 to 10 dB lower than the ECSP with BC stimulation. The result for the bony canal is not directly comparable with that of the TM radiation, since the resonances at the high frequencies differ, due to the canal length, for the two. In addition, at low frequencies, the acoustic impedance into the ear canal seen from the TM is influenced by the ear canal length.

The phase of the ECSP divided by the umbo relative velocity reveals differences between reverse ossicular stimulation in a temporal bone specimen and BC stimulation in a skull [Fig. 7(b)]. With reverse ossicular stimulation, the ECSP phase increases from 80 degrees at 0.1 kHz to 220 degrees at 0.5 kHz. This phase stays close to 200 degrees up to 2.0 kHz, after which it falls off to −100 degrees at 8 kHz; however, it recovers at the very high frequencies. With BC stimulation in a head, the phase of the ECSP divided by the relative umbo velocity increases from −220 degrees at 0.1 kHz to 150 degrees at 4.5 kHz. It then decreases slightly, after which it starts to rise rapidly above 6.0 kHz and becomes 400 degrees at 10 kHz. The phase of the ECSP with BC stimulation, after removal of the cartilage and soft tissue in the ear canal, stays close to zero degrees at low frequencies; above 1.2 kHz, it becomes similar to the phase of the ECSP in an intact ear canal with BC stimulation.

E. The occlusion effect

The occlusion effect on the ECSP with BC stimulation, (1) in an intact ear canal, (2) when the cartilage and soft tissue part of the ear canal is removed, and (3) in the bony canal when the TM is removed, is shown in Fig. 8. In addition, a curve for the occlusion effect in the artificial ear canal when the ossicles are driven in reverse is included in the figure. The strongest effect of occluding the ear canal is obtained with TM radiation alone in an artificial ear canal: an increase of over 20 dB at 100 Hz that falls off at about −6 dB/oct. The attenuation of about 20 dB around 3 kHz is mainly due to the quarter wavelength resonance for the open canal, which disappears when the canal is occluded; the increase of 15 dB at 5.5 kHz results from the half wavelength

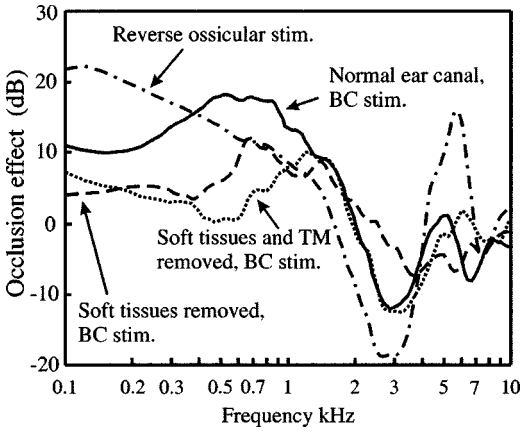


FIG. 8. The occlusion effect computed as the sound pressure level in the occluded ear canal relative to the sound pressure level in the open ear canal. Solid line: normal intact ear canal and stimulation by BC (mean of nine ears). Dashed line: cartilage and soft tissue of the ear canal removed and stimulation by BC (mean of nine ears). Dotted line: cartilage, soft tissue, and TM of the ear canal removed and stimulation by BC (mean of nine ears). Dash-dotted line: Sound radiation from the TM into an artificial ear canal when the ossicles were driven in reverse (mean of four temporal bone specimen). Frequency resolution is 50 points/decade.

resonance in an occluded canal. The curve for the intact ear canal with BC stimulation is similar to the one with TM radiation in an artificial ear canal. That the curves are not the same may be a result from the different geometries of the two canals: the artificial canal is straight whereas the normal ear canal is curved. Also, the damping imposed by the soft part of the intact ear canal can cause divergence between the two; the artificial ear canal is a hard plastic tube with little or no damping. When either the cartilage part of the canal or the TM is removed, the occlusion effect is less than for an intact ear canal.

IV. DISCUSSION

A. The ECSP with BC stimulation

The relative contributions of three factors to the ECSP with BC stimulation were investigated: (1) the cartilage and soft tissue part of the ear canal, (2) the bony part of the ear canal, and (3) the TM. The greatest ECSP in an open ear canal with BC stimulation is obtained in an intact ear canal (Fig. 3); removing the cartilage and soft tissue of the ear canal reduced the ECSP level 5 to 10 dB for frequencies below 1.0 kHz, 10 to 15 dB for frequencies between 1.0 and 4.0 kHz, and some 5 dB for higher frequencies. The reduction of the ECSP level at frequencies below 2.0 kHz, with removal of the cartilage and soft tissue, indicates that the cartilage part of the ear canal is the greatest source of the ECSP with BC stimulation (Fig. 5). It is difficult to compare the results at higher frequencies, since the shorter bony canal has resonance properties that differ from those of the longer intact canal.

The bony part of the ear canal has sometimes been considered the origin of the ECSP with BC stimulation (Huizing, 1960; Tonndorf, 1966). However, for frequencies below the first resonance of the skull (0.8–1.0 kHz, Håkansson *et al.*, 1994), the skull bone moves as a rigid body; there is no compression and expansion in the bony part of the external ear canal. The cartilage and soft tissue part of the ear canal are more compliant than the bone and can cause compression and expansion of the ear canal at frequencies far below the resonance frequency of the skull. This is a further indication that the cartilage and soft tissue part of the ear canal generates the ECSP with BC stimulation at the low frequencies, as suggested by Naunton (1963). Also, the finding that occlusion of the bony part of the ear canal minimizes the perceived occlusion effect concurs with the hypothesis that the soft tissue part of the ear canal dominates the ECSP with BC stimulation at the low frequencies (Békésy, 1941).

With a shorter ear canal, the resonance frequencies are altered: after the cartilage and soft tissue are removed, the quarter wavelength canal resonance should be between 4.5 and 5 kHz (corresponds to a length of approximately 17–20 mm). However, Fig. 5(a) does not show any clear resonance around that frequency for an open ear canal. An explanation is that the length of the bony ear canal differed among the measured ears, and the effect of the resonance disappears after averaging the results. This is supported by the individual data that show resonances between 4.0 and 5.5 kHz. The spread in resonance frequencies as a result of variation in the ear canal length cannot be verified, since ear canal length was not measured.

B. The ECSP from the TM

The contribution to the ECSP from the TM was assessed in two ways: the TM was removed, and the ECSP was measured with BC stimulation in the whole head; also, the ECSP from the TM was measured in temporal bone specimens when the ossicles were stimulated in reverse. An increase in ECSP level is found after removing the TM from the bony canal in the whole head experiments (Fig. 6); the ECSP levels are still lower than in the intact ear canal. Some of the difference in the ECSP after removal of the TM can be explained by the change in acoustic properties for the lengthened canal and by the fact that the position of the probe tube microphone opening is not altered; this opening is approximately in the middle of the new canal which encompasses both the bony part of the ear canal and the middle ear cavity. The measurement position in the ear canal can affect the ECSP obtained. However, the ECSP as a function of canal position was not investigated.

The increase in measured ECSP after removal of the TM from the bony canal can also be a result of sound radiation into the middle ear cavity. After removal of the TM, the sound radiated into the middle ear cavity increases the measured ECSP. This effect with BC stimulation was proposed by Groen (1962) to be a major contribution to hearing by BC at around 2.5 kHz. Stenfelt *et al.* (2002) did not find any significant sound radiation into the middle ear canal when temporal bone specimens were shaken. Tonndorf (1966), who measured the BC response in cats with both open and closed middle ear cavities, found no difference in the response. When he removed the TM, he found only small and nonsystematic changes of the ECSP. Hence, there is no support, either in the Tonndorf (1966) data on cats or in the Stenfelt *et al.* (2002) data on temporal bone specimens, for a contribution to hearing by BC from sound radiation into the middle ear cavity.

When the ossicles are driven in reverse, the low-frequency ECSP relative to umbo velocity level is 10 to 30 dB lower than in the bony canal with BC stimulation (Fig. 7). Since driving the ossicles in reverse causes an ECSP due only to the motion of the TM, the greater low-frequency ECSP obtained in the bony ear canal indicates that this sound pressure is caused by the bony part of the ear canal. However, for frequencies below the skull resonance, there is no expansion and compression in the skull bone, hence there should not be any sound radiation into the ear canal. The ECSP from temporal bone specimens with artificial ear canals (plastic tubes, the same as in this study) is included in Fig. 7; the specimens were shaken to simulate BC stimulation (Stenfelt *et al.*, 2002). The Stenfelt *et al.* (2002) data show low-frequency ECSP levels similar to those in the bony canal, however they also show phase data similar to that from stimulation of the ossicles in reverse. If the low-frequency ECSP from BC stimulation, as hypothesized, is caused by the TM in the bony ear canal, the three conditions, (1) bony ear canal in the skull with BC stimulation, (2) the ossicles driven in reverse, and (3) the shaking of a temporal

bone specimen with an artificial ear canal (Stenfelt *et al.*, 2002), should show similar low-frequency ECSP data. This is apparently not the case (Fig. 7). However, at least two of the phenomena involved in the three conditions differ, which could, at least partly, explain the discrepancies: (1) the inertia effects of the TM itself and (2) the sound radiated into the surrounding air and subsequently transmitted to the open (artificial) ear canal.

The TM has a small but nonzero mass (14 mg, Wever and Lawrence, 1954) which gives rise to a distributed force that acts on the TM for BC stimulation or when the temporal bone specimen is shaken. At the low frequencies, this force on the TM acts in phase with the inertial forces of the ossicles, which gives higher ECSP levels than if the TM were driven only by the ossicles. Since they act in phase, the phase of the ECSP would be the same. When an object is vibrating, it radiates sound into the surrounding air, i.e., both shaking a temporal bone specimen and stimulating a head by BC generates a sound in the surrounding air. This sound can be transmitted to the open ear canal. Since the head is a larger structure than the temporal bone specimen, it generates greater low-frequency sound for the same amount of bone vibration. The source of this sound is spatially different from the TM, and the resulting ECSP caused by sound radiation into the surrounding air has a phase different from the sound radiated from the TM. It should be noted that an occlusion of the ear canal removes this sound source from the ECSP. At low frequencies, the occlusion effect is lower for ear canals in the head with BC stimulation than for the reverse ossicular stimulation. This difference can be caused by the sound radiation into the surrounding air being transmitted to the open ear canal but not the occluded canal (Fig. 8).

The difference between low-frequency ECSP from the reverse ossicular stimulation and from the shaking of the temporal bone in Fig. 7 can be explained in part by the inertial effect of the TM: the ECSP level from shaking of the temporal bone specimen is greater than, but the ECSP phase is similar to, that of the reverse ossicular stimulation. Similarly, the low-frequency ECSP difference between the bony canal and the reverse ossicular stimulation can be a combination of sounds, one of which radiates from the head into the surrounding air and is transmitted to the open ear canal, while the other is the inertial effect of the TM: this means a greater low-frequency ECSP level for the bony canal and a different phase, when compared with the reverse ossicular stimulation. At the higher frequencies, the ECSP can be generated in the bony part of the ear canal, which is indicated by the similarity, in both magnitude and phase, of the intact ear canal data and data from the ear canal with the cartilage and soft tissue removed (Fig. 7).

C. Occlusion effect

The ECSP increase for an occluded intact ear canal is similar to the results of Huizing (1960), except at frequencies below 0.4 kHz where our data is about 10 dB lower. Huizing (1960) occluded, with a rubber plug, the outer end of the ear canal, a position similar to that used in this study. Elpern and Naunton (1963) showed that the occlusion effect depends on the type of device chosen: small supra-aural earphones gave

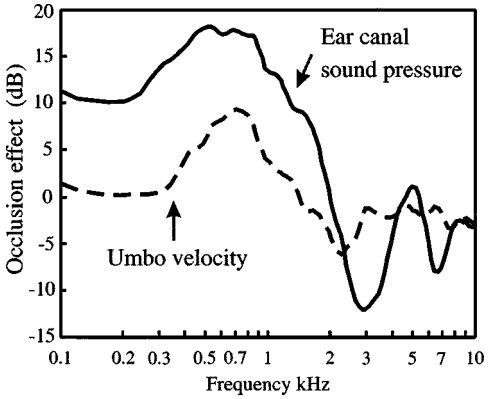


FIG. 9. Increase of the ECSP level (solid line) and umbo velocity level (dashed line) with BC stimulation after occlusion of an intact ear canal (mean of nine ears). Frequency resolution is 50 points/decade.

the greatest occlusion effect, whereas earphones containing large air volumes did not cause any occlusion effect. Békésy (1941) found that when the occlusion was deep seated, i.e., all the way down to the bony part of the ear canal, the occlusion effect disappeared. Figures 3 and 5 show that the ECSP with BC stimulation in an occluded bony ear canal is similar to the ECSP in an open intact ear canal. Thus, in an intact ear canal, an occlusion down to the bony part of the ear canal does not increase the ECSP with BC stimulation.

Onchi (1954) described two peaks in the occlusion effect of the ear canal: one at 200 Hz which he ascribed to the resonance of the ossicular chain, and the other at 800 Hz which was due to a resonance of the TM. The resonance frequency of the middle ear ossicles with BC stimulation was found to be around 1.5 kHz (Stenfelt *et al.*, 2002). Onchi's explanations seem questionable, since neither of the peaks was found for occlusion of the ear canal in this study.

Huizing (1960) explained the occlusion effect by resonance properties of the enclosed air in the ear canal: a closed tube has resonance properties that differ from those of an open tube. Tonndorf (1966) presented another explanation. The mass-effect of the air column in the ear canal together with the compliance of air in the ear canal and TM form a high-pass filter effect for the ECSP. When the ear canal is occluded, the high-pass filter effect is eliminated, which results in an increase in low-frequency sound. Tonndorf's explanation is correct for the low frequencies where the mass and compliance of the air in the ear canal determine the acoustic properties, whereas Huizing's explanation is correct at higher frequencies where resonances and antiresonances determine the acoustic properties of the ear canal (above 2 kHz for the human ear canal).

D. The ECSP contribution to hearing by BC

With BC stimulation, the middle ear ossicles move relative to the surrounding bone due to two phenomena. One is the inertial effect (mass effect) of the ossicles themselves, and the other is the ECSP acting on the TM, which results in a motion of the ossicles. In an effort to determine their relative influence on hearing by BC, the relative umbo velocity was compared with the ECSP. Figure 9 shows the ECSP and relative umbo velocity level increase after occlusion of the

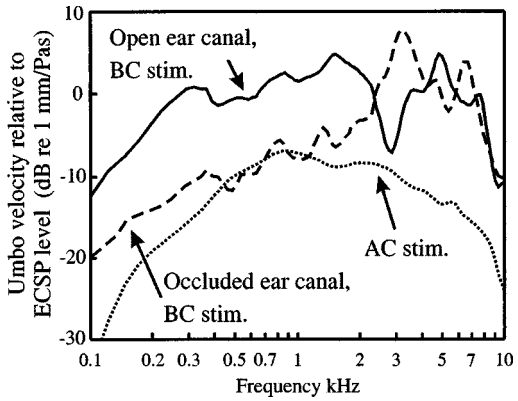


FIG. 10. Level of umbo velocity relative to ECSP. Solid line: open intact ear canal with BC stimulation (mean of nine ears). Dashed line: occluded intact ear canal with BC stimulation (mean of nine ears). Dotted line: AC stimulation in temporal bone specimens with an artificial ear canal (mean of nine temporal bone specimens). Frequency resolution is 50 points/decade.

intact ear canal. The ECSP increased by a maximum of about 18 dB for frequencies between 400 and 800 Hz. The rise of the relative umbo velocity at those frequencies is approximately 8 dB. If the ECSP is the only origin of the umbo motion, the increase of the relative umbo velocity should equal the increase of ECSP after occlusion of the ear canal. It should be noted here that an occlusion of the ear canal only minimally affects the inertial response of the ossicles (Stenfelt *et al.*, 2002). Hence, the contribution of the ECSP to the umbo velocity is 10 dB lower than that of the inertial effects for an intact open ear canal with BC stimulation. This concurs with the results of both Goldstein and Hayes (1971) and Huizing (1960), which show a greater ECSP increase for occlusion of the ear canal than the subjectively perceived occlusion effect: the difference was typically 5 to 15 dB for frequencies below 1 kHz.

Another way to investigate the ECSP influence on hearing by BC is to compare the input admittance (umbo velocity divided by ECSP) of AC and BC stimulation. Figure 10 shows the relative umbo velocity divided by the ECSP for three situations: (1) a normal intact open ear canal stimulated by BC, (2) an occluded intact ear canal also with BC stimulation, and (3) a temporal bone specimen with an artificial ear canal and stimulated by AC. The third is calculated from nine temporal bone specimens stimulated with a small AC receiver in the artificial ear canal. The ECSP is obtained at the same position for all three conditions (2 mm in front of the TM). For frequencies below 2 kHz, the relative umbo velocity divided by the ECSP is about 10 dB higher with an open ear canal than with an occluded one. For frequencies between 0.4 and 1.2 kHz, similar results are obtained for an occluded ear canal with BC stimulation as with AC stimulation.

This result indicates that, for an open ear canal stimulated by BC, the inertial effect of the ossicles is about 10 dB greater than the contribution from the ECSP. However, when the ear canal is occluded and stimulated by BC, for frequencies between 0.4 and 1.2 kHz, the relative velocity umbo is dominated by the ECSP. Consequently, the ECSP with BC stimulation can have a major influence on hearing by BC for frequencies between 0.4 and 1.2 kHz when the ear canal is occluded but not when it is open.

Huizing (1960) found that the ECSP with BC stimulation in an open ear canal was higher for frequencies below 500 Hz and lower for frequencies above 500 Hz than when an AC stimulus caused the same sensation. He believed that the sound pressure which arises in the external ear canal with BC stimulation should not be considered a stimulus that produces BC via the AC route. Moreover, in patients with otosclerosis of the stapes, the AC low-frequency hearing impairment can reach 60 dB, whereas BC thresholds are hardly affected (Ginsberg and White, 1994); this is yet another indication that the ECSP for an open ear canal is not a large contributor to the total BC sensation. In a study of cats, Brinkman *et al.* (1965) found that for both large and small perforations of the TM, there was only minor effect on the BC response, which further indicates that BC sound is not caused by the ECSP.

Khanna *et al.* (1976) placed the BC vibrator on the forehead and measured the ECSP with a probe tube microphone, placed inside the ear canal when the ear was occluded and just outside for an open ear canal. They reported the difference between the occluded and open ear canals to be over 20 dB for the entire frequency range, which is inconsistent with our findings, as well as most reported results, for occlusion of the ear canal (see Sec. IV C). They further reported that hearing by BC was dominated by the outer ear component for frequencies below 2 kHz, with the ear canal occluded, and below 800 Hz with an open ear canal. Figures 9 and 10 show that it is only when the ear canal is occluded, and for frequencies below 1.2 kHz, that the external ear component can dominate the BC response in humans. This difference in results could originate partly from the use of the mastoid as the stimulation position here, while the forehead is stimulated in the study by Khanna *et al.*; the middle ear inertial sensitivity for low frequencies can be 5 to 10 dB better when BC is stimulated at the mastoid instead of at the forehead (Studebaker, 1962; Dirks and Malmquist, 1969; Goodhill *et al.*, 1970; Stenfelt *et al.*, 2002). However, the skull moves in more or less all dimensions in space wherever a BC stimulation is applied to it; hence, it is unlikely that BC would be more sensitive to inertial effects of the middle ear for stimulation applied at the mastoid than for stimulation at the forehead (Stenfelt *et al.*, 2000).

E. Influence of the jaw on ECSP

The ear canal sound pressure did not show any significant change (in two ears) after sectioning of the lower jaw condyle: there was only a slight increase in ECSP of 4 dB around 3 kHz (Fig. 4). Franke *et al.* (1952) found that in an occluded ear the ECSP was 5 to 10 dB higher between 100 and 400 Hz when the mouth was open than for a closed mouth. This effect of the jaw position almost vanished when the ear canal was open. The position of the lower jaw affects the overall compliance of the cartilage part of the ear canal; it may be that the difference in the stiffness of the cartilage altered the ECSP when the lower jaw position was changed. However, Huizing (1960) reported that there was no systematic change of the ECSP with position of the lower jaw.

Huizing (1960) further reported about 10 dB higher ECSP after the jaw had been resected. Tonndorf (1966) found insignificant and nonsystematic changes of the ECSP at the high frequencies after removal of the jaw in cats. Howell and Williams (1989) showed that ECSP exists in the ear canal even after the jaw has been removed, while the results from the present study show no significant difference in ECSP with BC stimulation after resection of the jaw. Consequently, the relative motion between the lower jaw and the skull, which was suggested by Békésy (1932) to be a major contributor to the ECSP, has only a minor influence on the ECSP with BC stimulation, as predicted by Tonndorf (1966). Nevertheless, the relative motion between the lower jaw and the skull can influence BC hearing in another way. There is a ligament that connects the malleus with the temperomandibular joint (Pinto, 1962). This ligament can transmit vibrations from the jaw to the middle ear ossicles, which might contribute to BC hearing at low frequencies.

F. Clinical BC testing

It has been suggested that BC hearing threshold measurement should be conducted with the ear canal occluded (Onchi, 1954). One reason for this is the theory that the occlusion effect reduces the ambient masking. According to this theory, the true BC thresholds are better measured with the ears occluded; however, the theory is now known to be erroneous. Another reason to use occluded ear canal testing is that the masking procedures used in BC threshold measurement can give rise to occlusion effects, which suggests that it might be better to make all measurements with the ears occluded to reduce the uncertainties introduced by the masking procedure. The variability of the BC thresholds is similar for both open and occluded ear canals (Dirks and Swindeman, 1971). However, if the BC thresholds are measured with the ear canal occluded, the ECSP dominates the BC hearing at low frequencies. Hence, what one measure is an AC signal produced in the ear canal by a mechanical vibration; a middle ear lesion affects the AC and BC thresholds similarly. Consequently, if the measurement of BC thresholds is made with the ear canals occluded, it can be difficult to distinguish between the conductive and sensorineural losses from AC and BC threshold measurements. Also, since there is no standard for BC hearing thresholds with the ears occluded, the use of this method for BC testing is not advisable.

If, on the other hand, an insert phone is used to provide the masking stimulus, and the insert phone is positioned down to the bony part of the ear canal, the occlusion effect is minimized. As noted earlier, it is mainly vibrations in the cartilage part of the ear canal that cause the ECSP; when this source of the sound is removed, the ECSP caused by BC stimulation is lowered by some 15 dB. Consequently, the occlusion of the ear canal by an insert phone causes no BC sound increase that influences the hearing, provided the insert phone is tightly fitted and positioned in the bony part of the ear canal.

The occlusion of the ear canal may serve as a simple means of differentiating between conductive and sensorineural hearing losses when using the Bing test. In this test the stem of a tuning fork is applied to the mastoid, after which the patient's ear canal opening is gently occluded with the finger. If there is no conductive lesion of the ear, the tone (below 1 kHz) should become louder due to the occlusion effect. If no increase is detected, this should be taken as an indication of a conductive loss at that ear.

V. CONCLUSIONS

The ear canal sound pressure (ECSP) and malleus umbo velocity with bone conduction (BC) stimulation were measured in nine ears from five human cadaver heads. Furthermore, four temporal bone specimens were used to measure the sound radiation from the TM, and nine temporal bone specimens to measure the malleus umbo velocity with air conduction (AC) stimulation. With a transducer rigidly attached to the mastoid, the ECSP per unit applied force in an open ear canal is -35 dB *re* 1 Pa/N for low frequencies; above 500 Hz, the ECSP rises by 15 dB/oct until it reaches the quarter wavelength resonance at 2.7 kHz. With the ear canal occluded, the level of the ECSP is 15 to 20 dB higher at frequencies below 1 kHz. At higher frequencies, the difference between an open and an occluded ear canal is determined by the resonance properties of the ear canal.

Removing the lower jaw does not significantly influence the ECSP with BC stimulation. The major contribution to the ECSP is due to vibrations of the cartilage and soft tissues in the ear canal; when the cartilage and soft tissues are removed from the ear canal, the ECSP is reduced by 5–15 dB. With BC stimulation, the sound radiated from the TM is lower than the ECSP in a normal intact ear canal. The slightly greater ECSP found after removal of the TM is attributed to the influence on resonance properties of the effective length of the ear canal and possibly of sound radiation into the middle ear cavity.

Moreover, it was found that when the ear canal is occluded, the increase in relative umbo velocity with BC stimulation does not correspond to the increase in ECSP; the relative umbo velocity increase was 10 dB less than the ECSP level increase. Comparing the umbo velocity and ECSP by using both BC and AC stimulation shows that, for an intact open ear, the total BC stimulation of the inner ear is only minimally influenced by the ECSP. However, when the ear canal is occluded, the ECSP caused by BC stimulation can dominate hearing by BC for frequencies between 0.4 and 1.2 kHz.

ACKNOWLEDGMENTS

This work was supported in part by V.A. Merit Review Grant No. GDE0010ARG and the Swedish Institute.

Allen, G., and Fernandez, C. (**1960**). "The mechanism of bone conduction," Ann. Otol. Rhinol. Laryngol. **69**(1), 5–28.

Bárány, E. (**1938**). "A contribution to the physiology of bone conduction," Acta Oto-Laryngol., Suppl. **26**, 1–129.

Békésy, G. von (**1932**). "Zur theorie des hörens bei der schallaufnahme durch knochenleitung (Hearing theory of acoustic perception by bone conduction)," Ann. Phys. (Leipzig) **13**, 111–136.

Békésy, G. von (**1941**). ''Uber die schallausbreitung bei knochenleitung (About acoustic transmission by bone conduction),'' Z. Hals Nasen Ohrenheilk. **47**, 430–442.
Békésy, G. von (**1960**). *Experiments in Hearing*, edited by E. G. Wever (McGraw–Hill, New York), p. 745.
Brinkman, W., Marres, E., and Tolk, J. (**1965**). ''The mechanism of bone conduction,'' Acta Otolaryngol. **59**, 109–115.
Dirks, D., and Malmquist, C. (**1969**). ''Comparison of frontal and mastoid bone-conduction threshold in various conductive lesions,'' J. Speech Hear. Res. **12**, 725–746.
Dirks, D., and Swindeman, J. (**1971**). ''The variability of occluded and unoccluded bone-conduction thresholds,'' in *Hearing Measurement: A Book of Readings*, edited by I. Ventry, J. Chaiklin, and R. Dixon (Appleton-Century-Crofts, New York), pp. 158–169.
Elpern, B., and Naunton, R. (**1963**). ''The stability of the occlusion effect,'' Arch. Otolaryngol. **77**, 44–52.
Franke, E. (**1956**). ''Response of the human skull to mechanical vibrations,'' J. Acoust. Soc. Am. **28**(6), 1277–1284.
Franke, E., von Gierke, H., Grossman, F., and von Wittern, W. (**1952**). ''The jaw motions relative to the skull and their influence on hearing by bone conduction,'' J. Acoust. Soc. Am. **24**(2), 142–146.
Ginsberg, I., and White, T. (**1994**). ''Otologic disorders and examination'' in *Handbook of Clinical Audiology*, 4th ed., edited by J. Katz (Lippincott, Williams and Wilkins, Philadelphia), pp. 6–24.
Goldstein, D., and Hayes, C. (**1971**). ''The occlusion effect in bone-conduction hearing,'' in *Hearing Measurement: A Book of Readings*, edited by I. Ventry, J. Chaiklin, and R. Dixon (Appleton-Century-Crofts, New York), pp. 150–157.
Goodhill, V., Dirks, D., and Malmquist, C. (**1970**). ''Bone-conduction thresholds. Relationships of frontal and mastoid measurement in conductive hypacusis,'' Arch. Otolaryngol. **91**, 250–256.
Groen, J. (**1962**). ''The value of the Weber test,'' in *Otosclerosis*, edited by H. Schuknecht (Little, Brown and Company, Boston), pp. 165–174.
Håkansson, B., and Carlsson, P. (**1989**). ''Skull simulator for direct bone conduction hearing devices,'' Scand. Audiol. **18**, 91–98.
Håkansson, B., Carlsson, P., and Tjellström, A. (**1986**). ''The mechanical point impedance of the human head, with and without skin penetration,'' J. Acoust. Soc. Am. **80**(4), 1065–1075.
Håkansson, B., Brandt, A., Carlsson, P., and Tjellström, A. (**1994**). ''Resonance frequency of the human skull *in vivo*,'' J. Acoust. Soc. Am. **95**(3), 1474–1481.
Howell, P., and Williams, M. (**1989**). ''Jaw movement and bone-conduction in normal listeners and a unilateral hemi-mandibulectome,'' Scand. Audiol. **18**, 231–236.
Howell, P., Williams, M., and Dix, H. (**1988**). ''Assessment of sound in the ear canal caused by movement of the jaw relative to the skull,'' Scand. Audiol. **17**, 93–98.
Huizing, E. H. (**1960**). ''Bone conduction-The influence of the middle ear,'' Acta Oto-Laryngol., Suppl. **155**, 1–99.
Khanna, S. M., Tonndorf, J., and Queller, J. (**1976**). ''Mechanical parameters of hearing by bone conduction,'' J. Acoust. Soc. Am. **60**, 139–154.
Kirikae, I. (**1959**). ''An experimental study on the fundamental mechanism of bone conduction,'' Acta Oto-Laryngol., Suppl. **145**, 110.
Naunton, R. (**1963**). ''The measurement of hearing by bone conduction,'' in *Modern Developments in Audiology*, edited by J. Jerger (Academic, New York), pp. 1–29.
Onchi, Y. (**1954**). ''The blocked bone conduction test for differential diagnosis,'' Ann. Otol. Rhinol. Laryngol. **63**, 81–96.
Pinto, O. (**1962**). ''A new structure related to the temporomandibular joint and middle ear,'' J. Prosthet. Dent. **12**(**1**), 95–103.
Stenfelt, S., Håkansson, B., and Tjellström, A. (**2000**). ''Vibration characteristics of bone conducted sound *in vitro*,'' J. Acoust. Soc. Am. **107**, 422–431.
Stenfelt, S., Hato, N., and Goode, R. L. (**2002**). ''Factors contributing to bone conduction: The middle ear,'' J. Acoust. Soc. Am. **111**(2), 947–959.
Studebaker, G. (**1962**). ''Placement of vibrator in bone-conduction testing,'' J. Speech Hear. Res. **5**(4), 321–331.
Tjellström, A., Håkansson, B., and Granström, G. (**2001**). ''Bone-anchored hearing aids. Current status in adults and children,'' Otolaryngol. Clin. North Am. **34**, 337–363.
Tonndorf, J. (**1966**). ''Bone conduction. Studies in experimental animals,'' Acta Oto-Laryngol., Suppl. **213**, 1–132.
Tonndorf, J. (**1972**). ''Bonde conduction,'' in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), pp. 197–237.
Wever, G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton U.P., Princeton), p. 454.

Differential responses to acoustic damage and furosemide in auditory brainstem and otoacoustic emission measures

David M. Mills[a)]
V. M. Bloedel Hearing Research Center, Department of Otolaryngology, Head & Neck Surgery, University of Washington, Seattle, Washington 98195

(Received 17 August 2002; revised 4 November 2002; accepted 9 November 2002)

Characteristics of distortion product otoacoustic emissions (DPOAEs) and auditory brainstem responses (ABRs) were measured in Mongolian gerbil before and after the introduction of two different auditory dysfunctions: (1) acoustic damage with a high-intensity tone, or (2) furosemide intoxication. The goal was to find emission parameters and measures that best differentiated between the two dysfunctions, e.g., at a given ABR threshold elevation. Emission input–output or "growth" functions were used (frequencies f_1 and f_2, $f_2/f_1=1.21$) with equal levels, $L_1=L_2$, and unequal levels, with $L_1=L_2+20$ dB. The best parametric choice was found to be unequal stimulus levels, and the best measure was found to be the change in the emission threshold level, Δx. The emission threshold was defined as the stimulus level required to reach a criterion emission amplitude, in this case -10 dB SPL. (The next best measure was the change in emission amplitude at high stimulus levels, specifically that measured at $L_1\times L_2=90\times 70$ dB SPL.) For an ABR threshold shift of 20 dB or more, there was essentially no overlap in the emission threshold measures for the two conditions, sound damage or furosemide. The dividing line between the two distributions increased slowly with the change in ABR threshold, ΔABR, and was given by $\Delta x_t = 0.6\ \Delta\text{ABR}+8$ dB. For a given ΔABR, if the shift in emission threshold was more than the calculated dividing line value, Δx_t, the auditory dysfunction was due to acoustic damage, if less, it was due to furosemide. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1535942]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Wn [BLM]

I. INTRODUCTION

For clinical purposes, considerable research has been devoted to finding characteristics of emission measurements that provide the best prediction of auditory dysfunction as measured by pure-tone thresholds (e.g., Gorga *et al.*, 1997; Dorn *et al.*, 1999; Boege and Janssen, 2002). Even with the best choice of parameters, however, in the human population there remains a wide spread in emission responses at a given audiometric status. The overlapping distributions in emission responses as a function of hearing threshold make it difficult to adequately predict hearing function from emission measurements alone.

However, at least some of the differences observed between emission characteristics and hearing threshold measures could reflect differences in underlying pathologies. A simple argument suggests one reason that such differences should occur. Both animal experiments and theoretical considerations suggest that otoacoustic emissions depend primarily on outer hair cell (OHC) functionality (Neely and Stover, 1993; Trautwein *et al.*, 1996). By contrast, overall auditory function evidently depends not only on OHC integrity but on normal function of the inner hair cells (IHCs), eighth nerve, and auditory pathways in the brain. Because common dysfunctions could differentially affect different stages of this system, otoacoustic emissions should not necessarily be expected to correlate well with audiometric status in the general population. Rather than trying to find emission stimulus parameters that result in the best *agreement* with behavioral thresholds in the case of dysfunction, it may be preferable to find emission parameters and measures that best *distinguish* between different pathologies.

At present, it is not possible to establish such results through human studies alone. The normal human population has a mix of pathologies, and there is no easy way to distinguish between them. In fact, it is for this very reason that otoacoustic emissions could become of immense clinical diagnostic value. The results to be presented here suggest that emission responses combined with other appropriate tests might provide a noninvasive method of distinguishing between common dysfunctions in human hearing. Such differential diagnosis has the potential to result in better choices for treatment. For example, it may be possible to determine unequivocally whether a particular patient's symptoms were due to a lifetime accumulation of noise damage or were caused by progressive strial dysfunction (e.g., Gates *et al.*, 2002). Such possibilities are the motivation for the present line of inquiry.

As a first step, animal studies are required to create different pathologies of the auditory system and to investigate the best emission parameters to distinguish between them. This report describes the differences found in distortion product otoacoustic emission (DPOAE) characteristics and auditory brainstem responses (ABRs) with respect to two relatively simple induced dysfunctions. DPOAEs were chosen because these are strong and very consistent emissions in rodents, as well as being routinely measured in humans. The

[a)]Electronic mail: dmmills@u.washington.edu

two types of cochlear manipulation chosen were acoustic damage and furosemide injection.

The acoustic damage paradigm involved measurements before and 2 weeks after application of a high-intensity pure tone. The animal was allowed to recover for 2 weeks so that only permanent threshold shifts remained (Puel *et al.*, 1998). It is obviously hoped that such an acoustic damage paradigm will provide guidance to DPOAE and threshold measures that could be most useful for differential diagnosis involving noise damage in human subjects.

For comparison with the acoustic damage paradigm, systemic furosemide injection was employed. Such injections typically caused an immediate, profound loss of auditory function followed by a rapid but partial recovery, then a long plateau that was essentially flat, reflecting very slow recovery (e.g., Mills *et al.*, 1993). Auditory function was measured as late as possible during the plateau phase when the response was relatively stable, typically 3–8 h following injection.

From the point of view of simply providing a useful contrast, the mechanism of action of furosemide is not particularly relevant, as long as it affects cochlear mechanics differently than acoustic damage. However, note that by causing temporary dysfunction of the stria vascularis, furosemide appears to affect cochlear function almost entirely through the consequent decrease in the endocochlear potential (EP; Mills *et al.*, 1993, and references therein). Furosemide injection was therefore chosen for this study because of its potential as a model for differential diagnosis of strial dysfunction. However, because it was found that gerbil auditory function usually recovered completely from systemic furosemide injection in less than a day, it was only possible using this approach to create strial dysfunction persisting over short (8-h) time periods. Additional studies are required to determine if the results obtained here apply to chronic strial dysfunction in humans.

Finally, it is well established that ABR responses are a useful substitute for measurement of behavioral hearing thresholds in laboratory animals. The absolute ABR thresholds measured are usually somewhat higher than behavioral thresholds (Borg and Engstrom, 1983). However, as a response to an experimental manipulation, *changes* in ABR tone-pip thresholds have been shown to be a useful frequency-specific measure of the changes in behavioral thresholds (Davis and Ferraro, 1984). It is this latter aspect that is most relevant to the current experiments.

The purpose of the present study was not to provide specific guidelines for DPOAE measurements in human subjects. Rather, this first study was undertaken primarily as a proof of concept, as a demonstration. The results obtained demonstrate that the two different mechanisms chosen do produce two distinct cochlear dysfunctions, and that these cause differential effects on measurements of DPOAEs and auditory threshold. The study further demonstrates that it is possible to choose DPOAE parameters and to analyze the responses so as to *maximize* the differences between the DPOAE responses to the two conditions. In fact, parameters and specific measures were found which resulted in essentially nonoverlapping distributions in emission responses at a given ABR threshold shift. The goal to demonstrate the possibility of noninvasive differential diagnosis was therefore reached: it was shown to be possible to distinguish between the two underlying pathologies by comparing specific DPOAE responses at a given audiometric status.

II. METHODS

Subjects for the study were young adult Mongolian gerbils (*Meriones unguiculates*) obtained either from Charles River Laboratories (Wilmington, MA) or from breeding pairs obtained from the same source. All procedures were approved by the Animal Care Committee at the University of Washington.

For measurements of evoked responses, animals were initially anesthetized with a subcutaneous injection of a mixture of ketamine (Ketaject, Phoenix Pharm., 80 mg/kg) and xylazine (Xyla-ject, Phoenix Pharm., 5 mg/kg). Supplemental anesthesia was administered as needed. Typically, the same ketamine–xylazine mixture was given at about 1-h intervals, plus ketamine alone at the intervening half-hours, at dosages 1/4 to 1/2 the first injection. Animals were also administered a supportive dose of glucose solution (usually 1 cc) at 1-h intervals. In addition, a sterile ophthalamic ointment was placed in each eye at the start of the procedure.

The animal was attached to an adjustable surgical head holder (Kopf) employing a bite bar, and positioned on a heating pad. Animal temperature was maintained at 37 °C by an automatic system monitoring anal temperature. The pinna on the left side was removed, and a small amount of tissue removed over the bulla immediately dorsal-posterior to the ear canal. A small hole was drilled into the bulla and a tube (0.86 mm i.d. by 6.5 cm long) force-fit into the hole. The purpose of the small-diameter tube was to provide static pressure relief for the middle ear while maintaining normal middle-ear function.

The equipment used for emission and ABR measurements was the same as described previously (Mills and Shepherd, 2001; Mills, 2002). Briefly, a custom coupler was advanced under micromanipulator control and sealed to the ear canal. This coupler included a port for a low-noise microphone (Etymotics ER10B), a port for a probe tube reference microphone, and two ports for sound delivered through tubing from custom tweeters. Before beginning measurements, a wideband noise signal was generated by one tweeter, and the output of the low-noise microphone was calibrated *in situ* by comparison to the output of the probe tube microphone. The same coupler and calibration were used for both DPOAE and ABR measurements.

For all animals, emission measurements were made first. Stimuli were two tones at frequencies f_1 and f_2, with $f_2/f_1=1.21$. Baseline measurements included input–output, or ''growth'' functions for both the equal-level stimulus case, $L_1=L_2$, and the unequal case with $L_1=L_2+20$ dB. Growth functions were measured at octave intervals with f_2 frequencies from 1 to 32 kHz. Half-octave intervals were added between 4 and 32 kHz for the acoustic damage experiments. All growth functions were measured with 5-dB steps in stimulus levels, for stimulus levels chosen for emissions below detectability, up to levels of $L_1 \times L_2$ equal to either 80

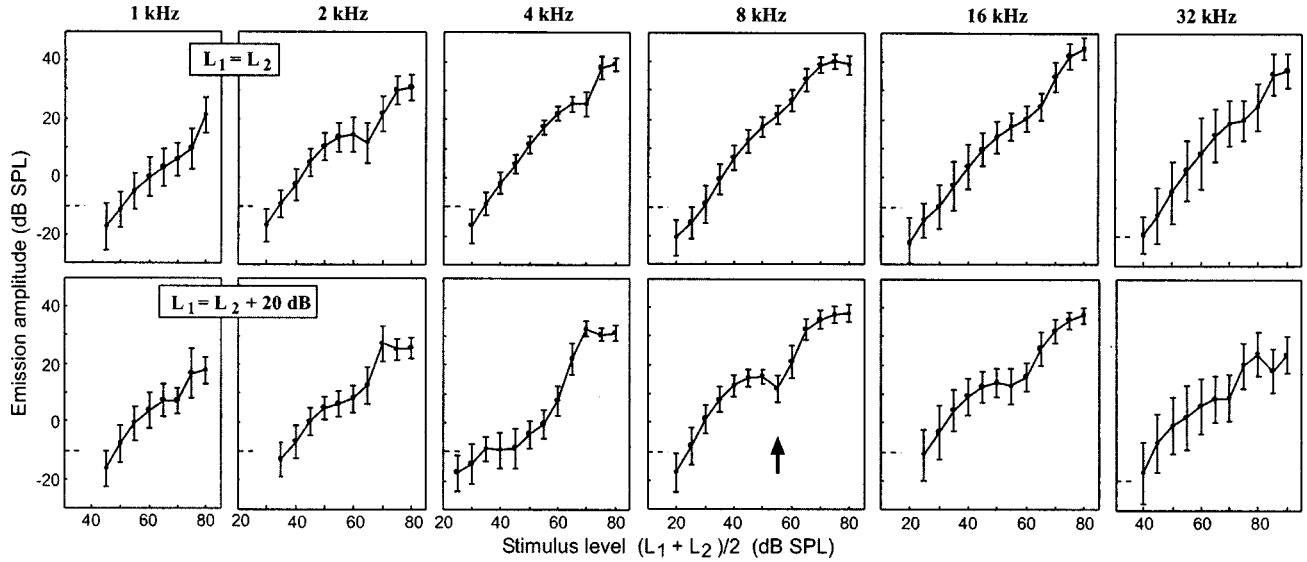


FIG. 1. Emission input–output, or "growth" function responses for normal young adult gerbils. The mean plus and minus one standard deviation is shown at each stimulus level. Above each column of panels is the f_2 frequency in kHz. The top row of panels give the responses for equilevel stimuli, $L_1=L_2$, and the bottom row of panels for unequal levels, with $L_1=L_2+20$ dB. The vertical axis in each panel is the distortion product emission at $2f_1-f_2$. The horizontal axis is the stimulus level, defined to be the geometric mean equal to $(L_1+L_2)/2$ when measured in dB SPL. The horizontal dashed line in each panel indicates the emission amplitude equal to -10 dB SPL, the level chosen as the criterion threshold emission level in subsequent analysis. Note that the 32-kHz panels are shifted 10 dB in both axes relative to all other panels. For clarity, mean values measured at stimulus levels where there was significant contamination by noise were omitted.

$\times 80$ or 90×70 dB SPL. At 32 kHz, the upper limits were always 10 dB higher. Higher limits were also employed at all frequencies for animals with emissions significantly weakened by experimental manipulation.

In all animals, ABR measurements were made using the same stimulus frequencies as the f_2 frequencies chosen for the emission measurements. Tone pips were generated with a 5-ms total duration, with 1-ms $\cos^2$ rise/fall times, and presented with alternating polarity. Typically, each measurement consisted of the average of the response to 1000 pips presented at a rate of 33 Hz. Stimulus levels were varied at 10-dB intervals to find the approximate threshold, then threshold was bracketed by repeating measurements at least twice at each stimulus level at 5-dB intervals. All traces were recorded and thresholds determined offline by visual inspection.

In most animals, following baseline measurements, cochlear dysfunction was initiated either with acoustic damage or furosemide injection. For acoustic damage measurements, a pure tone at 11.3 kHz was applied at 130-, 135-, or 140-dB SPL for either 15 or 30 min. For control animals, the tone was at 80 dB SPL, but all other procedures were the same. Following application of the tone, the bulla tube was removed, the hole sealed with bone wax, and the skin sutured over the bulla. The animals were allowed to recover for 2 weeks, and then remeasured with the same procedures. During the 2-week recovery, they were housed individually but allowed normal access to food and water.

For the furosemide-treated animals, the application of furosemide (American Regent Labs, Shirley, NY) was usually made following baseline measurements while the animal was still anesthetized, and was applied by either subcutaneous or intraperitoneal injection. For this group, typical furosemide applications ranged from 1–3 injections of 100 mg/kg at 40-min intervals. The animal was kept anesthetized and the emission and ABR responses were measured for 3–6 h following the application. Measurements were always continued until measured emission responses were stable within a few dB over an hour's time.

In eight additional animals, furosemide was instead given 3–7 h prior to anesthesia and no baseline measures were made. Typical applications were 3–4 injections of 100–200 mg/kg at 40-min intervals. These animals were subsequently anesthetized and evoked responses measured, at a time typically 4–8 h after the last injection. For these animals, baseline responses were taken to be the mean of the normal responses (as summarized in Figs. 1 and 2).

Instrumental noise and distortion levels were estimated as previously reported (Mills and Rubel, 1996; Mills, 2002), and all emission amplitudes shown in this report and used for analysis were 10 dB or more above the measured noise and instrumental distortion levels.

III. RESULTS

A. Normal pretreatment responses

Measured initial emission growth functions are summarized in Fig. 1. These represent the mean and standard deviation of the emission responses measured in 28 normal young adult gerbils. Figure 2(A) presents the mean ABR frequency-threshold curve (FTC) for the same set of animals, and Fig. 2(B) presents the calculated emission threshold responses for a criterion emission level of -10 dB SPL. Note that for these measurements, the gerbil bulla was not open, but was vented with a small-diameter tube to allow static pressure relief. The emission and ABR responses in Figs. 1 and 2 therefore can be compared directly to measures obtained under natural conditions. For such comparison, Fig. 2(A) also summarizes

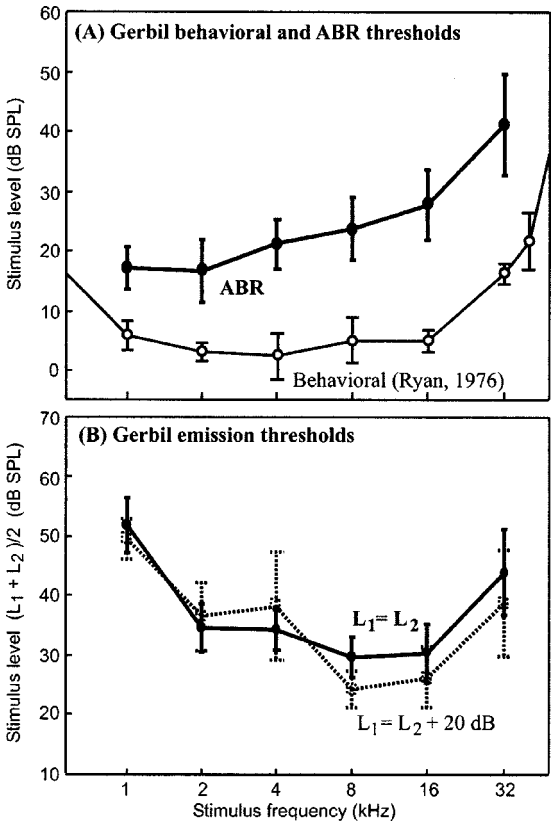


FIG. 2. Normal adult gerbils: Frequency-threshold curves (FTCs) from behavioral, auditory brainstem response (ABR), and emission measurements. Same subjects as Fig. 1 except that behavioral thresholds are from Ryan (1976). Means and standard deviations are shown. The emission threshold was defined as the stimulus level required to reach a criterion emission amplitude, here taken to be −10 dB SPL as noted in Figs. 1 and 3. In (B), the solid lines and filled circles indicate the equilevel case, $L_1 = L_2$, and the dashed lines the unequal case, with $L_1 = L_2 + 20$ dB.

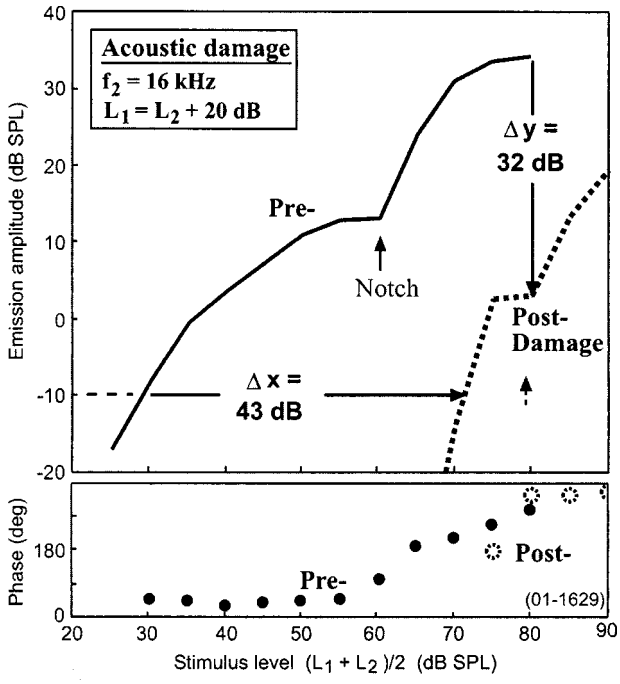


FIG. 3. Typical example of change in a growth function due to acoustic damage. The upper panel shows the emission amplitude as in Fig. 1. The lower panel displays the phase angle of the emission. Note that the absolute phase angle is not relevant, but the changes shown as a function of stimulus level are. For clarity, only phase responses clearly unaffected by noise were included. Exposure for this animal was an 11.3-kHz pure tone at 135 dB SPL for 30 min, and postdamage measurements were made 13 days following exposure. The change in growth function is shown for the frequency $f_2 = 16$ kHz and the unequal stimuli case, as noted.

the behavioral auditory threshold determinations by Ryan (1976). Note also that the emission stimulus level used for the axes in this report is the geometric mean of the two stimulus levels, i.e., the scale is equal to $(L_1 + L_2)/2$ when the levels are expressed in dB. This is done to provide a useful comparison of emission responses when comparing stimuli with different level ratios in a way that avoids choosing either of the individual levels, L_1 or L_2, alone. Using this representation, for example, the equal level stimulus, $L_1 \times L_2 = 80 \times 80$ dB SPL, is plotted at the same level, i.e., at "80," as that with the unequal levels, $L_1 \times L_2 = 90 \times 70$ dB SPL.

It can be seen from Fig. 1 that, for the adult gerbil, the variances were remarkably small at many frequencies and stimulus levels. That is, the emission responses were quite consistent from animal to animal, even to the location of the relative minima, or "notches," seen in many of the growth functions (e.g., Mills, 1997, 2002). For example, note the notch at 8 kHz consistently found at $L_1 \times L_2 = 65 \times 45$ dB SPL, noted by the short vertical arrow in Fig. 1.

It is worthwhile to consider the few situations seen in Figs. 1 and 2 where the variances were relatively large. The reason is that it may be advisable to avoid relying on measurements at such parameter choices, especially in situations where one has to compare individuals to the norms. At 4 kHz, for example, for the unequal level case there was a relatively large variance for low stimulus levels, between 25 and 45 dB SPL (Fig. 1, lower row). This was associated with a particularly wide variance in the emission threshold at that frequency [dashed line, Fig. 2(B)]. These larger variances were not due to contamination by noise, but to the fact that at this frequency there was typically found a weak, and quite variable, relative peak in the emission amplitude. This peak was similar to, but much weaker than, the relative peak seen at other frequencies, e.g., at 8 kHz below the notch level. The weakness and variability of the response at low stimulus levels for the unequal stimuli at 4 kHz is known, in fact, to be due to the intrusion of a phase cancellation "notch" into this region (Mills and Rubel, 1994; Mills, 2002).

It can be seen from Fig. 2(A) that mean ABR thresholds were slightly higher than behavioral thresholds, and that the difference increased slowly with frequency. In contrast, emission thresholds [Fig. 2(B)] were considerably higher than behavioral at the lowest frequencies tested, but their correspondence improved as the frequency increased. Such differences in absolute threshold, however, are relatively unimportant in the present study design, in which the focus is on the changes in threshold caused by experimental manipulations.

B. Differential changes in threshold measures

Figures 3–5 summarize an example of changes typically observed following acoustic damage. Responses were mea-

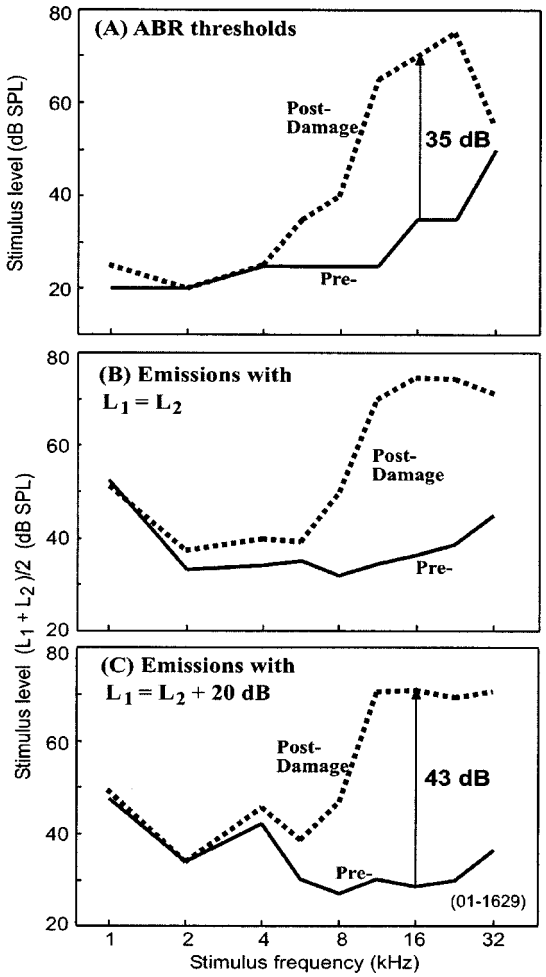


FIG. 4. Typical change in ABR and emission FTCs with acoustic damage. Same animal as in Fig. 3. The top panel shows the observed change in ABR response. The emission threshold responses are shown in the middle panel for the equilevel stimuli case, and in the bottom panel for the unequal case. The shift of 43 dB noted in the bottom panel at 16 kHz was that illustrated in Fig. 3. The corresponding shift in the ABR was 35 dB as noted in the top panel (A).

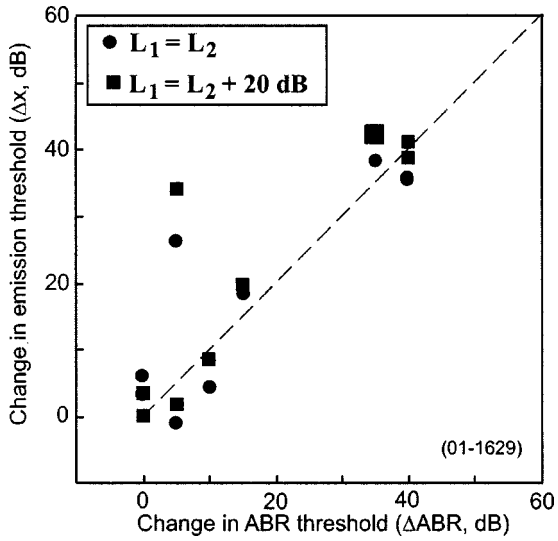


FIG. 5. Comparison of typical changes in ABR with the changes in emission threshold following acoustic damage, for the same animal as in Figs. 3 and 4. The horizontal axis is the change in ABR threshold in dB obtained from the top panel in Fig. 4. The vertical axis is the change in emission threshold at the same frequency, taken from the lower two panels in Fig. 4. Results for all frequencies measured from 1 to 32 kHz are included. The large symbol indicates the same point emphasized in Fig. 4, for which the change in ABR was 35 dB and the change in emission threshold was 43 dB. For reference, the dashed line indicates the line for equality between changes in ABR and emission thresholds.

sured immediately before and two weeks after exposure to an intense tone at 11.3 kHz. Figure 3 shows growth functions for $f_2 = 16$ kHz, approximately the frequency of maximum effect. Note that the postdamage emission amplitude was considerably reduced at all stimulus levels. For a criterion threshold equal to -10 dB SPL, there was an increase, denoted Δx, of 43 dB in the emission threshold stimulus level as noted. As indicated, there was also a decrease, denoted Δy, of 32 dB in the emission amplitude as measured at the level identified as 80 dB SPL, i.e., at $L_1 \times L_2 = 90 \times 70$ dB SPL. Note that, for convenience, the quantity Δy is defined positive in the normal case of an amplitude decrease.

As has been noted elsewhere (e.g., Mills, 2002), when a notch is seen in the normal emission growth function (identified by the solid vertical arrow in Fig. 3), it is usually associated with a rapid change in the emission phase angle (lower panel). In this animal, the postdamage growth function showed a similar notch, with similar characteristics, but located at a stimulus level almost 20 dB higher (dashed vertical arrow in upper panel of Fig. 3). Note that the shift in the notch location was less than the shift in threshold, and that there was an increase in the steepness of the growth function following damage. The overall increase in slope is illustrated by the fact that $\Delta x > \Delta y$. Also note that notches were not always observable in postdamage growth functions, particularly in cases where extensive damage had occurred.

Figures 4(B) and (C) summarize emission thresholds determined for all the measured f_2 frequencies, for the same animal as in Fig. 3. For comparison, Fig. 4(A) presents the ABR thresholds measured pre- and postdamage. Clearly, the ABR and both emission measures all tracked the changes very similarly, and on a frequency-specific basis. The 43-dB shift noted for the emission measure in Fig. 3 corresponds to a 35-dB change in ABR threshold as noted in Fig. 4(A).

A direct comparison of the shifts in threshold caused by this experimental manipulation is given in Fig. 5. On the vertical axis is the change in emission threshold, Δx, and on the horizontal axis the change in ABR threshold, ΔABR, measured at the same frequency. For this animal, all frequencies that were measured from 1 to 32 kHz are included. In this, and similar plots, emission measurements using equal-level stimuli are consistently represented by circles, and unequal stimuli by squares. In this type of figure, filled symbols are used to indicate changes due to acoustic damage (e.g., Fig. 5) to contrast with unfilled symbols used to indicate changes due to furosemide (e.g., Fig. 8).

For the acoustic damage example in Fig. 5, the shift in the emission threshold was seen to be about the same for the equilevel and unequal case, which in turn was about equal to or greater than the corresponding shift in the ABR threshold.

Following in the same pattern as Figs. 3–5, Figs. 6–8 summarize the changes seen in one representative animal following application of furosemide. In this case, following

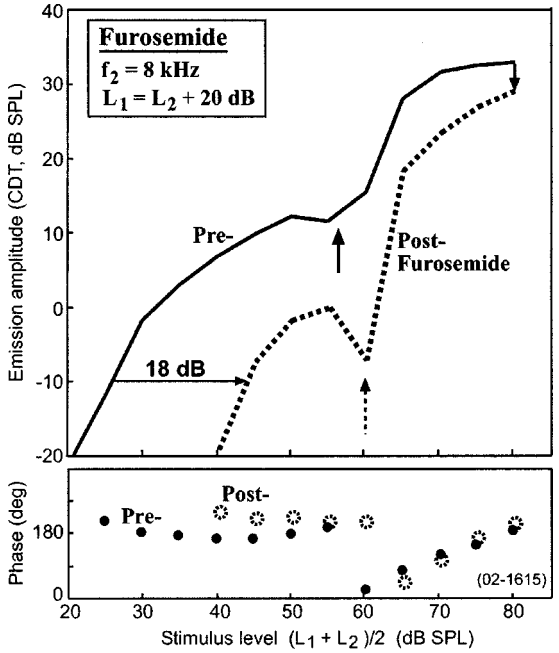


FIG. 6. Typical example of change in a growth function following application of furosemide. Same layout as Fig. 3. Following initial measurements, there were three subcutaneous injections of furosemide of 100 mg/kg at 40-min intervals. Postinjection measurements were made after 4 h of recovery following the last injection. In this case, the shift in emission threshold was only 18 dB at the emission criterion level of −10 dB SPL, as indicated.

baseline measures, three injections of 100 mg/kg furosemide were made at 40-min intervals. Postfurosemide measures were taken 4 h following the last injection. In contrast to the acoustic damage case, Fig. 6 shows that the changes in emission growth functions were relatively modest. The notch location moved only a few dB (vertical arrows), and the threshold shift, Δx, was only 18 dB. The emission amplitude shift, Δy, was even smaller, less than 5 dB. It should not be assumed that the auditory function was similarly only mildly impaired, however. The ABR threshold at the same frequency was elevated 55 dB [Fig. 7(A)].

For the furosemide intoxication case, thresholds for both emission and ABR measures were elevated across frequency, and both were relatively more elevated at higher frequencies. This result contrasts with the much more frequency-specific elevations seen with a pure-tone acoustic damage (Fig. 4). As in Fig. 5, Fig. 8 compares directly the changes in the emission and ABR thresholds. In contrast to the acoustic damage case, the changes in emission threshold were typically smaller than the ABR shifts.

Comparison of the two examples in Figs. 3–8 suggests two potentially useful differences between acoustic damage and furosemide intoxication in their effects on emission measures. (1) With furosemide, the shift in emission threshold appears to be comparatively smaller than the shift in ABR threshold, but with acoustic damage, the emission threshold shift is equal to or larger than the ABR threshold shift. (2) There seems to be a difference in the relative change in slope of the emission growth function, in that the emission amplitude at high stimulus levels, Δy, appears to decrease less with furosemide than it does with acoustic damage, for similar changes in emission threshold, Δx.

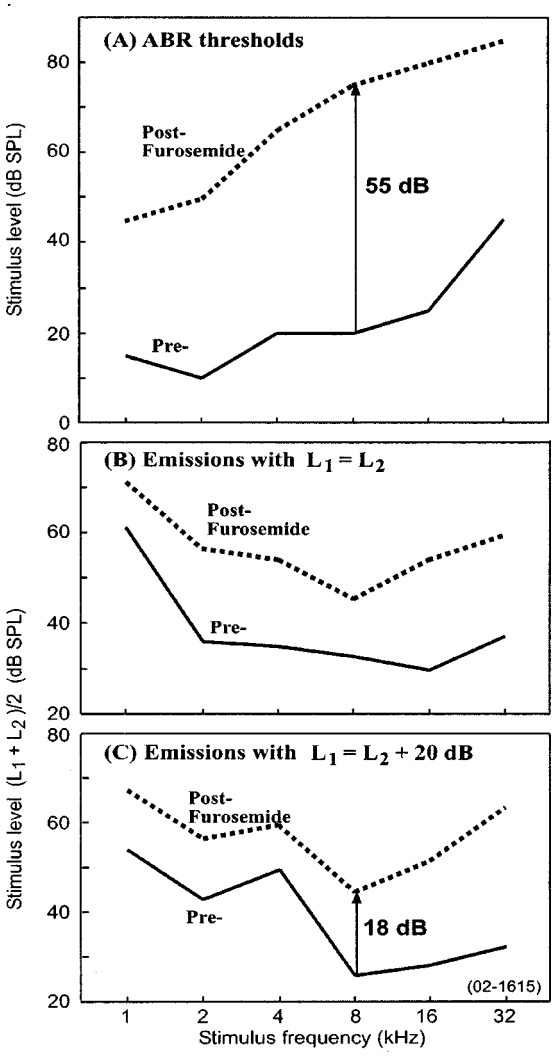


FIG. 7. Typical change in ABR and emission FTCs with furosemide. Same animal as in Fig. 6, and same layout as Fig. 4. The 18-dB shift in emissions seen in Fig. 6 is noted in the lower panel of Fig. 7. Note in the upper panel that the corresponding shift in ABR was 55 dB.

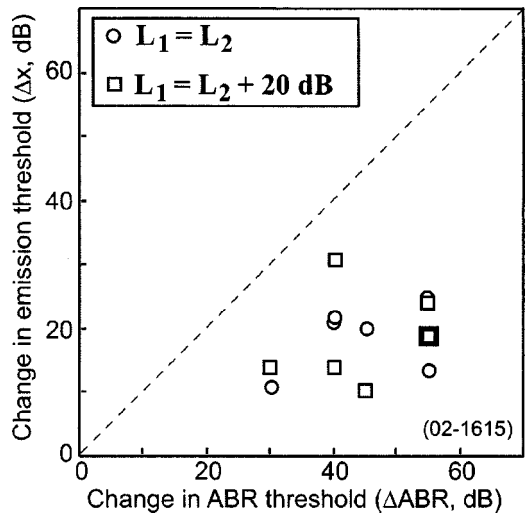


FIG. 8. Comparison of typical changes in ABR with the changes in emission threshold following furosemide. Same animal as Figs. 6 and 7, same layout as Fig. 5. The bold point is that emphasized in Fig. 7, with the ABR change of 55 dB and emission threshold change of 18 dB. All frequencies from 1 to 32 kHz were included.

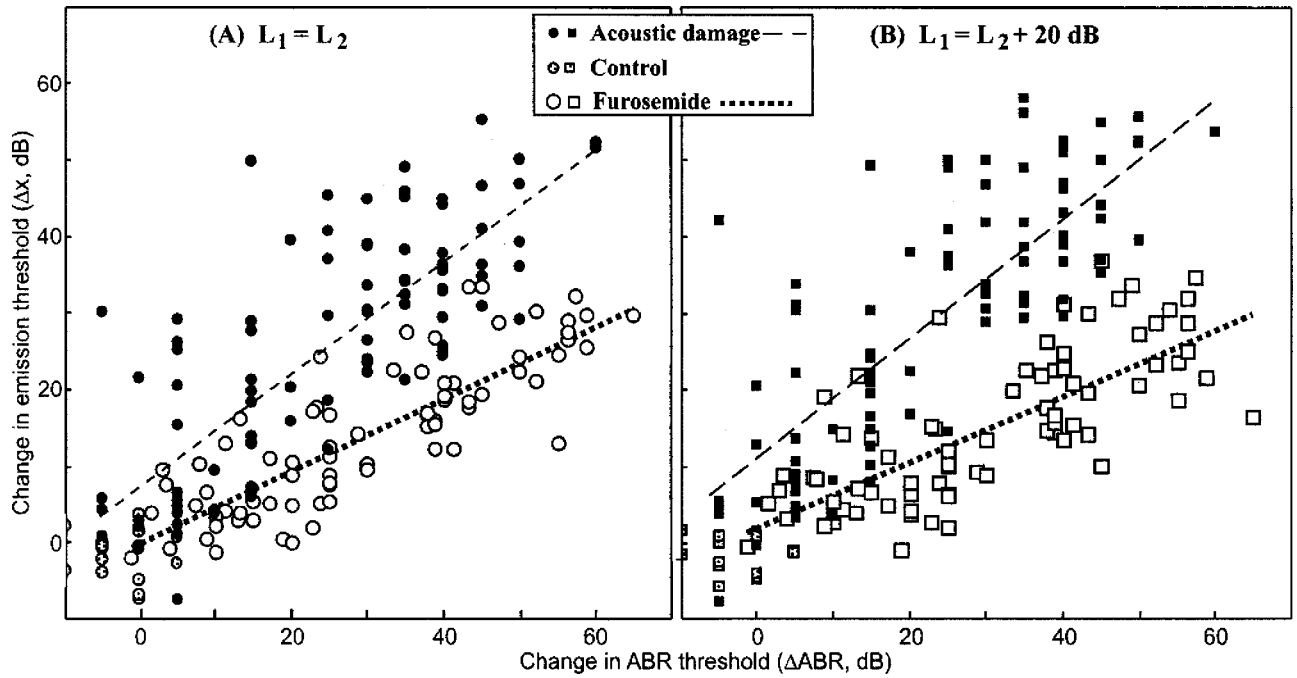


FIG. 9. Comparison of change in ABR and emission thresholds for all animals. In each panel, the horizontal axis is the change in ABR threshold, and the vertical axis the change in emission threshold (Δx). As in previous figures, the solid filled symbols are for acoustic damage with an 11.3-kHz tone. For clarity, these responses are included for f_2 frequencies above 4 kHz only in this figure. Since low frequencies responses were essentially unaffected by the 11.3-kHz tone, including these responses would only add a dense cluster of points near the origin (as illustrated in Figs. 4 and 5). The unfilled symbols are for furosemide effects, with all f_2 frequencies included from 1 to 32 kHz. The shaded symbols represent changes in controls remeasured two weeks later, with all frequencies included. Dashed lines indicate the least-square linear fit to the experimental data for each of the two experimental conditions.

To test the first possibility, Fig. 9 summarizes the comparison of threshold response changes for all the animals in the study. Symbols are the same as in previous graphs (Figs. 5 and 8). In addition, shaded symbols represent control measurements using the same acoustic damage protocol, but with only 80 dB SPL applied (see Sec. II). Dashed lines in Fig. 9 represent the least-squares linear fits to the responses of each experimental group. Clearly, the first possibility listed above is adequately verified. For the same change in ABR threshold, acoustic damage generally causes a significantly larger increase in the emission threshold measure than does furosemide intoxication. In fact, the two data sets overlap very little for unequal level emission stimuli (right panel). Discussion of Fig. 9 is resumed in the summary of the results below.

C. Differential changes in other measures

Figure 10 represents a test of the possibility of a differential change in the overall slope of the emission growth function. Conventions are the same as in Fig. 9, except that here both axes represent emission measures. The horizontal axis in Fig. 10 represents the change in the emission threshold, Δx, and the vertical axis the shift in the emission amplitude, Δy, as defined in Fig. 3. Figure 10 therefore summarizes changes in the emission growth function responses themselves, with no ABR information utilized. Only the unequal-level case is shown. In the same way as in Fig. 9, there was more overlap of the two distributions for the equal-level case, but the trends were the same.

The dashed lines represent the least-square linear fits to the responses in the two exposure conditions. The trend is clearly in the direction suggested by Figs. 3 and 6. For a given emission threshold shift, Δx, there was a slightly smaller mean decrease in the high-level emission amplitude, Δy, in the furosemide case compared to acoustic damage. However, there was also a very large overlap of the distributions. Figure 10 makes it clear that, at least for the parameter choices considered in this study, one *cannot* use the observed change in emission growth functions alone to distinguish be-

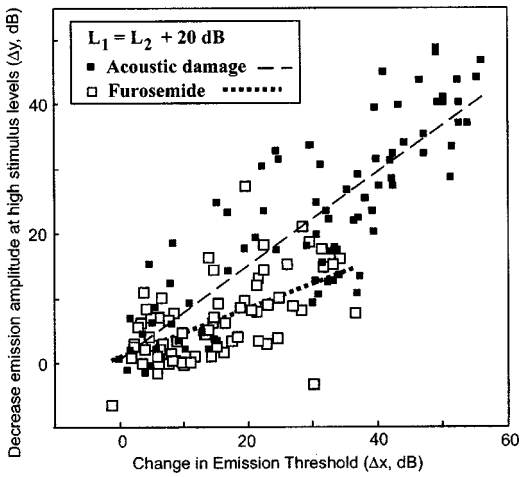


FIG. 10. Comparison of overall changes in emission growth functions for all animals. The horizontal axis is the change in emission threshold, Δx, and the vertical axis the decrease in emission amplitude at high stimulus levels, defined as Δy in Fig. 3. The filled symbols indicate acoustic damage, with data included only for f_2 frequencies above 4 kHz for clarity. The unfilled symbols indicate changes due to furosemide injection, with data at all f_2 frequencies from 1 to 32 kHz. The dashed lines indicate least-squares linear fits within each of the two experimental conditions.

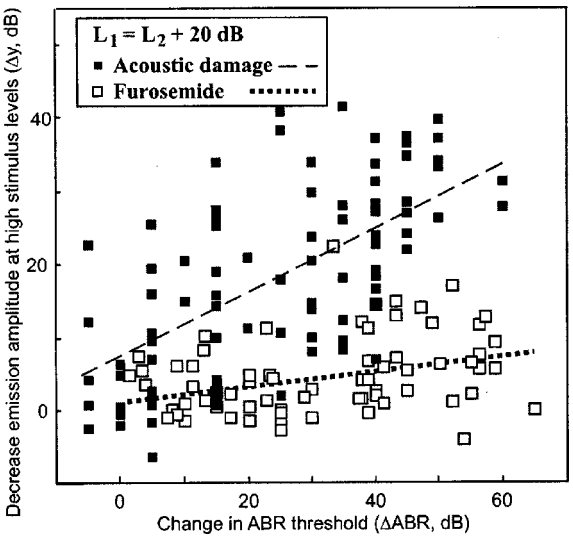


FIG. 11. Changes in ABR versus decreases in emission amplitude measured at high stimulus levels. The horizontal axis is the ABR threshold shift, ΔABR. The vertical axis is the shift in emission amplitude, Δy, as defined in Fig. 3, for the stimulus level $L_1 \times L_2 = 90 \times 70$ dB SPL. Dashed lines are least-squares linear fits for each condition. Note that the horizontal axes in all panels in Figs. 9 and 11 are the same, equal to the shift in ABR threshold, ΔABR. The vertical axis in Fig. 9 is the increase in emission threshold, Δx, while the vertical axis in Fig. 11 is the decrease in emission amplitude, Δy, measured at stimulus levels $L_1 \times L_2 = 90 \times 70$ dB SPL (the highest level consistently measured). These quantities are defined in Fig. 3.

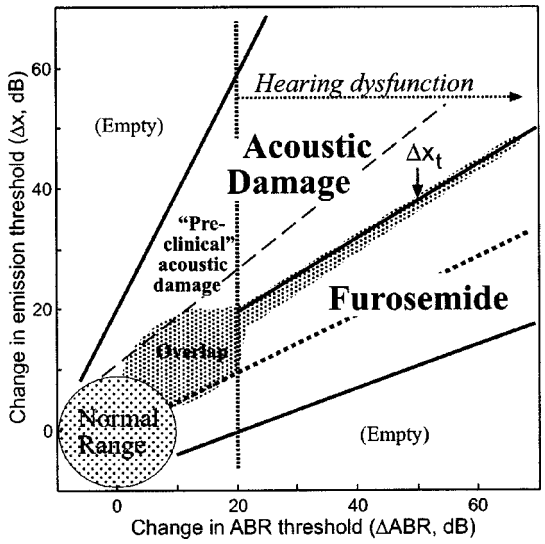


FIG. 12. Schematic of distribution of shifts in emission threshold responses (vertical axis) plotted against elevations of ABR threshold (horizontal axis). This diagram is for emission parameters, $L_1 = L_2 + 20$ dB, and is based on the raw data shown in Fig. 9(B). The heavy line shown, which approximately separates the two distributions and is labeled Δx_t, is defined by Eq. (1).

tween the acoustic damage and furosemide cases.

However, these two experimental conditions separate considerably more when the ABR threshold shift is substituted for the emission threshold shift at the same frequency. For the display in Fig. 11, the shift in high-level emission amplitude, Δy, on the vertical axis is compared to the ABR threshold shift (horizontal axis) measured at the same frequency.

There is still more overlap between the two distributions for the emission amplitude measure, Δy, in Fig. 11 compared to the emission threshold measure, Δx, in Fig. 9(B). Of course, it is possible that the choice of stimulus parameters for either or both measures could be further improved to provide better separation between the two experimental conditions.

D. Summary of results

With acoustic damage, emission growth functions tend to decrease at all stimulus levels, while with furosemide the emission amplitudes at high stimulus levels tend to decrease relatively less (Figs. 3 and 6). This means that, at the same shift in emission threshold, Δx, the emission growth function is typically slightly steeper for the furosemide case than for acoustic damage (Fig. 10). However, the difference is not large or consistent enough to allow this characteristic to be used for differential diagnosis between these two dysfunctions.

For a given change in ABR threshold, the emission growth function amplitude typically decreases much more in the case of acoustic damage than for furosemide. This is true whether one measures the change in growth function by the increase in emission threshold (Δx; Figs. 5, 8, and 9) or by the decrease in emission amplitude at a high stimulus level (Δy; Figs. 10 and 11). Note that the sensitivity of the emission *at high stimulus levels* to acoustic damage is in direct contrast to its relative invulnerability at such levels when using furosemide (Mills *et al.*, 1993). This apparent invulnerability led to the earlier impression of the presence of a ''passive'' emission component at high stimulus levels, no longer considered a useful distinction (e.g., Mills, 2002).

For purposes of differential diagnosis, of the possibilities considered in this study the best separation between the two experimental conditions is given using the emission threshold shift, Δx, with unequal stimulus levels ($L_1 = L_2 + 20$ dB). Note that, in Fig. 9, there is essentially no overlap between the two cases once the change in ABR threshold is 20 dB or more. The results from Fig. 9(B) are summarized schematically in Fig. 12. As suggested by this summary, very few (<1%) of the responses fall within the empty regions in this diagram, considering damage from either sound or furosemide injection. If there is acoustic damage, the points fall within the region noted, where the increase in emission threshold exceeds or is approximately equal to the change in ABR threshold. Conversely, if the dysfunction is caused by furosemide intoxication, there is typically much greater rise in ABR threshold compared to emission threshold change.

The emission responses to the two experimental conditions overlap only in the small region marked in Fig. 12. As shown, there is nearly complete separation between the two distributions for ABR threshold shifts of 20 dB or more. The border separating the two regions starts approximately at the location where the value of emission threshold elevation, Δx, is equal to 20 dB and where the ABR shift is also 20 dB. The border then increases linearly until it has the value of Δx equal to 50 dB for a ΔABR value equal to 70 dB. That is, the two distributions are separated by the heavy line drawn be-

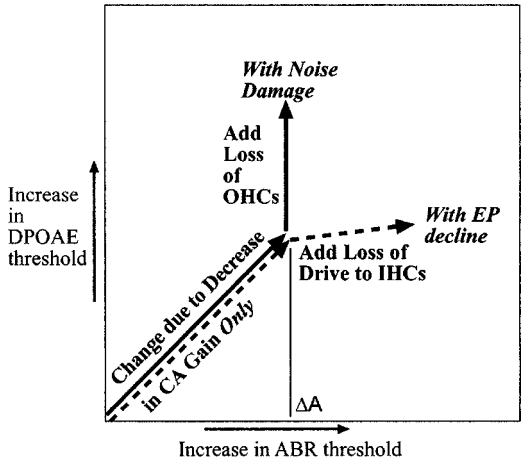


FIG. 13. Schematic illustrating differential effects in two cases, one of acute EP decline (dashed lines), the other of noise damage (solid lines), where both have experienced an identical decrease in cochlear amplifier (CA) gain.

tween the two regions in Fig. 12, a line that obeys the relationship

$$\Delta x_t = 0.6\Delta\mathrm{ABR} + 8\ \mathrm{dB}. \quad (1)$$

For example, for an ABR threshold increase of 50 dB, the corresponding separation between the two regions occurs at an emission threshold shift of 38 dB, as noted by the vertical arrow in Fig. 12.

It was also discovered that the emission amplitude shift at the highest level routinely measured in this study, 90 $\times$70 dB SPL, also distinguished the two cases moderately well (Fig. 11). That is, there was only slightly more overlap in Fig. 11 than in Fig. 9(B). Because emission amplitudes at a given stimulus level are easier to determine than emission thresholds, this kind of measure could legitimately be considered in situations where time for measurement is severely restricted. However, full emission growth functions are still to be preferred whenever possible, both because they offer improved separation between the two dysfunctions considered here (Fig. 9) and because the additional information that growth functions provide may turn out to be even more important when other dysfunctions are included.

IV. DISCUSSION

A useful way to conceptualize the general results to this point is illustrated in Fig. 13. First, consider the effect of the two different dysfunctions on the cochlear amplifier (CA) gain. The CA gain is defined as the increase of peak basilar-membrane response compared to the passive response, at low stimulus levels (e.g., Mills, 1997). Unfortunately, the direct measurement of CA gain requires invasive measurements (e.g., Ruggero and Rich, 1991). However, both dysfunctions introduced here appear to decrease the CA gain, among other effects. As a thought experiment, one could imagine selecting cases for comparison from each type of dysfunction that have identical decreases in the CA gain, at a given frequency. An effective decrease in CA gain of ΔA *by itself* will cause an increase in the ABR threshold of ΔA, as indicated in Fig. 13. If threshold stimulus levels in the two different measurements (emissions and ABR) are approximately equivalent, the decrease of CA gain by itself will also cause an increase in the DPOAE threshold of about same amount. Considering only the effect of the decrease in CA gain, the two different dysfunctions cause the same shift, represented by the two parallel arrows in Fig. 13.

Next, add in the differential effects of the changes that occur in the two different types of dysfunction. If noise damage (at these sound levels) operates mainly through loss or damage to OHCs, at a given traveling wave amplitude there would be an added decrease in emission amplitude solely due to the decrease in the number of sources of the emissions, namely in the numbers of functional OHCs. However, because the effect of the OHCs on the IHCs operates only through the CA gain, in this case (solid vertical line) there would no additional change in ABR threshold. In contrast, with acute EP decline there is no known loss of OHCs. At a given traveling wave amplitude, it is at least plausible to argue that the resulting emission amplitude might be unchanged or changed little. However, there is evidence that a decrease in EP has direct effects on IHC and associated eighth-nerve responses that are well in excess of those due to the decrease in cochlear amplifier gain alone (Sewell, 1984a, b, c; Rübsamen *et al.*, 1995). Therefore, for an EP decrease the additional change (dashed line) to be added in Fig. 13 is primarily an increase in ABR threshold due to the decrease in drive to the IHCs, with little or no change in the emissions.

Now consider the implications of these results in more detail (Fig. 9). There is substantial separation between the responses measured in the two experimental conditions, each taken by themselves. While there was clearly some intrinsic scatter, the correlations between the emission and ABR measures were found to be reasonable. For example, the r^2 values for each of the two least-square linear fit lines shown in Fig. 9(B) were about 0.6. However, if one *combines* the two conditions, the r^2 value for the whole distribution in Fig. 9(B) drops to 0.3. Making similar measurements in a human group, Boege and Janssen (2002) found an r^2 of about 0.4 for an equivalent correlation. The present results suggest that the underlying reason for such a relatively poor correlation is not a defect in either emissions or threshold measurements, but the fact that the two methods are measuring different things. Rather than attempt to derive one from the other, it seems preferable to use both. The different answers obtained from each measure then provide the possibility of noninvasive, differential diagnosis of human hearing dysfunction.

To demonstrate further how this might work, suppose that the two conditions tested were the main causes of hearing dysfunction in some natural population, and that an elevation of 20 dB in auditory threshold was considered the threshold for hearing problems. This threshold is indicated in Fig. 12 by the dashed vertical line. What would be the implications for differential diagnosis in this population?

Since nearly all of the overlap region is located where changes in emission or ABR threshold shifts are less than 20 dB, any combination of threshold shifts that are actually in the region denoted as hearing dysfunction could be unequivocally assigned to one of the two conditions. From Eq. (1), if the hearing threshold (ABR) is elevated 40 dB, for

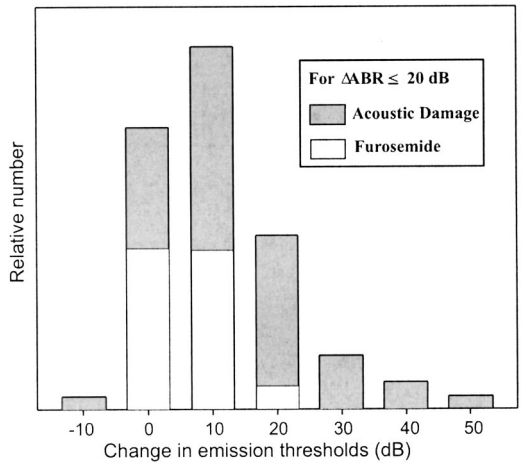


FIG. 14. Distribution of emission thresholds found with ABR changes of 20 dB or less, for case with $L_1 = L_2 + 20$ dB [Fig. 9(B)]. No normal, unexposed animals were included in this figure.

example, the line separating the two conditions is located at an emission threshold shift, $\Delta x_t = 32$ dB. If the measured increase in emission threshold, Δx, is less than 32 dB, then the cause is furosemide intoxication. If it is greater, then the cause would be ascribed to acoustic damage.

There is an additional implication in the data shown in Figs. 9 and 12. The distribution of emission shifts found suggests that there would be a number of individuals who have hearing thresholds in the ''normal'' range, i.e., with thresholds elevated less than 20 dB, but who have emission thresholds elevated from 20 to 50 dB. The distributions of the emission thresholds in this group from this study are summarized in Fig. 14.

Even in the normal population, the responses shown in Figs. 9, 12, and 14 suggest that there will be observed a large variation of emission threshold. That is, for ΔABR $\leqslant$20 dB, emission thresholds are seen to range from 10 dB below the normal, unexposed mean to 50 dB above. This suggests that if a study population includes a distribution of individuals with exposure to cochlear stresses, the correlation between pure-tone thresholds and emission responses even in the normal-hearing group would be expected to be poor. That is, in any ''normal-hearing'' population with underlying and varying degrees of cochlear exposure, there would be expected to be a significant number of individuals with elevations of emission thresholds above normal, in the range of 20 to 40 dB SPL. For the two exposures studied here, nearly all such elevations would be due to subclinical acoustic exposure (Fig. 14). This is because furosemide effects typically cause more elevation in hearing (or ABR) thresholds than in emission thresholds, where the opposite is true for acoustic damage. Not all acoustic exposure causes this signature, as some moderate acoustic exposure subjects were found in the ''overlap'' region noted (Fig. 12). Nonetheless, elevations of emission threshold into the region marked ''preclinical acoustic damage'' might represent a potentially useful early warning sign of noise damage. Such a result has been suggested for humans (e.g., Lucertini *et al.*, 2002), based on measurements in normal-hearing ears of individuals who have noise damage evident in the other ear. Of course, only one type of cochlear dysfunction was examined in the Lucertini *et al.* study, and only two types of cochlear dysfunction were examined in the present study. It is possible that there are additional clinically relevant conditions which could cause a similar weakening of emissions while hearing thresholds remained in the normal range (loss$\leqslant$20 dB)

To summarize more concretely the findings to this point, consider that an investigator is presented with subjects who have potentially been exposed to one of the two conditions studied here, and is challenged to make a diagnosis on the basis of noninvasive tests. The clinician would measure the pure-tone or ABR thresholds, and the emission growth functions for $L_1 = L_2 + 20$ dB, compare these to the norms in Fig. 2. He or she would then make the diagnosis as follows, using Fig. 12 as a guide.

(1) *Both emission and ABR thresholds normal within ±10 dB.* Normal, unexposed individual.
(2) *ABR threshold shift $\leqslant$20 dB, but emission threshold elevated more than 20 dB.* Exposure to sound has caused moderate damage, but hearing function has not yet been seriously affected.
(3) *ABR threshold elevated above 20 dB.* Given the measured ABR threshold, calculate the dividing emission threshold shift, Δx_t, from Eq. (1). If the observed emission threshold is above Δx_t at this frequency, the diagnosis is hearing dysfunction caused by acoustic damage. Otherwise, the dysfunction is attributable to furosemide.

While suggestive, the applicability of the results here to differential diagnosis of human hearing dysfunction remains to be firmly established. Given the measurements of Lucertini *et al.* (2002), it seems likely that the signature of the differential effect of noise damage on pure-tone thresholds and DPOAEs in humans may be found to be quite similar to that seen in this study, as a result of acoustic damage in gerbils. It is less obvious that the differential effects of chronic strial dysfunction will be identical to that seen here with acute furosemide injection. That a similar signature is at least possible, however, is suggested by results of a 3-year study of quiet-aged gerbils (Schmiedt and Schulte, 1992; Boettcher *et al.*, 1995; Gratton *et al.*, 1996). This study concluded that the main cause of the increase in ABR thresholds in these animals was not loss of OHCs, but declines in the EP caused by strial dysfunction. Significantly, there were found to be large increase in ABR thresholds compared to modest shifts in emission amplitudes. While the results of this aging study are therefore consistent with those found here with furosemide, a direct comparison is not possible because equivalent parametric emission measurements were not made.

This initial study was offered as a proof of concept. The concept is that noninvasive differential diagnosis for auditory system dysfunction might be possible using the combination of specific otacoustic emission measurements and standard auditory threshold measures. It was shown that two different experimental manipulations did have different effects on DPOAEs compared to ABRs, and that it was possible to find stimulus parameters and emission measures which adequately divided the observed responses into two mostly

nonoverlapping distributions. At least for these two cochlear dysfunctions, then, the combination of audiometric status and specific emission measurements provides a method of noninvasive differential diagnosis that can clearly distinguish between the two underlying conditions.

ACKNOWLEDGMENTS

Thanks to Barbara Cone-Wesson, George Gates, Ed Rubel, and Rick Schmiedt for useful discussions, and to Natalie Hardie for assistance with recovery surgery techniques and for useful suggestions on a draft of the manuscript. This work was supported by grants DC 04077 and DC 04661 from the National Institute for Deafness and Other Communication Disorders, National Institutes of Health.

Boege, P., and Janssen, T. (**2002**). ''Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears,'' J. Acoust. Soc. Am. **111**, 1810–1818.

Boettcher, F. A., Gratton, M. A., and Schmiedt, R. A. (**1995**). ''Effects of noise and age on the auditory system,'' Occup. Med. Rev. **10**, 577–591.

Borg, E., and Engstrom, B. (**1983**). ''Hearing thresholds in the rabbit,'' Acta Oto-Laryngol. **95**, 19–26.

Davis, R. I., and Ferraro, J. A. (**1984**). ''Comparison between AER and behavioral thresholds in normally and abnormally hearing chinchillas,'' Ear Hear. **5**, 153–159.

Dorn, P. A., Piskorski, P., Gorga, M. P., Neely, S. T., and Keefe, D. H. (**1999**). ''Predicting audiometric status from distortion product otoacoustic emissions using multivariate analysis,'' Ear Hear. **20**, 149–163.

Gates, G. A., Mills, D., Nam, B.-h., D'Agostino, R., and Rubel, E. W. (**2002**). ''Effects of age on the distortion-product otoacoustic emission growth functions,'' Hear. Res. **163**, 53–60.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). ''From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss,'' Ear Hear. **18**, 440–455.

Gratton, M. A., Schmiedt, R. A., and Schulte, B. A. (**1996**). ''Age-related decreases in endocochlear potential are associated with vascular abnormalities in the stria vascularis,'' Hear. Res. **102**, 181–190.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). ''On the detection of early cochlear damage by otoacoustic emission analysis,'' J. Acoust. Soc. Am. **111**, 972–978.

Mills, D. M. (**1997**). ''Interpretation of distortion product otoacoustic emission measurements. I. Two stimulus tones,'' J. Acoust. Soc. Am. **102**, 413–429.

Mills, D. M. (**2002**). ''Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response,'' J. Acoust. Soc. Am. **112**, 1545–1560.

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.

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., and Rubel, E. W. (**1996**). ''Development of the cochlear amplifier,'' J. Acoust. Soc. Am. **100**, 428–441.

Mills, D. M., and Shepherd, R. K. (**2001**). ''Distortion product otoacoustic emission and auditory brainstem responses in the echidna (*Tachyglossus aculeatus*),'' J. Assoc. Res. Otolaryngol. **2**, 130–146.

Neely, S. T., and Stover, L. J. (**1993**). ''Otoacoustic emissions from a nonlinear, active model of cochlear mechanics,'' in *Proceedings of the International Symposium on Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, River Edge, NJ), pp. 64–71.

Puel, J.-L., Ruel, J., d'Aldin, C. G., and Pujol, R. (**1998**). ''Excitotoxicity and repair of cochlear synapses after noise-trauma induced hearing loss,'' NeuroReport **9**, 2109–2114.

Rübsamen, R., Mills, D. M., and Rubel, E. W. (**1995**). ''Effects of furosemide on distortion product otoacoustic emissions and on neuronal responses in the anteroventral cochlear nucleus,'' J. Neurophysiol. **74**, 1628–1638.

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.

Ryan, A. (**1976**). ''Hearing sensitivity of the mongolian gerbil, *Meriones unguiculatis*,'' J. Acoust. Soc. Am. **59**, 1222–1226.

Schmiedt, R. A., and Schulte, B. A. (**1992**). ''Physiologic and histopathologic changes in quiet- and noise-aged gerbil cochleas,'' in Noise-Induced Hearing Loss, edited by A. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik (Mosby-Year Book, St. Louis).

Sewell, W. F. (**1984a**). ''The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats,'' Hear. Res. **14**, 305–314.

Sewell, W. F. (**1984b**). ''Furosemide selectively reduces one component in rate-level functions from auditory-nerve fibers,'' Hear. Res. **15**, 69–72.

Sewell, W. F. (**1984c**). ''The relation between the endocochlear potential and spontaneous activity in auditory nerve fibers of the cat,'' J. Physiol. (London) **347**, 685–696.

Trautwein, P., Hofstetter, P., Wang, J., Salvi, R., and Nostrant, A. (**1996**). ''Selective inner hair cell loss does not alter distortion product otoacoustic emissions,'' Hear. Res. **96**, 71–82.

Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability

Benjamin Munson
Department of Communication Disorders, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Drive SE, Minneapolis, Minnesota 55455

Gail S. Donaldson and Shanna L. Allen[a)]
Department of Otolaryngology, University of Minnesota, 8323 Philips Wangensteen Building, 516 Delaware Street SE, Minneapolis, Minnesota 55455

Elizabeth A. Collison[b)]
Department of Communication Disorders, University of Minnesota 115 Shevlin Hall, 164 Pillsbury Drive SE, Minneapolis, Minnesota 55455

David A. Nelson
Department of Otolaryngology, University of Minnesota, 8323 Philips Wangensteen Building, 516 Delaware Street SE, Minneapolis, Minnesota 55455

(Received 22 March 2002; revised 14 November 2002; accepted 18 November 2002)

Many studies have noted great variability in speech perception ability among postlingually deafened adults with cochlear implants. This study examined phoneme misperceptions for 30 cochlear implant listeners using either the Nucleus-22 or Clarion version 1.2 device to examine whether listeners with better overall speech perception differed qualitatively from poorer listeners in their perception of vowel and consonant features. In the first analysis, simple regressions were used to predict the mean percent-correct scores for consonants and vowels for the better group of listeners from those of the poorer group. A strong relationship between the two groups was found for consonant identification, and a weak, nonsignificant relationship was found for vowel identification. In the second analysis, it was found that less information was transmitted for consonant and vowel features to the poorer listeners than to the better listeners; however, the pattern of information transmission was similar across groups. Taken together, results suggest that the performance difference between the two groups is primarily quantitative. The results underscore the importance of examining individuals' perception of individual phoneme features when attempting to relate speech perception to other predictor variables.
[DOI: 10.1121/1.1536630]

PACS numbers: 43.64.Me, 43.66.Ts, 43.66.Lj, 43.71.Ky, 43.71.Es [CWT]

I. INTRODUCTION

A large research literature has determined that cochlear implantation has the potential to greatly improve the speech perception of postlingually deafened adults. Postimplantation, improvements can be noted in phoneme perception scores, open-set word recognition, and sentence recognition. As implant technology has evolved, all of these measures have improved, with some listeners who utilize the most recent devices demonstrating speech perception in quiet approaching that of normal-hearing listeners [e.g., Wilson (2000)]. Nevertheless, there continues to be considerable variability in speech perception across individuals.

A large number of investigations have attempted to identify factors accounting for the variability in speech perception by individuals with cochlear implants. Studies have examined factors related to the implanted device [e.g., Tyler *et al.* (1996)], the speech processing strategy [e.g., Kompis *et al.* (1999); Osberger and Fisher (1999); Skinner *et al.* (1996, 1999)], listeners' psychophysical abilities [e.g., Busby and Clark (1999); Donaldson and Nelson (2000)], and higher-level cognitive and linguistic factors [e.g., Collison *et al.* (2002); Lyxell *et al.* (1998); Sarant *et al.* (1997)]. Although studies have related each of these factors to variability in speech perception performance, there is no consensus on their relative contributions to overall performance.

Most investigations of speech perception in individuals with cochlear implants have used percent-correct scores for vowel and consonant identification as the dependent measures of phoneme perception. These measures have the advantage of providing summary scores of phoneme perception, and appear to correlate with various factors purported to underlie speech perception. However, they have the disadvantage of collapsing performance across different phonemes, *potentially obscuring differences among listeners' error patterns*. Related to this, percent-correct scores may obscure qualitative differences between groups of listeners related to their use of particular acoustic and perceptual fea-

[a)]Currently affiliated with the Henry Ford Hospital, Detroit, MI.
[b)]Currently affiliated with the Indiana University—Purdue University of Indiana Medical Center, Indianapolis, IN.

tures. To address these deficits, a few studies have included analyses of consonant and vowel confusion matrices [e.g., Skinner *et al.* (1996, 1999)]. Confusion matrices catalogue the amount and types of misperceptions made by individual listeners or groups of listeners. They provide useful information regarding the specific phonemes that are most difficult for listeners to perceive, and common patterns of misperceptions. Additionally, confusion matrices can be analyzed using a variety of multivariate statistical techniques, such as log-linear modeling and multidimensional scaling, which provide information regarding the specific perceptual features that are most problematic for listeners.

Very few studies have compared patterns of phoneme errors in better- versus poorer-performing individuals with cochlear implants. Recently, one such study was reported by Van Wiringen and Wouters (1999). These investigators examined the phoneme perception errors of a group of individuals using the Laura cochlear implant. Listeners were divided arbitrarily into three groups based on their speech perception performance. Information transmission and multidimensional scaling analyses suggested that the three groups of listeners were using different features to perceive vowels and consonants, and that the better-performing listeners perceived features more efficiently. This study was limited, in that the high-performing listeners had only moderate speech perception performance: mean correct vowel and consonant identification for their best group of listeners was 66% and 49%, respectively. As a result, subgroups were not well separated on speech perception measures, and findings for the highest-performing group may not be applicable to the best users in current clinical populations.

The purpose of the present study is to further examine the differences between better-and poorer-performing cochlear implant users, by analyzing consonant and vowel misperceptions for two groups that were well separated in overall speech perception ability. A finding that the two groups differ qualitatively in their speech perception might suggest that different intervention strategies be used to enhance speech perception in the two groups. For example, a finding that poorer-performing listeners have more difficulty than better-performing listeners with vowel-height features but not vowel-backness features would imply that a remediation strategy should focus on the speech coding strategies targeting the first-formant frequency region. Analyses were designed to determine whether the speech perception errors made by poorer listeners were only quantitatively different than those made by better listeners (i.e., the poorer listeners made the same types of errors as better listeners, only more of them) or both quantitatively and qualitatively different (i.e., the poorer listeners made more errors than better listeners, and made errors on different features than the better listeners). In the first analysis, simple regression is used to compare better-and poorer-performing listeners' percent-correct scores for individual phonemes. The second analysis examines information transmission for vowel and consonant features to the two groups. Results from the two analyses suggest that the two groups differ primarily quantitatively in their perception of vowel and consonant sounds.

II. METHODS

A. Participants

Thirty postlingually deafened adults with cochlear implants participated. Summary data for the 30 participants are given in Table I. Twelve listeners were implanted with the Nucleus-22 device and used the SPEAK processing strategy. The other 18 listeners were implanted with the Clarion device; 13 of these listeners used the continuous interleaved sampling (CIS) strategy, four used the paired pulsatile (PPS) strategy, and one used the simultaneous analog (SAS) strategy. Listeners were heterogeneous with respect to many demographic variables, including etiology, age at onset, and duration of implant use.

B. Speech perception tests

Vowel and consonant recognition data were obtained using a standard phoneme-confusion procedure. Vowel stimuli were 11 /hVd/ monosyllables from the database of Hillenbrand *et al.* (1995), spoken by three male talkers. Only male talkers were used, because the female talkers in Hillenbrand *et al.*'s database demonstrated much greater variability in formant frequency than the male talkers, both within and between vowels. Vowels tested were /i, ɑ, ɛ, eɪ, ɚ, ɪ, æ, oʊ, ʊ, ʌ, u/ as in *heed*, *hod*, *head*, *hayed*, *heard*, *hid*, *had*, *hoed*, *hood*, *HUD* and *who'd*. Consonant stimuli were 19 /aCa/ disyllables from the stimulus set of Van Tasell *et al.* (1992), spoken by three male and three female talkers. Consonants tested were /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, m, n, r, l, j/.

Testing was conducted in a sound-isolated room, with the participant seated approximately 1 meter in front of a pair of high-quality loudspeakers and a video screen. Digitized speech stimuli were played out from computer memory, low-pass filtered at half the digitization rate, amplified, and presented through the speakers. The stimulus was presented once on each trial, and the subject used a computer mouse to select his or her response from a list of possible alternatives displayed on the video screen. The alternative choices presented in the vowel identification task were real words (e.g., *had, heed*), whereas alternatives for the consonant task were nonsense words (e.g., *awa, atha*). As in Van Tasell *et al.* (1992), correct-answer feedback was provided immediately after each stimulus presentation. Van Tasell *et al.* (1992) argued that it was appropriate to use feedback with examining performance of listeners who may not be well practiced in speech-perception tasks. The large number of different stimuli per block (six tokens for each of the 19 consonants or the 11 vowels) likely prevented subjects from remembering idiosyncratic features of individual tokens and identifying them based on those features. Stimulus level was calibrated such that average speech peaks of a file containing a concatenated sample of the speech perception stimuli reached 60 dB SPL on the slow, A scale of a Bruel & Kjær type 2203 sound-level meter at the location of the listener's head.

Vowel and consonant data were obtained in separate test sessions. For each stimulus type (vowels and consonants), one practice block and five standard blocks of data were obtained. Practice blocks were comprised of two trials per vowel phoneme (33 trials) or three trials per consonant pho-

TABLE I. Subject demographic information. Subjects ordered from highest mean consonant and vowel identification (top) to lowest (bottom).

Subject code	Age at onset (years)	Duration of profound deafness (years)	Duration of implant use (years; months)	NU-6 % words correct[a]	Etiology of deafness	Device-strategy	Group
N13	50	2	10;5	46	Progressive	Nucleus-SPEAK	Better
C04	36	0	0;3	66	Progressive	Clarion-CIS	Better
N29	3	28	1;7	90	Progressive	Nucleus-SPEAK	Better
N12	32	8	10;5	54	Progressive	Nucleus-SPEAK	Better
C15	33	7	1;0	68	Progressive	Clarion-CIS	Better
C12	34	13	0;10	···	Otosclerosis	Clarion-CIS	Better
C07	23	0	1;11	34	Progressive	Clarion-CIS	Better
N32	5	24	2;8	44	Rubella	Nucleus-SPEAK	Better
C14	16	47	1;6	76	Unknown	Clarion-CIS	Better
N14	49	0	6;9	88	Progressive	Nucleus-SPEAK	Better
C05	42	0	3;1	66	Unknown	Clarion-CIS	Better
C03	22	27	3;0	74	Progressive	Clarion-PPS	Better
C02	18	19	2;7	86	Unknown	Clarion-CIS	Better
C01	29	13	3;10	74	Progressive	Clarion-SAS	Better
N31	50	21	10;9	0	Noise	Nucleus-SPEAK	Poorer
N05	50	5	8;4	0	Otosclerosis	Nucleus-SPEAK	Poorer
C09	0	41	0;6	···	Unknown	Clarion-PPS	Poorer
N07	50	4	11;10	6	Cogan's	Nucleus-SPEAK	Poorer
C13	72	6	1;2	50	Noise	Clarion-CIS	Poorer
N30	50	8	3;11	24	Otosclerosis	Nucleus-SPEAK	Poorer
C19	30	32	0;5	36	Progressive	Clarion-PPS	Poorer
C10	25	23	1;3	16	Meningitis	Clarion-PPS	Poorer
C06	50	12	2;1	24	Progressive	Clarion-CIS	Poorer
N09	56	0	11;0	24	Meniere's	Nucleus-SPEAK	Poorer
C08	5	26	3;9	···	Ototoxicity	Clarion-CIS	Poorer
C20	28	31	1;3	48	Progressive	Clarion-CIS	Poorer
C17	23	17	0;10	32	Progressive	Clarion-CIS	Poorer
N22	66	1	6;3	30	Noise	Nucleus-SPEAK	Poorer
N28	57	0	4;9	58	Meningitis	Nucleus-SPEAK	Poorer
C11	43	0	0;10	···	Kearns-Sayre	Clarion-CIS	Poorer

[a]··· indicates data not available.

neme (38 trials). Standard blocks were comprised of six trials per phoneme (66 vowels or 114 consonants) presented in random sequence. Occasionally, performance was observed to improve over the first few standard blocks. When this occurred, additional standard blocks were obtained until performance was stable for five consecutive blocks, and the final five blocks were retained. A merged confusion matrix was created from the five standard blocks of data for a particular subject. Each merged matrix represented 30 observations (5blocks×6tokens) per stimulus.

Subjects used their own speech processors for all testing. Nucleus subjects had a Spectra speech processor programmed in the SPEAK strategy. Clarion subjects had a v1.2 or S-series speech processor programmed in the CIS, PPS, or SAS strategy. Nucleus subjects adjusted the sensitivity controls on their processors to achieve comfortable loudness for the test stimuli. Clarion subjects adjusted the volume control, leaving the sensitivity control set to a level (''10:30'' for v 1.2 users, ''11:00'' for S-series users) at which AGC compression would not be activated for the stimulus levels used.

The peak level chosen for presentation of stimuli in this study, 60 dB, is lower than that used in some other studies. A separate study [Donaldson and Smith (1999)] examined the effect of presentation level on perception of the same stimuli as were used in the current study. Because of the different ways that intensity is coded by the Nucleus and Clarion devices, presentation level may have impacted users of the two devices differentially. The Nucleus-22 SPEAK processor has only a sensitivity control, which adjusts the dynamic range of sounds encoded by the device upward and downward with respect to absolute sound levels. The dynamic range of the Nucleus device is approximately 30 dB: a given sensitivity setting might encode sounds between 30 and 60 dB SPL, whereas a higher sensitivity setting might encode sounds between 20 and 50 dB SPL. Sounds greater than those coded at the top of the dynamic range are all mapped to the top of the electric dynamic range, whereas sounds softer than the selected range are not coded at all. For testing, each Nucleus-22 subject adjusted the sensitivity of his/her processor so that the 60-dB phoneme stimuli were comfortably loud. The same sensitivity setting was used for the testing of consonants, vowels, and NU-6 words. Donaldson and Smith (1999) found that sound-field thresholds were quite variable across individuals utilizing the Nucleus-22 device. For the speech frequencies, they ranged from about 35 to 50 dB SPL, with the average thresholds for speech frequencies being 42.5 dB, and most occurring between 35 and 45 dB. Based on this observation, audibility may have been an issue for some cues; however, data on identification of consonants at different intensity levels did not support this hypothesis.

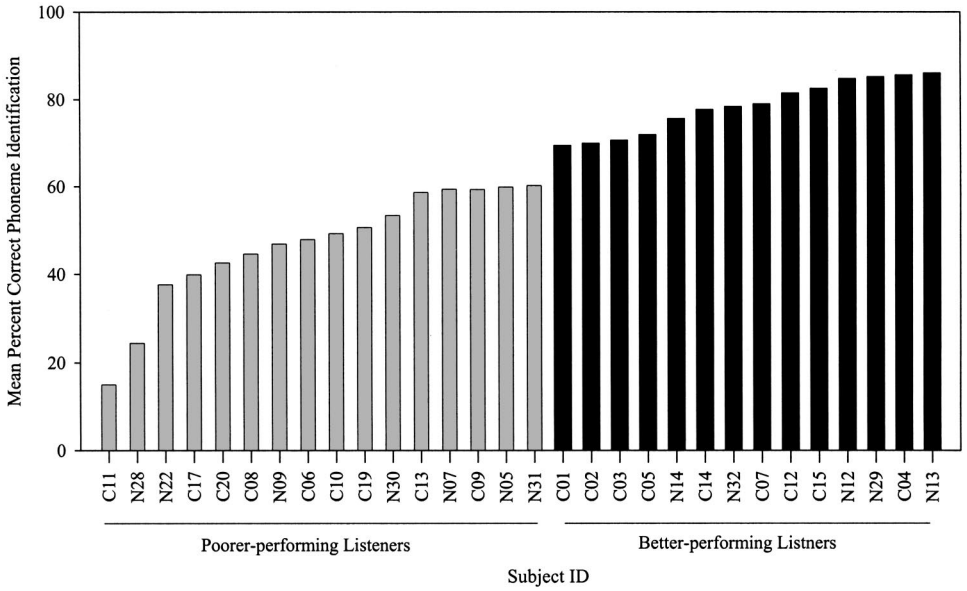


FIG. 1. Total percent-correct vowel and consonant identification for individual listeners.

For consonants, subjects showed small or absent improvements in performance when presentation level was increased from 60 to 70 dB SPL. When analyzed by feature, the average data showed no change between 60 and 70 dB for voicing and place features, but did show about 10% improvement in transmitted information for manner features. For vowels, subjects showed fairly constant performance between 50 and 70 dB, so audibility would not seem to be an issue at 60 dB.

In contrast, the Clarion device has both a sensitivity control and a volume control. Subjects set their sensitivity to the default setting (''10:30'' or ''11:30'' depending on processor type), which sets the top of the encoded dynamic range to approximately 70 dB SPL. They then adjusted the volume control to achieve comfortable loudness for 60-dB stimuli. The volume control for this device moves the top of the *electric* dynamic range up or down to alter loudness. For the Clarion device, the range of speech sounds encoded (referred to as *input dynamic range* [IDR]) is adjustable for values between 30 and 60 dB. Almost all subjects used an IDR of 50 or 60 dB. As would be expected from this wider input dynamic range, Clarion subjects had more sensitive soundfield thresholds than Nucleus subjects, average 28 dB SPL for the speech frequencies. Clarion subjects reached asymptotic performance for consonants by 60 dB and for vowels by 50 dB. For many individual subjects, asymptotic performance was reached for a lower presentation level.

In summary, data from Donaldson and Smith (1999) suggest that audibility could be a minor factor influencing consonant performance in Nucleus subjects, but not for vowel performance in Nucleus subjects, and not for consonant or vowel performance in Clarion subjects.

III. RESULTS

A. Descriptive data

1. Percent-correct scores

The analyses presented in this section compare the performance of better- and poorer-performing listeners with cochlear implants. For the purposes of this study, we defined better- and poorer-performing listeners using the following technique. First, we calculated total percent-correct identification for consonants and vowels. Second, we examined individual listeners' scores on this composite measure. In doing so, we observed that no listeners had composite scores between 60% and 69%. This is illustrated in Fig. 1.

We divided the larger group of listeners into two subgroups. The better-performing group of listeners ($n=14$) had a mean composite score of 78.5% (s.d.$=6.1$, range $=[69\%-86\%]$). The poorer-performing group of listeners ($n=16$) had a mean composite score of 46.9% (s.d.$=13$, range$=[16\%-60\%]$). The two groups were well separated in their percentage-correct scores for vowels ($M=86.6\%$, s.d.$=5.8$ for the better listeners, $M=53.7\%$, s.d.$=16$ for the poorer listeners) and consonants ($M=70.4\%$, s.d.$=7.8$ for the better listeners, $M=40\%$, s.d.$=12.8$ for the poorer listeners). Nucleus-22 and Clarion users were equally distributed across the two groups, as indicated by a chi-square test ($\chi^2=0.36, p>0.05$). Mann-Whitney U tests indicated that the two groups did not differ in their age at onset of deafness, duration of profound deafness, or duration of implant use. Percent-correct scores for identification of NU-6 words in quiet were available for 13 of the poorer listeners and 13 of the better listeners. Mann-Whitney U tests confirmed that the two groups differed in their word perception ($z=-3.774$, $p<0.001$). The better listeners had a mean percent-correct score of 67% (s.d.$=18\%$). The poorer-performing group had a mean score of 27% (s.d.$=18\%$). Each listener's group membership is noted in Table I.

2. Common patterns of phoneme confusions

Prior to completing statistical analyses, we conducted a descriptive analysis of the most common patterns of phoneme confusions made by the two groups of listeners. The confusion matrices for vowels and consonants by the poorer- and better-performing groups are presented in Tables II–V.

Phoneme confusions are reported in the form response/stimulus (i.e., *t*/*k* means *the stimulus* [*k*] *was misperceived to*

TABLE II. Vowel confusion matrix, poorer listeners.

	æ	ɑ	ɛ	eɪ	ɚ	ɪ	i	oʊ	ʊ	ʊ	u
æ	303	28	66	20	30	16	12	4	2	4	1
ɑ	143	207	11	6	61	2	1	22	6	22	5
ɛ	19	5	186	14	4	186	11	4	25	30	2
eɪ	21	6	16	305	12	17	98	3	2	2	4
ɚ	8	11	29	8	240	13	2	36	47	27	65
ɪ	6	0	66	12	3	356	30	1	4	7	1
i	10	0	11	49	6	18	387	4	1	0	0
oʊ	2	17	7	4	33	2	2	231	31	8	149
ʊ	7	10	28	4	22	39	4	31	257	47	37
ʌ	11	19	64	3	19	38	6	15	164	134	13
u	10	12	7	7	54	5	1	94	38	12	246

be [*t*]). The ten most common vowel confusions for the poorer-performing group of listeners were æ/ɑ, ɛ/ɪ, ɛ/ʌ, ɚ/ɑ, ɪ/ɛ, i/eɪ, oʊ/u, ʊ/ʌ, u/ɚ, and u/oʊ. These confusions suggest problems perceiving a variety of vowel features. The pairs i/eɪ and oʊ/u suggest problems with height; æ/ɑ, ɛ/ɪ, ɛ/ʌ, and, ɪ/ɛ suggest problems with backness, and ɚ/ɑ and u/ɚ indicate a difficulty perceiving the r-coloring feature of the vowel /ɚ/. The ten most common misperceptions made by the better-performing group of listeners were æ/ɑ, ɛ/ɪ, ɛ/æ, ɚ/u, ɪ/ɛ, oʊ/u, ʊ/ɚ, ʊ/ʌ, u/ʌ, and u/ʊ. Five of the most frequent misperceptions were common to the two groups. Again, these confusions show a problem perceiving a variety of vowel features.

The ten most common consonant confusions for the poorer-performing group of listeners were t/p, k/p, t/k, g/d, f/θ, f/s, ʃ/s, v/ð, m/n, and l/r. These confusions indicate particular difficulty with obstruent sounds: eight of the ten target stimuli were stops or fricatives, six of which were voiceless. The ten most common consonant confusions for the better-performing group of listeners were t/p, t/k, g/d, f/θ, θ/f, θ/s, v/ð, ð/v, ð/z, and m/n. Again, these errors show a particular difficulty perceiving obstruent sounds, as nine of the targets were stops or fricatives. Unlike the poorer group of listeners, the target obstruent sounds most often misperceived by the better-performing listeners were equally likely to be voiced and voiceless; the poorer-performing listeners were most likely to make errors perceiving voiceless obstruents. In general, both groups showed similar patterns of confusion for consonants. Six of the most common misperceptions were evidenced by both of the two groups. In addition, two of better-performing users' most common confusions, θ/f and θ/s, were among the poorer-performing listeners' most common misperceptions, albeit not among the ten most common. Only two of the better-performing listeners' errors, ð/v and ð/z, were not among the poorer-performing listeners' most common misperceptions. This is not surprising; given that [ð] was the poorer-performing listeners' least-common response to the set of stimuli.

In general, the commonality in error vowel and consonant error patterns between the two groups suggests that perception of these two classes of sounds differs quantitatively between the two groups. Both groups demonstrate difficulty in perceiving place of articulation features in obstruent consonants, and a variety of vowel features.

B. Regression analysis of percent-correct scores

In the first statistical analysis, we computed mean percent-correct scores for each phoneme for both the better- and poorer-performing groups of listeners, and measured the degree of association between groups using simple regression analyses. We reasoned that the strength of the association would be a measure of the extent to which two groups' error patterns were qualitatively similar. If the poorer-performing users' errors differ only quantitatively from the better-performing users', then we would expect a close-to-perfect measure of association (i.e., an R^2 approaching 1.0). In contrast, if the performance of the poorer-performing group of listeners were to some degree qualitatively different from the better-performing listeners', then a less-than-perfect measure of association would be expected. In addition, we examined the pattern of outlying residuals in the regression

TABLE III. Vowel confusion matrix, better listeners.

	æ	ɑ	ɛ	eɪ	ɚ	ɪ	i	oʊ	ʊ	ʌ	u
æ	358	4	26	2	0	0	0	0	0	0	0
ɑ	25	357	1	0	1	0	0	1	0	4	1
ɛ	8	1	310	1	0	55	1	0	3	11	0
eɪ	0	0	1	374	3	2	10	0	0	0	0
ɚ	0	0	3	0	342	3	1	3	22	3	13
ɪ	0	0	48	4	2	332	3	0	1	0	0
i	0	0	3	14	0	0	371	2	0	0	0
oʊ	0	0	0	0	0	1	1	328	1	3	56
ʊ	1	1	6	1	9	1	1	7	320	9	34
ʌ	0	2	10	0	2	0	0	1	52	322	1
u	0	0	0	1	22	0	0	60	8	4	295

TABLE IV. Consonant confusion matrix, poorer listeners.

	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	m
p	188	98	106	12	6	15	17	13	5	1	3	3	2	2	0
t	55	234	135	2	2	5	8	22	3	4	1	3	2	2	0
k	72	115	242	1	1	4	7	11	9	4	0	5	1	0	1
b	32	3	6	190	55	60	31	21	3	0	33	10	3	0	16
d	13	12	9	42	151	200	9	11	0	0	7	7	0	0	4
g	14	16	15	15	80	268	7	9	3	1	11	5	6	2	2
f	39	14	13	11	1	4	177	86	55	17	20	13	5	8	3
θ	44	16	12	9	8	4	171	127	50	5	5	19	4	0	0
s	24	8	6	7	7	4	102	79	128	46	10	21	11	18	1
ʃ	2	5	0	0	1	0	5	1	49	353	1	2	11	46	1
v	15	9	8	55	16	39	16	42	19	2	101	22	25	6	26
ð	10	7	6	66	24	28	22	51	7	2	101	23	21	8	17
z	9	2	5	22	11	14	17	39	31	32	57	9	79	43	5
ʒ	1	4	2	1	0	3	4	3	7	103	5	2	54	258	2
m	7	0	7	23	2	5	19	4	4	1	14	2	6	2	217
n	7	6	6	8	21	12	4	4	5	1	16	2	12	2	67
r	4	0	2	14	2	7	5	6	5	0	56	5	10	5	22
l	5	1	0	17	1	4	5	1	3	1	12	2	5	0	78
j	2	5	5	1	4	10	4	3	6	12	11	2	24	13	6

analysis. Our rationale for this analysis was that the phonemes perceived differently by the two groups would be associated with the largest residuals. A similar technique has been used previously to examine differences in reaction times between children with slower and faster processing of linguistic and nonlinguistic stimuli [Lahey *et al.* (1999)].

Vowels and consonants were examined separately. In both analyses, the mean scores for phoneme identification for the poorer-performing group served as the dependent variable in the regression, with the mean scores for the better-performing group serving as the independent variable. For consonants, a strong, significant association was found between scores for the two groups of listeners [$F(1,17) = 117.823$, $p < 0.001$, $R^2 = 0.874$]. The relationship between the two groups is shown in Fig. 2. We chose to examine phonemes whose standardized residuals were greater than 1.0 or less than −1.0, indicating sounds that fell outside of the 68% confidence interval of the regression line. When the standardized residuals were examined, it was found that the consonants /ʃ/ and /f/ had standardized residuals greater than 1.0 ($z = 2.23$ for /ʃ/, $z = 1.19$ for /f/). Thus, the poorer-performing listeners were identifying the consonants /ʃ/ and /f/ with greater accuracy than would be predicted by the performance of the better-performing listeners: mean correct identification for /ʃ/ was 74% for the poorer-performing group of listeners and 93% for the better-performing group of listeners; /f/ was identified with 37% accuracy for the poorer-performing group of listeners and 58% by the better-performing group of listeners. In contrast, the consonants /l/, /n/, /r/, and /p/ were found to have standardized residuals less than −1 ($z = -1.05$ for /n/, $z = -1.24$ for /r/, $z = -1.3$ for /l/, $z = -1.73$ for /p/). That is, the poorer-performing listeners identified the consonants /l/, /n/, /r/, and /p/ with lower accuracy than would be predicted by the performance of the better-performing group of listeners. Mean percent-correct scores for the two groups were as follows: for /l/, $M = 27\%$

TABLE V. Consonant confusion matrix, better listeners.

	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	m
p	342	50	13	1	0	1	3	2	1	1	0	0	0	0	0
t	30	311	64	0	0	0	0	3	0	6	0	0	0	0	0
k	25	52	329	0	0	1	2	2	0	0	0	1	0	0	0
b	0	0	0	300	50	7	11	12	0	0	15	11	1	0	4
d	0	0	0	7	238	160	1	0	0	0	0	6	2	0	0
g	0	2	2	5	27	371	2	2	1	0	0	2	2	0	0
f	5	2	0	7	1	2	240	60	61	7	15	8	3	0	2
θ	2	0	0	7	2	1	135	198	49	0	8	12	1	0	1
s	1	1	1	2	0	3	49	94	237	6	3	11	7	0	0
ʃ	0	3	0	0	0	0	0	0	18	385	0	0	1	8	0
v	8	3	0	47	8	0	11	13	1	1	187	61	29	0	16
ð	0	0	0	40	21	18	2	32	0	0	98	130	39	0	6
z	1	0	0	1	12	13	8	16	27	11	15	60	202	13	0
ʒ	0	1	0	0	0	0	0	0	2	15	0	0	23	364	0
m	1	0	0	2	0	0	0	2	0	0	2	2	0	0	315
n	1	0	0	0	3	0	0	0	0	0	1	3	0	0	19
r	0	0	0	0	0	0	0	3	0	0	11	1	1	0	0
l	1	0	2	0	0	0	0	0	0	0	4	0	1	0	19
j	0	0	0	0	0	0	0	0	0	0	3	0	3	0	0

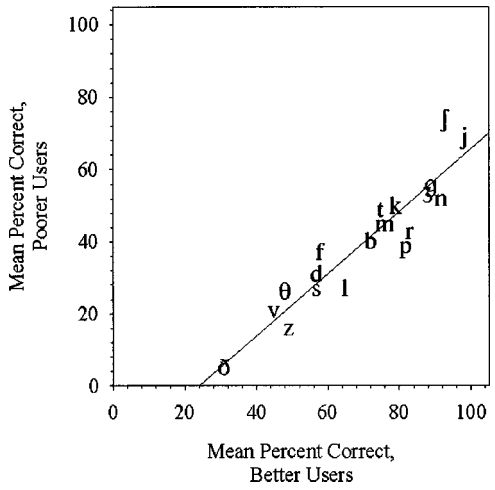


FIG. 2. Mean percent of consonants correctly perceived by the better listeners (*x* axis) and the poorer listeners (*y* axis).

for the poorer group and $M=65\%$ for the better-performing group; for /n/ $M=52\%$ for the poorer-performing group and $M=92\%$ for the better-performing group; for /r/, $M=43\%$ for the poorer-performing group and $M=83\%$ for the better-performing group; for /p/, $M=39\%$ for the poorer-performing group and $M=82\%$ for the better-performing group.

When vowels were examined, a weak, statistically nonsignificant relationship between the two groups was found ($F[1,9]=3.213$, $p=0.107$, $R^2=0.263$). The relationship between the two groups is shown in Fig. 3. Residuals were not examined for vowels: because the slope of the regression line was not significantly different from zero, the magnitude of the standardized residuals could not be considered stable estimates.

A possible explanation for the difference in regression results for consonants and vowels is that a greater range of performance was noted in consonants than in vowels. For consonants, poorer-performing listeners had a range of 69% in their percent-correct scores, from 5% for the phoneme /ð/ to 74% for the phoneme /ʃ/. Better-performing listeners had a range of 68%, from 31% for the phoneme /ð/ to 98% for the

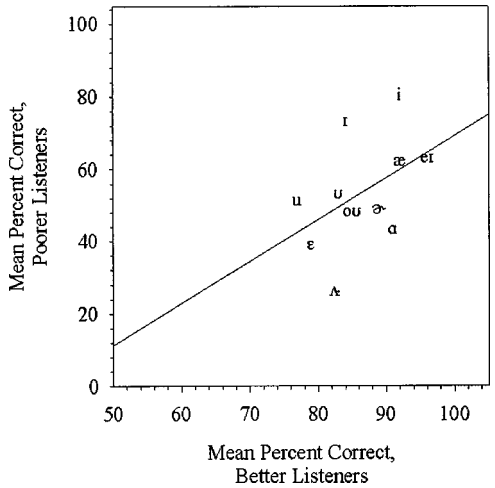


FIG. 3. Mean percent of vowels correctly perceived by the better listeners (*x* axis) and the poorer listeners (*y* axis).

phoneme /j/. Performance on vowels was less variable. For the poorer listeners, performance ranged 53%, from 27% for /ʌ/ to 80% for the vowel /i/. Performance of the better listeners ranged only 19%, from 77% for the vowel /u/ to 96% for the vowel /eɪ/. In addition, fewer vowels than consonants were examined, suggesting that the lack of significance in the regression on vowel performance may have been due to statistical power.

To examine whether these factors influenced the different regression results for vowels and consonants, we ran two additional regressions on subsets of the consonant performance data. First, we ranked the consonants based on the percent-correct scores of the poorer-performing listeners. Based on this, we divided the consonants into two subsets of nine consonants: poorly perceived consonants (/z/, /v/, /θ/, /s/, /l/, /d/, /f/, /p/, and /b/) and better-perceived consonants (/r/, /m/, /t/, /k/, /n/, /ʒ/, /g/, /j/, /ʃ/). The consonant demonstrating the poorest perception by both groups, /ð/, was omitted from this analysis so that the two subsets of consonants could be equal in size. The range of performance for the poorly perceived consonants was 24% for the poorer listeners (from 16% for /z/ to 40% for /b/) and 37% for the better listeners (from 45% for /v/ to 82% for /p/). The range of performance for the better-perceived consonants was 31% for the poorer listeners (from 43% for /r/ to 74% for /s/) and 33% for the better listeners (from 75% for /t/ to 98% for /j/). Separate regressions were run on these two subsets of consonants. These separate regressions were more comparable to the regression analyses of vowel data, both in terms of range of performance and number of items. Even with fewer items and a smaller range of performance, both of the regressions were significant [$F(1,7)=11.962$, $p=0.011$, $R^2=0.63$ for the poorly perceived consonants; $F(1,7)=10.363$, $p=0.015$, $R^2=0.60$ for the better-perceived consonants]. These analyses suggest that the difference in vowel and consonant regression results between the two groups may not be due to differences in range of performance or number of items. However, this conclusion is limited by the fact that the smallest range of performance on the consonant subsets (24% for the poorest-perceived consonants by the poorer-performing listeners) was still larger than the 19% range in the better-performing listeners' vowel perception.

A second set of analyses compared the percent-correct data for individual poorer-performing listeners to the mean data for the better-performing group of listeners. Each of the 16 poorer-performing users' percent correct vowel and consonant identification served as the dependent variable in a series of regressions, with the mean scores for the better-performing group of listeners serving as the independent variable. The results of these 32 regressions, including R^2 and p values, supported the group analysis: the majority of poorer-performing listeners ($n=13$) showed significant relationships for consonants, but not for vowels. In contrast to the descriptive analysis of confusion matrices, these results suggest that the two groups differ qualitatively and quantitatively in the perception of vowels, but not in the perception of consonants. Again, this conclusion is limited by the fact that larger ranges in performance in the better-performing

TABLE VI. Features used for vowels and consonants in the information transmission analysis.

Vowels			Consonants		
Feature	Categories	Items	Feature	Categories	Items
Tense/lax	tense	/i,e,æ,ɑ,o,u,/	voicing	voiced	/b,d,g,v,ð,z, ʒ,m,n,j,r,l/
	lax	/ɪ,ɛ,ʊ,ʌ/		voiceless	/p,t,k,ʃ,f,θ,s,S/
*f*0	low	/i,e,æ,ɑ/	duration	short	/d,g,k,b,p,t,j/
	mid	/ɪ,u,ɛ/		medium	/l,n,m,v,ð,r/
	high	/o,ʊ,ɚ,ʌ/		long	/ʒ,z,f,ʃ,θ,s/
Height	low	/æ,ɑ/	place	labial	/p,b,m,f,v/
	mid	/e,ɛ,ʌ,ɚ,o/		coronal	/θ,ð,t,d,n,l,r/
	high	/i,ɪ,u,ʊ/		palatal	/ʃ,ʒ,j/
Backness	back	/o,ʊ,u/		velar	/k,g/
	central	/ɑ,ʌ,ɚ/	manner	plosive	/p,b,t,d,k,g/
	front	/i,ɪ,e,ɛ,æ/		fricative	/f,v,θ,ð,s,z,ʃ,ʒ/
r-coloring	r-colored	/ɚ/		sonorant	/m,n,j,r,l/
	non-r-colored	/i,ɪ,e,ɛ,æ,ɑ,o, ʊ,u,ʌ/			

listeners were noted for consonant perception than for vowel perception.

C. Information transmission for better- and poorer-performing listeners

The second statistical analysis used sequential information transmission analysis [SINFA, Wang and Bilger (1973)] to examine whether better- and poorer-performing users' vowel and consonant perception differed for individual features. We reasoned that if the two groups' perception of vowels and consonants differed qualitatively, an analysis of variance (ANOVA) would reveal a significant interaction between group and feature.

The features used for the analysis of consonants and vowels are listed in Table VI. The features used for consonants were voicing, place of articulation, manner of articulation, and duration. The first three of these features are traditionally used in describing consonants [e.g., Ladefoged (2001)]. The feature duration was included in case some of the poorest listeners were unable to use enough spectral and envelope cues to receive any information about voicing, place, or manner, and were able to perceive only the duration of consonants. Values for the duration feature were determined by measuring the mean durations of the consonantal portion of the /ɑCɑ/ stimuli and categorizing them as short (mean duration of 100 to 110 ms), medium length (mean duration of 123 to 136 ms), or long (mean duration of 152 to 222 ms). These categories were chosen arbitrarily, based on the observation that there were no consonants with mean values between 110 and 123 ms, or between 136 and 152 ms. The features height, backness, r-coloring, tense/lax, and *f*0 were used for vowels. For the feature *f*0, mean *f*0's were measured for the stimuli, and vowels were classified as either low pitch (mean *f*0 of 104 to 112 Hz), medium pitch (mean *f*0 of 118 to 122 Hz), or high pitch (mean *f*0 of 131 to 152 Hz).

SINFA was used to calculate the information transmitted by each feature to each listener. Information transmission scores for individual listeners were then submitted to a two-way mixed-model ANOVA. Vowels and consonants were analyzed separately. For consonants, feature (voicing, place, manner, and duration) was the within-subjects factor and group (better versus poorer listeners) was the between-subjects factor. Mean information transmitted for the four features to listeners in the two groups can be found in Fig. 4. A significant main effect of group was found [$F(1,28) = 38.308$, $p < 0.001$], with less information transmitted to the poorer listeners than to the better listeners. A significant main effect of feature was also found [$F(3,84) = 53.455$, $p < 0.001$]. All *post hoc* differences were significant, with

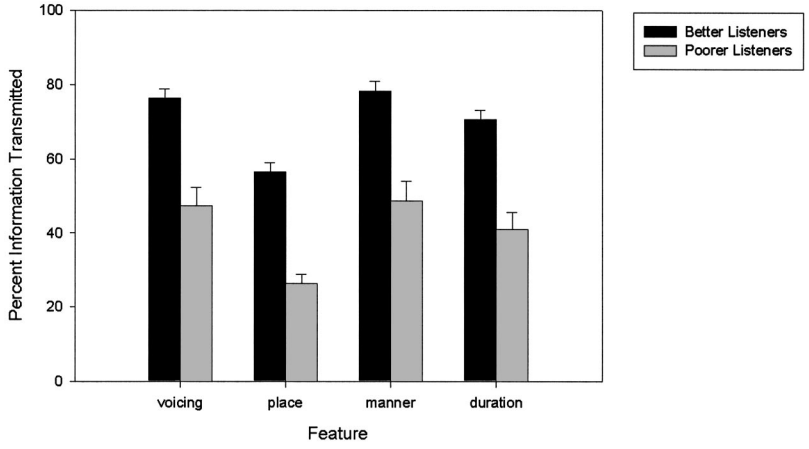


FIG. 4. Percentage of information transmitted for consonant features. Error bars represent one standard error of measurement.

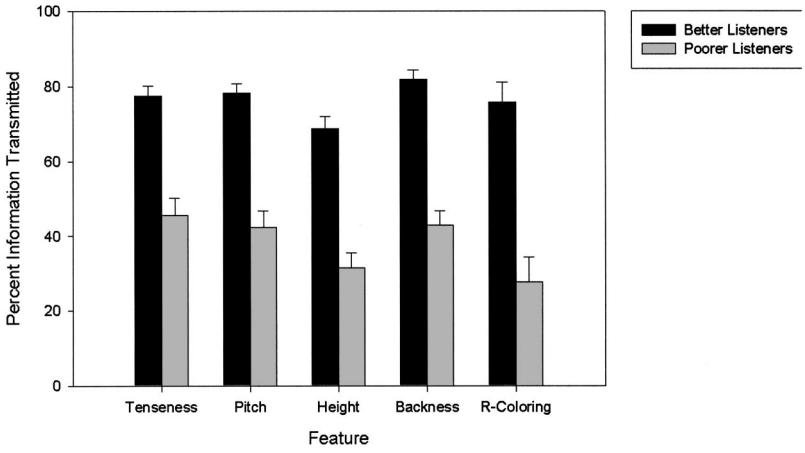


FIG. 5. Percentage of information transmitted for vowel features. Error bars represent one standard error of measurement.

the exception of that between voicing and manner of articulation. No significant group by feature interaction was found [$F(3,84)=0.049$, $p=0.985$], suggesting that the two groups did not differ qualitatively in their perception of consonant features.

The second ANOVA examined information transmitted for vowels. Again, information transmission scores were submitted to a two-way mixed model ANOVA, with feature (tense/lax, $f0$, height, backness, and r-coloring) as the within-subjects factor and group (poorer versus better) as the between-subjects factor. Mean information transmitted for the five features to the listeners in the two groups can be found in Fig. 5. A significant main effect of group was found [$F(1,28)=60.414$, $p<0.001$]. Less information was transmitted to the poorer-performing listeners than to the better-performing listeners. A significant main effect of feature was also found [$F(4,112)=8.796$, $p=0.001$]. Less information about the feature height (corresponding to $F1$ frequency) was transmitted than the other features. The group by feature interaction was nearly significant [$F(4,112)=2.353$, $p=0.057$]. Visual inspection of the data suggests that the group difference between better- and poorer-listeners was greater for the features height ($F1$ frequency) and r-coloring ($F3$ frequency) than for the other three features, with large within-group variability noted for the r-coloring feature. Although this interaction was not significant, the trend is consistent with the findings of the regression analyses, suggesting that the two groups may differ qualitatively in their perception of vowels.

The participants of this study included both Nucleus-22 and Clarion users. As noted earlier, there were no significant differences in the number of Clarion and Nucleus-22 users in the better- and poorer-performing groups of listeners. However, given that these two devices and their associated speech-processing strategies code sound very differently, we were interested in examining whether information transmitted would be different for the two groups. Two mixed-model ANOVAs examined the influence of group membership (the between-subjects factor) and feature (the within-subjects factor) on information transmitted. No significant effect of group was found for either analysis. In addition, no group-by-feature interactions were found, despite the potential small difference in audibility of consonant features by users of the Nucleus-22 and Clarion devices.

The consonant results are similar to those in previous studies, some of which used a different set of consonant stimuli and consonant features [Fu and Shannon (1999); McKay and McDermott (1993); Skinner *et al.* (1996, 1999); Tyler and Moore (1992); Van Wiringen and Wouters (1999); Vandali (2001)]. In all of these studies, place of articulation information was more poorly transmitted than information about manner and voicing. There is less previous research concerning information transmission in vowel perception to listeners with cochlear implants. A previous study comparing vowel feature perception for Nucleus users utilizing SPEAK speech-processing strategies was presented by Skinner *et al.* (1999). Consistent with the current study, Skinner *et al.* (1999) found that more information was transmitted for backness ($F2$) features than for height or r-coloring features. An earlier study examined information transmission in vowels for listeners with an Ineraid cochlear implant [Tyler *et al.* (1992)], and found that the greatest amount of information was transmitted for duration and $F1$ frequency. This result contrasts with the results for the current study, in which the least information was transmitted for the $F1$ feature. This may be attributable to differences between the devices and speech processing strategies used in that study and in the current study.

The results of the information-transmission analysis suggested that the two groups of listeners did not differ qualitatively in their perception of either vowel or consonant features. No group-by-feature interactions were found, although a nonsignificant trend was noted for poorer perception of height and r-coloring features by the poorer-performing group of listeners.

IV. DISCUSSION

This study examined patterns of phoneme misperceptions by two groups of cochlear implant listeners, those with relatively better speech perception performance, and those with relatively poorer performance. The three analyses performed in this paper gave conflicting findings. Regression analyses of the two groups' percent-correct vowel and consonant identification scores showed a strong, significant relationship for consonant perception and a weak, nonsignificant relationship for vowel perception. This suggests that poorer-performing listeners' consonant identification differed

only quantitatively from better-performing listeners', while the vowel perception differed both quantitatively and qualitatively. A larger range of performance was noted in consonant perception than in vowel perception for the two groups. In addition, more consonants were included in the analysis than vowels. Thus, the apparent qualitative difference between the two groups' vowel perception may have been due in part to the restriction-of-range and ceiling effects in the better-performing group of listeners.

In contrast to the regression findings, descriptive analysis of confusion matrices found similar patterns of confusion for both vowels and consonants, and information transmission analysis suggested that the same relative amount of information of vowel and consonant features was being transmitted to listeners in the two groups. Results of the information transmission analysis suggested that the poorer-performing listeners were receiving less information about height (which is correlated with mean $F1$ values) and r-coloring (which is correlated with mean $F3$ values) than $F0$, $F2$, and duration. Taken together, these findings suggest that the poorer-performing listeners had greater difficulty perceiving spectral information in the frequency region of $F1$ and $F3$. The better-performing listeners were able to use spectral information from three distinct spectral bands, $F1$, $F2$, and $F3$, while the poorer-performing listeners were able, in general, to hear differences best in $F2$. This finding may suggest that the poorer-performing listeners would benefit from a speech-processing strategy in which more electrodes are dedicated to representation of $F1$ and $F3$ information. This is finding contrasts with those of Henry *et al.* (2000) and McKay and Henshall (2002), who found that low-frequency information was best transmitted to the poorest-performing listeners with Nucleus-22 cochlear implants.

In general, the information suggesting that the two groups did not differ qualitatively on either vowel or consonant perception was stronger than evidence to the contrary. This finding is important, because it suggests that techniques to enhance the speech perception of people who use cochlear implants need not take into account overall level of functioning. The results of this study expand on the results of Van Wiringen and Wouters's (1999) study of the speech perception of better-, intermediate-, and poorer-performing individuals with cochlear implants. Van Wiringen and Wouters (1999) found that better-, intermediate-, and poorer-performing listeners used different features to perceive consonants, while the two groups of listeners in the current study appeared to use the same features. The differences between that study and the current one may be due to differences in the overall levels of performance in the two studies; to differences the different devices used by the listeners; to differences in the homogeneity of performance within the different groups; and to differences in the languages being examined.

In contrast to the regression analyses, the SINFA analysis found that the two groups used the same features to perceive both vowels and consonants. One problematic aspect of SINFA analysis is that it assumes that articulatory features are relevant for speech perception. While early theories of speech perception [e.g., Liberman and Mattingly (1985)] emphasized the possible articulatory basis of speech perception, more recent theories have emphasized the importance of acoustic information and auditory processing [e.g., Kluender and Lotto (1999)]. Consonant features such as [place] and [voice] are much more uniform articulatorily than acoustically. For example, the feature [place] has different acoustic correlates depending on the manner of articulation of the consonant being produced. Measures of the reception of the feature [place] in consonants provides limited information regarding the specific acoustic parameters that were used, and thus provides relatively little information on which parameters should be enhanced in individuals demonstrating poor reception of place features. Vowel features used in SINFA analyses are much more transparently related to acoustic features. For example, the feature [height] is well correlated with $F1$ frequency, and the feature [back] is well correlated with $F2$ frequency. The results of the information-transmission analysis might have been affected by the different relationships between articulation and acoustics for consonants and vowels.

In summary, these results provide tentative support for the hypothesis that listeners varying in overall performance do not differ qualitatively in their consonant and vowel perception. The fact that not all analyses arrived at this conclusion underscores the methodological difficulties of assessing the perception of specific vowel and consonant features with tests of vowel and consonant identification accuracy. Researchers wishing to understand the features that are misperceived by listeners with cochlear implants must make inferences based on patterns of consonant and vowel identification in these tasks. As an alternative to this, researchers could measure speech perception by examining the identification and discrimination of a small number of speech contrasts whose perception depends on a single acoustic parameter, rather than with mean data on accuracy of identification of a large set of phonemes.

ACKNOWLEDGMENTS

We gratefully acknowledge James Hillenbrand for providing the vowel identification stimuli, and for sharing the formant frequency and duration measures for the vowels used in the speech perception experiments. We thank Dianne Van Tasell for providing the stimuli used in the consonant identification experiment. We thank Peggy Nelson, Arlene Carney, the two anonymous reviewers, and the editor for valuable comments on this work. We also thank Edward Carney for technical assistance. This project was supported by NIDCD grant number P01-DC00110 and by the Lions 5M International Hearing Foundation.

Bosman, A. (**1996**). "Confusion analysis in the assessment of speech perception and hearing aids," in *Psychoacoustics: Speech and Hearing Aids*, edited by B. Kollmeier (World Scientific, Singapore).

Busby, P. A., and Clark, G. M. (**1999**). "Gap detection by early-deafened cochlear-implant subjects," J. Acoust. Soc. Am. **105**, 1841–1852.

Collison, E., Munson, B., and Carney, A. (**2002**). "Relationships among vocabulary size, nonverbal cognition, and spoken word recognition in adults with cochlear implants," J. Acoust. Soc. Am. **111**, 2428(A).

Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.

Donaldson, G. S., and Smith, S. L. (**1999**). ''Speech performance-intensity functions for Nucleus SPEAK and Clarion CIS cochlear implant listeners,'' American Academy of Audiology Annual Convention, Miami Beach, FL.

Fu, Q.-J., and Shannon, R. (**1999**). ''Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant,'' Ear Hear. **20**, 321–331.

Henry, B., McKay, C., McDermott, H., and Graeme, C. (**2000**). ''The relationship between speech perception and electrode discrimination in cochlear implantees,'' J. Acoust. Soc. Am. **108**, 1269–1280.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K., (**1995**). ''Acoustic characteristics of American English Vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Kluender, K., and Lotto, A. (**1999**). ''Virtues and perils of an empiricist approach to speech perception,'' J. Acoust. Soc. Am. **105**, 503–511.

Kompis, M., Vischer, M. W., and Hausler, R. (**1999**). ''Performance of compressed analogue (CA) and continuous interleaved sampling (CIS) coding strategies for cochlear implants in quiet and noise,'' Acta Oto-Laryngol. **119**, 659–664.

Ladefoged, P. (**2001**). *A Course in Phonetics*, 4th ed. (Harcourt College Publishers, New York).

Lahey, M., Edwards, J., and Munson, B. (**1999**). ''SLI: General or Specific Processing Limitations?'' Miniseminar presented at the 1999 Annual Convention of the American Speech-Language-Hearing Association, San Francisco, CA.

Liberman, A., and Mattingly, I. (**1985**). ''The motor theory of speech perception revisited,'' Cognition **21**, 1–36.

Lyxell, B., Andersson, J., Andersson, U., Arlinger, S., Bredberg, G., and Harder, H. (**1998**). ''Phonological representation and speech understanding with cochlear implants in deafened adults,'' Scand. J. Psychol. **39**, 175–179.

McKay, C. M., and Henshall, K. R. (**2002**). ''Frequency-to-electrode allocation and speech perception with cochlear implants,'' J. Acoust. Soc. Am. **111**, 1036–1044.

McKay, C. M., and McDermott, H. J. (**1993**). ''Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP),'' Ear Hear. **14**, 350–367.

Osberger, M. J., and Fisher, L. (**1999**). ''SAS–CIS preference study in postlingually deafened adults implanted with the Clarion cochlear implant,'' Ann. Otol. Rhinol. Laryngol. **177** (Suppl), 74–79.

Sarant, J. Z., Blamey, P. J., Cowan, R. S., and Clark, G. M. (**1997**). ''The effect of language knowledge on speech perception: What are we really assessing?'' Am. J. Otol. **18**, (Suppl), 135–137.

Skinner, M., Fourakis, M., Holden, T., Holden, L., and Demorest, M. (**1996**). ''Identification of speech by cochlear implant recipients with the Multipeak (MPEAK) and Spectral Peak (SPEAK) speech coding strategies. I. Vowels,'' Ear Hear. **17**, 182–197.

Skinner, M., Fourakis, M., Holden, T., Holden, L., and Demorest, M. (**1999**). ''Identification of speech by cochlear implant recipients with the Multipeak (MPEAK) and Spectral Peak (SPEAK) speech coding strategies. II. Consonants,'' Ear Hear. **20**, 443–460.

Tyler, R. S., Gantz, B. J., Woodworth, G. G., Parkinson, A. J., Lowder, M. W., and Schum, L. K. (**1996**). ''Initial independent results with the Clarion cochlear implant,'' Ear Hear. **17**, 528–536.

Tyler, R. S., and Moore, B. C. (**1992**). ''Consonant recognition by some of the better cochlear-implant patients,'' J. Acoust. Soc. Am. **92**, 3068–3077.

Tyler, R. S., Preece, J. P., Lansing, C. R., and Gantz, B. J. (**1992**). ''Natural vowel perception by patients with the Ineraid cochlear implant,'' Audiology **31**, 228–239.

Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). ''Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants,'' J. Acoust. Soc. Am. **92**, 1247–1257.

Van Wiringen, A., and Wouters, J. (**1999**). ''Natural vowel and consonant recognition by Laura cochlear implantees,'' Ear Hear. **20**, 89–103.

Vandali, A. (**2001**). ''Emphasis of short-duration acoustic speech cues for cochlear implant users,'' J. Acoust. Soc. Am. **109**, 2049–2061.

Wang, M., and Bilger, R. (**1973**). ''Consonant confusions in noise: A study of perceptual features,'' J. Acoust. Soc. Am. **54**, 1248–1266.

Wilson, B. (**2000**). *Sixth Quarterly Progress Report, Speech Processors for Auditory Prostheses* (Research Triangle Institute, Research Triangle Park, NC).

The importance of cochlear processing for the formation of auditory brainstem and frequency following responses

Torsten Dau[a)]
Carl von Ossietzky Universitat Oldenburg, Medizinische Physik, D-26111 Oldenburg, Germany

(Received 12 February 2002; accepted for publication 28 October 2002)

A model for the generation of auditory brainstem responses (ABR) and frequency following responses (FFRs) is presented. The model is based on the concept introduced by Goldstein and Kiang [J. Acoust. Soc. Am. **30**, 107–114 (1958)] that evoked potentials recorded at remote electrodes can theoretically be given by convolution of an elementary unit waveform (unitary response) with the instantaneous discharge rate function for the corresponding unit. In the present study, the nonlinear computational auditory-nerve model recently developed by Heinz *et al.* [ARLO **2**(3), 91–96 (2001)] was used to calculate the instantaneous discharge rate $r_i(t)$ for fibers i in the frequency range from 0.1 and 10 kHz. The summed activity across frequency was convolved with a unitary response which is assumed to reflect contributions from different cell populations within the auditory brainstem, recorded at a given pair of electrodes on the scalp. Predicted potential patterns are compared with experimental data for a number of stimulus and level conditions. Clicks, chirps as defined in Dau *et al.* [J. Acoust. Soc. Am. **107**, 1530–1540 (2000)], long-duration stimuli comprising the chirp, as well as tones and slowly varying tonal sweeps were considered. The results demonstrate the importance of considering the effects of the basilar-membrane traveling wave and auditory-nerve processing for the formation of ABR and FFR. Specifically, the results support the hypothesis that the FFR to low-frequency tones represents synchronized activity mainly stemming from mid- and high-frequency units at more basal sites, and not from units tuned to frequencies around the signal frequency. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1534833]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.64.Bt [LHC]

I. INTRODUCTION

Evoked responses represent the summation of responses from many neurons, recorded from electrodes placed on the surface of the head (e.g., Jewitt, 1970), i.e., remote to the individual neurons. Auditory evoked potentials can be recorded from all levels of the auditory system. They are usually grouped by the time of occurrence after the onset of the stimulus, and this grouping corresponds roughly to the site of generation. The click-evoked auditory brainstem response (ABR) waveform generally consists of seven peaks, all occurring within the first 10 ms after the signal onset. Of the seven peaks, waves I, III, and V are sufficiently robust for clinical use. The most robust peak, wave V, can be elicited at near-threshold levels.

The cellular generators of these short-latency potentials reside in the lower auditory system (e.g., Buchwald and Huang, 1975), but exactly which cells generate the various components of the ABR is not yet fully understood. Goldstein and Kiang (1958) introduced the concept that remote responses generated by auditory-nerve (AN) neurons can theoretically be given by the convolution of an elementary unit waveform, the unitary response, with the instantaneous rate function at which the AN cell discharges in response to the stimulus. Following this concept, Melcher and Kiang (1996) suggested in a more general description that the potential produced by *any* cell in the auditory pathway can be described by the convolution of the instantaneous rate function at which this cell responds to the stimulus with the unitary potential which is produced each time this cell discharges.

Such a concept has been used in the past to model whole-nerve action potentials (APs; de Boer, 1975; Biondi *et al.*, 1975; Elberling, 1974). Simplified models based on the tuning curves of the AN fibers acting as linear filters, and containing some basic characteristics of the auditory periphery, were used to predict the characteristics of APs in response to clicks and low-and high-frequency brief tones. Using these models several basic properties of the AP could be theoretically predicted. It was shown that the assumed type of frequency selectivity, coupled with the AN fiber latency as a function of the characteristic frequency (CF), is the major determinant of the AP (e.g., de Boer, 1975).

The role of the CF-dependent delays, imposed by the traveling wave on the basilar membrane (BM), for the generation of the compound AP (CAP) was investigated experimentally in a study by Shore and Nuttall (1985). They used tone bursts of exponentially rising frequency to hypothetically activate synchronous discharges of AN fibers along the length of the cochlear partition. Their equations defining the chirps were calculated to be the inverse of the delay-line characteristic of the guinea pig partition. Shore and Nuttall (1985) recorded CAPs in response to the rising chirp and compared them to CAP waveforms evoked by corresponding falling chirps as well as clicks. Their analysis of the CAP waveforms showed narrower widths and larger amplitudes for rising sweeps when compared to falling sweeps. Their results supported the hypothesis underlying the derivation of

a)Electronic mail: torsten.dau@medi.physik.uni-oldenburg.de

the rising chirp: spectral energy with the appropriate temporal organization, determined by BM traveling wave properties, increases CAP synchrony.

Based on the BM model of de Boer (1980), Dau *et al.* (2000) developed a chirp stimulus which theoretically produces synchronous discharge maxima of AN fibers along the length of the *human* cochlear partition, and studied ABR generation using this stimulus. Compatible with the results of Shore and Nuttall in the guinea pig, ABR elicited by the rising chirp showed a larger wave-V amplitude than did click-evoked responses for most stimulation levels tested. Like the CAP, the ABR depends on the synchronous activity of the underlying neural responses. Since the ABR wave V reflects a neural response from the brainstem, the effect of an optimized synchronization at the peripheral level thus can also be observed at the brainstem level. The data demonstrated that the ABR cannot be considered as an electrophysiological event purely evoked by the onset or offset of a stimulus. Instead, an appropriate temporal organization, determined by BM traveling wave properties, can increase neural synchrony (Dau *et al.*, 2000; Wegner and Dau, 2002).

Temporal dispersion on the BM has also been postulated to play an important role for the generation of frequency following responses (FFR; Janssen *et al.*, 1991). The FFR mimics the periodicity of tonal stimuli up to about 1 kHz and can be evoked at intensities substantially above the threshold of hearing (e.g., Marsh *et al.*, 1975; Stillman *et al.*, 1976; Moushegian *et al.*, 1973). Based on (*i*) recordings of tone-pulse evoked responses at low frequencies in the presence of high-pass noise maskers and (*ii*) model predictions assuming a linear bandpass filtering stage for BM transformation (de Boer, 1980), a simulation of inner hair-cell transduction (Russel and Cody, 1985), and a spike generation mechanism (Jones *et al.*, 1985), Janssen *et al.* (1991) concluded that FFR result from synchronized neural activity in basal frequency channels; activity in low-frequency channels around the signal frequency is less synchronized and does not contribute effectively to the response. However, only one signal level (95 dB SPL) was considered in the experiments and a linear BM model was used such that level-dependent mechanisms in the cochlea could not be investigated.

In the meantime, a number of studies have been published investigating the nonlinear properties of BM and AN processing. BM tuning has been shown to broaden with increases in level and to demonstrate compressive magnitude responses and nonlinear phase responses (Rhode, 1971; Ruggero *et al.*, 1997). Two-tone suppression is another nonlinear response property of the auditory periphery, and refers to the ability of an off-frequency tone to suppress BM and AN responses to CF tones (Sachs and Kiang, 1968; Delgutte, 1990; Ruggero *et al.*, 1992) whereby many of the observed nonlinear AN response properties are likely to result from the same mechanism, the cochlear amplifier (Sachs and Abbas, 1974; Sewell, 1984; Patuzzi *et al.*, 1989; Ruggero and Rich, 1991; Ruggero *et al.*, 1992).

The goal of the present study is to relate ABR and FFR waveforms to AN activity using the basic modeling concept introduced by Goldstein and Kiang (1958) and Melcher and Kiang (1996), i.e., that the evoked potential is equal to the convolution between a function dependent on the ''excitation'' by the stimulus and a function dependent on the basic activity in the contributing auditory neurons, the unitary response. The recently developed nonlinear computational AN model by Heinz *et al.* (2001) is used in the present study to calculate the instantaneous discharge rate functions (representing the excitation) for fibers in the range from 0.1 to 10 kHz. The Heinz *et al.* model describes many of the important response properties mentioned above associated with the cochlea amplifier, including broadened tuning with increases in level and the compressive-magnitude responses (see also Heinz, 2000). These aspects were not considered in the earlier modeling studies on potential generation (e.g., de Boer, 1975; Elberling, 1976) but are likely to be important for a realistic description. For each stimulus, the summed neural activity pattern, reflecting the whole AN activity summed across CF, is convolved with a unitary response function which is intended to represent contributions from the different cell populations along the auditory brainstem associated with the generation of the ABR waves. The assumed unitary response is estimated once, in advance, by deconvolution of click ABR data (obtained at a high stimulation level) with the corresponding neural activity function for the click, generated by the AN model. This unitary response function is then held constant throughout the study and is used to model the intensity-dependent aspects for a variety of stimuli.

First, the general modeling approach for the generation of ABR and the structure of the present model will be described. Predicted potential patterns will then be shown for transient stimuli like clicks and chirps at various levels. The same model will also be applied to more complex long-duration stimuli comprising the chirp as well as tones and tonal sweeps that generate ''classical'' FFR. The predictions will be compared with corresponding experimental data from the literature. Finally, capabilities and limitations of the modeling approach will be discussed.

II. MODEL FOR ABR GENERATION

A. The general modeling approach

Following the concept of Goldstein and Kiang (1958), Melcher and Kiang (1996) developed a model for the generation of auditory brainstem responses. The ABR was assumed to be the sum of potentials v_i produced by individual cells i in response to the stimulus s, combined across all the cells in the auditory brainstem:

$$\mathrm{ABR}(t,\overline{x_1},\overline{x_2},s)=\sum_i v_i(t,\overline{x_1},\overline{x_2},s). \tag{1}$$

The potential v_i depends on time (t), the locations of the recording electrodes ($\overline{x_1},\overline{x_2}$), and the stimulus ($s$). v_i produced by any given cell depends on two terms: The first one is the potential produced between $\overline{x_1}$ and $\overline{x_2}$ each time the cell discharges, the unitary response $u_i(t,\overline{x_1},\overline{x_2})$. The second one is the instantaneous rate function, $r_i(t,s)$, at which the cell discharges in response to the stimulus. Thus,

$$v_i=u_i(t,\overline{x_1},\overline{x_2})*r_i(t,s), \tag{2}$$

where $*$ denotes convolution. Both quantities are related to measurable cellular properties. Instantaneous discharge rates are directly derivable from measurements of poststimulus time (PST) histograms in response to acoustic stimuli (Johnson, 1978). The unitary potential waveform depends on the morphological and electrical properties of the cell within the context of the entire head (Melcher and Kiang, 1996). Thus, combining (1) and (2) leads to

$$\mathrm{ABR}(t,\overline{x_1},\overline{x_2},s)=\sum_i u_i(t,\overline{x_1},\overline{x_2})*r_i(t,s). \tag{3}$$

This form of the model requires that all the cells in the auditory brainstem are considered individually. Melcher and Kiang (1996) suggested considering groups of cells collectively, based on established anatomical and physiological criteria. The ABR can then be written as sum of potentials V_p, produced by any population p of cells:

$$\mathrm{ABR}(t,\overline{x_1},\overline{x_2},s)=\sum_{p=1}^{P} V_p(t,\overline{x_1},\overline{x_2},s) \tag{4}$$

$$=\sum_{p=1}^{P}\sum_{i=1}^{N_p} v_{i,p}(t,\overline{x_1},\overline{x_2},s) \tag{5}$$

$$=\sum_{p=1}^{P}\sum_{i=1}^{N_p} u_{i,p}(t,\overline{x_1},\overline{x_2})*r_{i,p}(t,s), \tag{6}$$

where N_p is the number of cells in population p, and P is the number of contributing populations. It can be assumed that each cell i in a given population p produces the same unitary potential $u_{i,p}=u_p$ since cells of a given physio-anatomical type generally have, by definition, similar morphological and electrical properties, the factors which determine the unitary potential waveform (Melcher, 1995; Melcher and Kiang, 1996):

$$\mathrm{ABR}(t,\overline{x_1},\overline{x_2},s)=\sum_{p=1}^{P}\left(u_p(t,\overline{x_1},\overline{x_2})*\sum_{i=1}^{N_p} r_{i,p}(t,s)\right) \tag{7}$$

$$=\sum_{p=1}^{P}\left(u_p(t,\overline{x_1},\overline{x_2})*R_p(t,s)\right). \tag{8}$$

Thus, in order to calculate the ABR waveform for a particular stimulus and electrode configuration, one needs to know the unitary waveforms, u_p, of the cellular generators which generate the specific extrema in the ABR, as well as the corresponding summed discharge rate functions, R_p, for the cell populations p. The ''stationary peaks'' in the ABR are most likely the result of propagating action potentials (e.g., Kimura *et al.*, 1984; Özdamar and Delgedo, 1990). Specifically, Stegeman *et al.* (1987) demonstrated that propagating action potentials in homogeneous populations of neurons do produce a far field response when there is (*i*) a change in electrical conductivity of the medium, (*ii*) a change in the volume conductor's geometry, or (*iii*) a change in the direction of the action potential. This provides strong constraints on the possible generator sites of the peaks, as will be addressed further below.

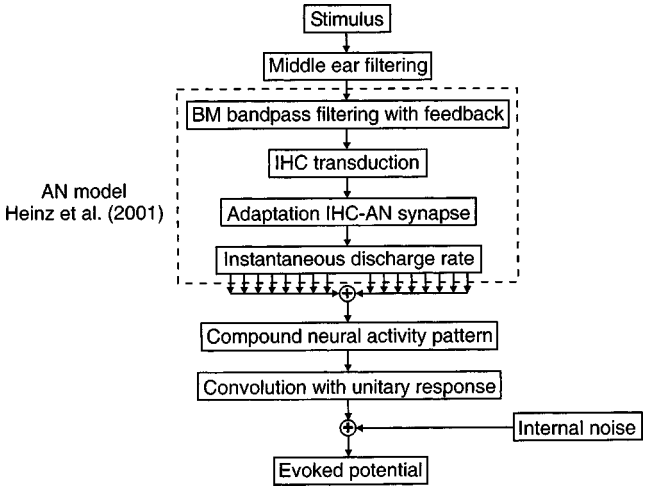


FIG. 1. Structure of the model for the generation of brainstem and frequency following responses. After middle-ear filtering, the stimuli are processed through the auditory-nerve model by Heinz *et al.* (2001). The instantaneous discharge rate functions are then summed across frequency. This summed activity pattern is convolved with the unitary response function shown in Fig. 2. Finally, internal noise is added that reflects background neural noise activity. Details are described in the text.

B. The overall model structure

Figure 1 shows the block diagram of the overall model used in the present study for the simulation of ABR. The first stage roughly simulates the middle-ear transformation. The incoming stimulus is filtered by a second-order Butterworth bandpass filter with cutoff frequencies of 0.3 and 7 kHz, respectively. It follows the processing through the computational AN model described below. The next stage in the model calculates the summed neural activity pattern, R_{AN}, at this level of processing, by adding up all discharge rate functions across CF. This pattern is then convolved with the unitary response waveform described below. Internal noise is added to the pattern to limit the resolution within the model. Pink noise was used since it was shown that the spectral energy density of the average ''background'' noise decreases at a rate of about 6 dB per octave (e.g., Granzow *et al.*, 2001). The internal noise may be considered as reflecting neural noise activity. The variance of the noise was determined once in advance to predict click-ABR data, and was held constant for any other stimulus, for a given number of averages. The output of the model represents the simulated evoked response pattern for the considered stimulus.

1. The auditory-nerve model

The computational AN model developed by Heinz *et al.* (2001) was used—without variation—in the present study to calculate the instantaneous discharge rate functions. The model is a modification of the physiologically based AN model by Zhang *et al.* (2001), which was developed for the cat and is itself an extension of the original Carney (1993) model. As indicated in Fig. 1, the model consists of nonlinear BM filtering, inner hair-cell (IHC) transduction, adaptation at the IHC-AN synapse, and generation of the instantaneous discharge rate as a function of CF. A detailed description of

the model and its implementation can be found in Heinz *et al.* (2001). In the following, some of the main characteristics are summarized.

The input stage is a filter bank that simulates the mechanical tuning of the basilar membrane. The model uses a *human* cochlear map according to Greenwood (1990), and the auditory filter bandwidths have been matched to humans based on psychophysical estimates of auditory filters (Glasberg and Moore, 1990). The parameters of these filters vary continuously as a function of stimulus level via a feedback mechanism, simulating the compressive nonlinearity associated with the mechanics of the basilar membrane. Level-dependent gain (compression), bandwidth, and phase properties are implemented with a control path that varied the gain and bandwidth of tuning in the signal-path filter. When the amplitude of the input is low, the filter is relatively sharply tuned. As the amplitude of the input signal increases, the bandwidth of the filter increases. The properties required of the feedback signal are similar to the response properties of outer hair cells (OHCs): A compressive magnitude response occurs near the characteristic frequency (CF), with the maximum compression occurring at the CF and an essentially linear response well away from the CF. The compressive response begins to occur at 20 dB SPL and is fully compressive by 40 dB SPL, consistent with physiological data (Ruggero *et al.*, 1997). The compression in the model begins to decrease above 80 dB SPL, but filter responses are still slightly compressive up to 120 dB SPL. The amount of compression (or cochlear-amplifier gain) in the model is largest for high CFs and decreases towards lower frequencies, consistent with data from basal and apical turns of the chinchilla cochlea (Ruggero *et al.*, 1997; Cooper and Rhode, 1997).

The time-varying AN discharge rate, $r_{i,\mathrm{AN}}(t)$, is calculated by passing the output of the signal-path filter through an asymmetric saturating nonlinearity, a low-pass filter, and a synapse model. The saturating nonlinearity and low-pass filter produce response properties associated with inner hair cell (IHC) transduction, whereas the synapse model includes adaptation effects such as the extended dynamic range at onset relative to the steady-state response. The model was specifically designed to describe the time-varying discharge rate of the AN fiber, $r_{i,\mathrm{AN}}(t)$, rather than producing discharge times which is more computationally extensive.

For the simulations of the present study, a set of 500 model CFs was used. The CFs ranged from 0.1 to 10 kHz, and were spaced according to a human cochlear map (Greenwood, 1990).

2. Discharge rate functions at brainstem level

Based on the combination of model-derived insights with experimental data from lesion studies in cats, and on comparison studies between cat and human ABR, it has been suggested that the three most robust peaks in the human ABR are predominantly generated by spiral ganglion cells in the auditory nerve (peak I), spherical cells in the anterior ventral part of cochlear nucleus (AVCN; peak III), and principal cells in the medial superior olive (MSO; peak V) which project to the inferior colliculus (IC; Fullerton *et al.*, 1987; Moore, 1987a, b; Melcher and Kiang, 1996).[1]1

The terminals of many AN fibers projecting to the AVCN contain large endbulbs which provide a secure synaptic connection with the large spherical bushy cells. Because of this secure synaptic connection, the firing behavior of primarylike units resemble AN fibers in most respects (Bourk, 1976; Pfeiffer, 1966; Rhode and Greenberg, 1994). Bushy cell axons terminate on cells in the MSO (Cant and Casseday, 1986; Osen, 1969; Warr, 1966, 1982). The (monaural) responses of cells in the MSO were shown to be similar to those seen in the AN: threshold tuning curves are V-shaped, rate-level functions are monotonic and saturating, and post-stimulus time histograms are primarylike in shape (Goldberg, 1975; Yin and Chan, 1990). Thus, phase locking by primary fibers is reliably followed by these more central centers. The information that converges in the MSO is then transmitted to the IC.

Based on these results, it is assumed in the present study that the instantaneous discharge rate functions are the *same* for the different cell populations considered, such that $R_{\mathrm{MSO}}(t,s)=R_{\mathrm{AVCN}}(t,s)=R_{\mathrm{AN}}(t,s)=:R(t,s)$. It follows from Eq. (8) that

$$\mathrm{ABR}(t,\overline{x_1},\overline{x_2},s)=R(t,s)*\sum_{p=1}^{P} u_p(t,\overline{x_1},\overline{x_2}). \tag{9}$$

Thus, the differences between the contributions from the different populations to the scalp potential are assumed to be reflected in the shape of the corresponding individual unitary responses u_p.

3. The unitary response function

In order to predict the ABR, specific assumptions about the individual unitary response functions are needed. While the contribution of a single neuron to ''gross'' responses has been measured (in cat) for the AN (Kiang *et al.*, 1976; Wang, 1979), no unitary responses have so far been obtained experimentally at neural centers higher than the AN. Thus, any specific assumptions about the waveforms of the components of the unitary response, chosen in order to simulate human ABR, would remain somewhat arbitrary. Instead, a different modeling strategy was used here, described in the following. Figure 2 (solid curve) shows the overall unitary response used in the present study. The function was calculated by *deconvolution* of the mean experimental click ABR data at 60 dB SL (corresponding to a peak equivalent sound pressure level of 106 dB) with the summed neural activity pattern for the click, generated by the AN model.[2] Thus, the stimulus-dependent neural excitation function in the brainstem is assumed to be approximated by the single function $R=R_{\mathrm{AN}}(t,s)$, as described in the previous section. The average click data, taken from Dau *et al.* (2000), are indicated by the dashed curve. The data show the typical pattern with clear waves I, III, and V at latencies that correspond well to a large body of literature data.

The obtained unitary response shows some characteristics that are consistent with results from other studies. The

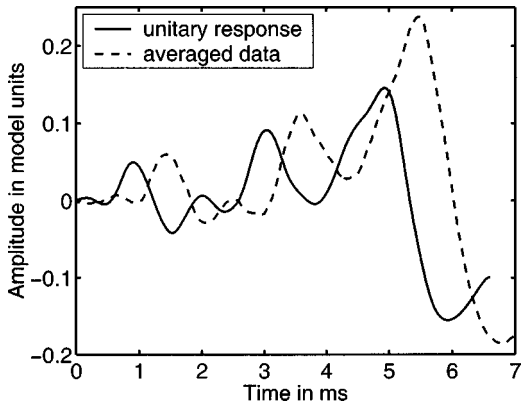


FIG. 2. The solid curve represents the unitary response function used in the present study. It was calculated by deconvolution of the mean experimental click ABR data for 60 dB SL (dashed curve, taken from Dau *et al.*, 2000) with the summed neural activity pattern for the click, generated by the AN model. In the present study, the same unitary response function was used for all stimuli at all levels.

biphasic shape of the first component (with its positive peak at about 1 ms) corresponds qualitatively to the shape of the (cat) AN unit responses measured by Wang (1979). Also, derived responses obtained from compound action potentials (CAPs) by high-pass masking (Elberling, 1974; Eggermont, 1976) showed a biphasic shape that resembles that of the first component of the unitary response in Fig. 2. The second main component of the function (with a peak at about 3 ms) reflects essentially a monophasic response. This is consistent with Melcher's (1995) calculated contribution from AVCN cell populations to evoked potentials (in cat) using a physico-mathematical cell model. It is also consistent with the results of Stegemann *et al.* (1987): the auditory nerve bifurcates at the level of the CN which represents a change in the volume conductor's geometry and should therefore, theoretically, generate a monophasic far-field response. The third component has a biphasic shape again. In this case, no estimates from experimental or modeling studies are available as for the earlier responses. However, a biphasic shape of this component seems necessary in order to produce phase cancellation as observed in ABR (and also FFR).

Within the present study, the above unitary response function will be used for *any* input stimulus at *any* level, implying the assumption of linearity at this stage of processing. All nonlinearity in the model is assumed to be restricted to the processing of the stimulus-dependent rate functions in the auditory periphery. In the following it will be explored whether the model accounts for the intensity-dependent aspects of ABR and FFR in various stimulus conditions.

III. STIMULI

First, click- and chirp-evoked responses are considered and compared to the experimental results shown in Dau *et al.* (2000). The chirp stimulus represents the ''approximate'' chirp as defined in Dau *et al.* (2000). The chirp duration is 10.48 ms; its instantaneous frequency starts at 100 Hz and increases up to 10.4 kHz in a nonlinear fashion. The click has a duration of 80 μs. Second, long-duration stimuli comprising the chirp are considered, which allow the investigation of the transition between traditional ABR obtained with transient stimuli and ''sustained''responses such as FFRs obtained with tones. Both stimulation paradigms were taken from a recent experimental study by Junius *et al.* (2000). In one experiment, the stimulus consisted of a continuous alternating sequence of chirps: each rising chirp was followed by the temporally reversed (falling) chirp. In another experiment, a single rising chirp was temporally and spectrally embedded in a long-duration stimulus comprised of a 30-ms 320-Hz tone, a 5.2-ms 320-to-8000-Hz chirp, and a 30-ms 8-kHz tone. This stimulus is referred to as the ''combination stimulus'' in the following. In the two experiments, the transitions between stimulus components were continuous. Third, FFRs obtained with 300-Hz tones and a 200-to-1200-Hz tonal sweep are considered and compared to the experimental data of Junius *et al.* (2000).

In most cases, the simulated patterns for all of the following stimulus configurations are shown in 10-dB steps in the range from 40 and 100 dB SPL. All cited experimental comparison data were obtained with electrodes at vertex and ipsilateral mastoid, and a ground electrode at the forehead. In some of the simulations, two independent waveforms of the internal noise are superimposed at each stimulation level. This was done in those stimulus conditions where two buffers were also shown in the corresponding experimental data to indicate response replicability (such as in Dau *et al.*, 2000).

IV. MODELING RESULTS

A. Click- and chirp-evoked ABR

Figure 3 shows the simulated patterns for click stimulation (left panel) in comparison to experimental click data for one individual subject (right panel), taken from Dau *et al.* (2000). In contrast to the Dau *et al.* study, the responses are indicated in dB SPL instead of dB sensation level (SL). The upper trace in the two panels shows the stimulus. Some of the characteristics of the simulated response patterns agree well with the data: The peaks I, III, and V are distinct for the higher stimulation levels (86 and 96 dB SPL) while, for a level of 56 dB SPL, wave V is the only prominent peak. There is no stimulus evoked activity visible below this level. Within the model, this is a consequence of the addition of the internal noise. Figure 4 shows corresponding results for chirp stimulation. The left panel represents the simulations while the experimental data for the same subject as in Fig. 3 are shown in the right panel. Again, some of the main characteristics of the data are reflected in the simulations. First, at high levels no clear formation of the three peaks I, III, and V can be observed. Instead, the wave components are temporally more dispersed than for the click. Second, wave-V amplitude becomes more distinct at lower levels and remains clearly detectable down to 43 dB SPL.

To illustrate the effects of neural synchrony across frequency due to cochlear processing, Fig. 5 shows the simulated activity at the output of the AN model. The activities for click (left panel) and chirp stimulation (right panel) are shown for ten frequency channels in the range from 0.1 to 10 kHz, together with the summed activity across channels in

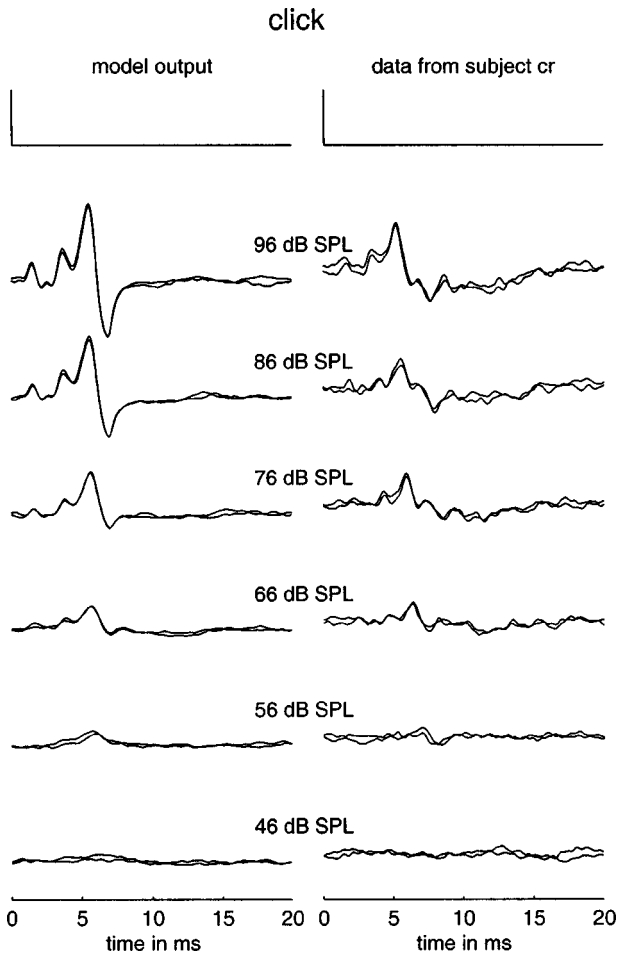


FIG. 3. Simulated responses for click stimulation (left panel) and experimental click-evoked responses for one individual subject (right panel), taken from Dau *et al.* (2000) (their Fig. 2, p. 1534). The stimulation level was in the range from 46 to 96 dB SPL, in steps of 10 dB. As for the data, the simulation shows two independent realizations for each level condition. The upper trace shows the stimulus.

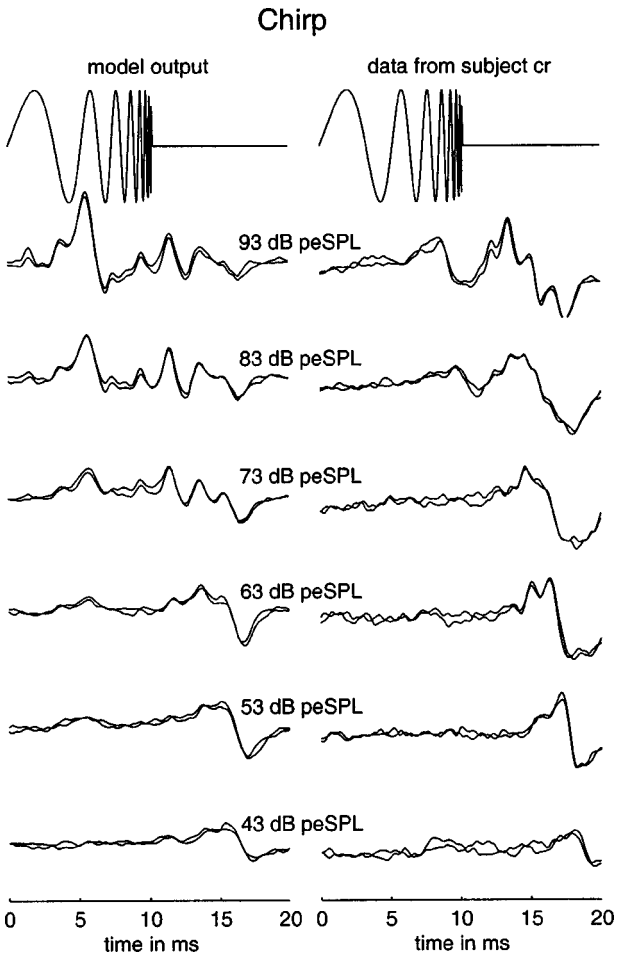


FIG. 4. Same as in Fig. 3 but for the chirp stimulus. In this case, the level ranged from 43 to 93 dB SPL. The data were also taken from Dau *et al.* (2000).

the top trace of each panel. The two upper panels show the results for low-level stimulation (50 dB SPL). For this level, neural activity is highly synchronized across frequency for the chirp, while it is temporally dispersed for the click due to the travel-time differences on the BM, particularly at the lower CFs. The situation changes for high-level stimulation (90 dB SPL) as shown in the lower panels of the figure. While neural activity is still desynchronized for the click in the low-frequency channels, it is well synchronized in the mid-and high-frequency region. In contrast, neural activity becomes more complex for the chirp at this high level because of the basal spread of excitation of the early low-frequency signal components in the chirp. This superposes with the later activity from the mid and high frequencies in the chirp, resulting in the more complex summed activity pattern.

While these main features are covered well by the model, at least qualitatively, there are several quantitative discrepancies between simulations and data. This may be illustrated best for the click results. Only the behavior of wave V is considered. The left panel of Fig. 6 shows wave-V amplitude as a function of the stimulation level. The open symbols indicate the average of the experimental data, representing the average of the values for the nine subjects from Dau *et al.* (2000), and the filled symbols show the results for the simulated responses. The simulations show a smaller increase with level at low levels and a steeper increase at medium levels than the experimental data. The values are similar at very low and high levels. The right panel of Fig. 6 shows the results for wave-V latency. The predicted latency change with level is much smaller (0.26 ms for a level change of 50 dB) in the simulations than in the experimental data (2.28 ms). Possible explanations for these discrepancies will be discussed in Sec. V B.

B. Responses to long-duration stimuli comprising the chirp

In the following, responses to the up–down chirp series and the combination stimulus are investigated, as defined in Sec. II. These stimuli illuminate the relation between ''classical'' ABR evoked by transient stimuli on the one hand, and ''sustained'' responses evoked by long-duration stimuli such as the FFR, on the other hand.

The left panel of Fig. 7 shows the simulated response patterns for the up–down chirp series. For direct comparison, the right panel of the figure shows corresponding experimental results for one representative subject, taken from Junius *et al.* (2000). While there are again obvious differences be-

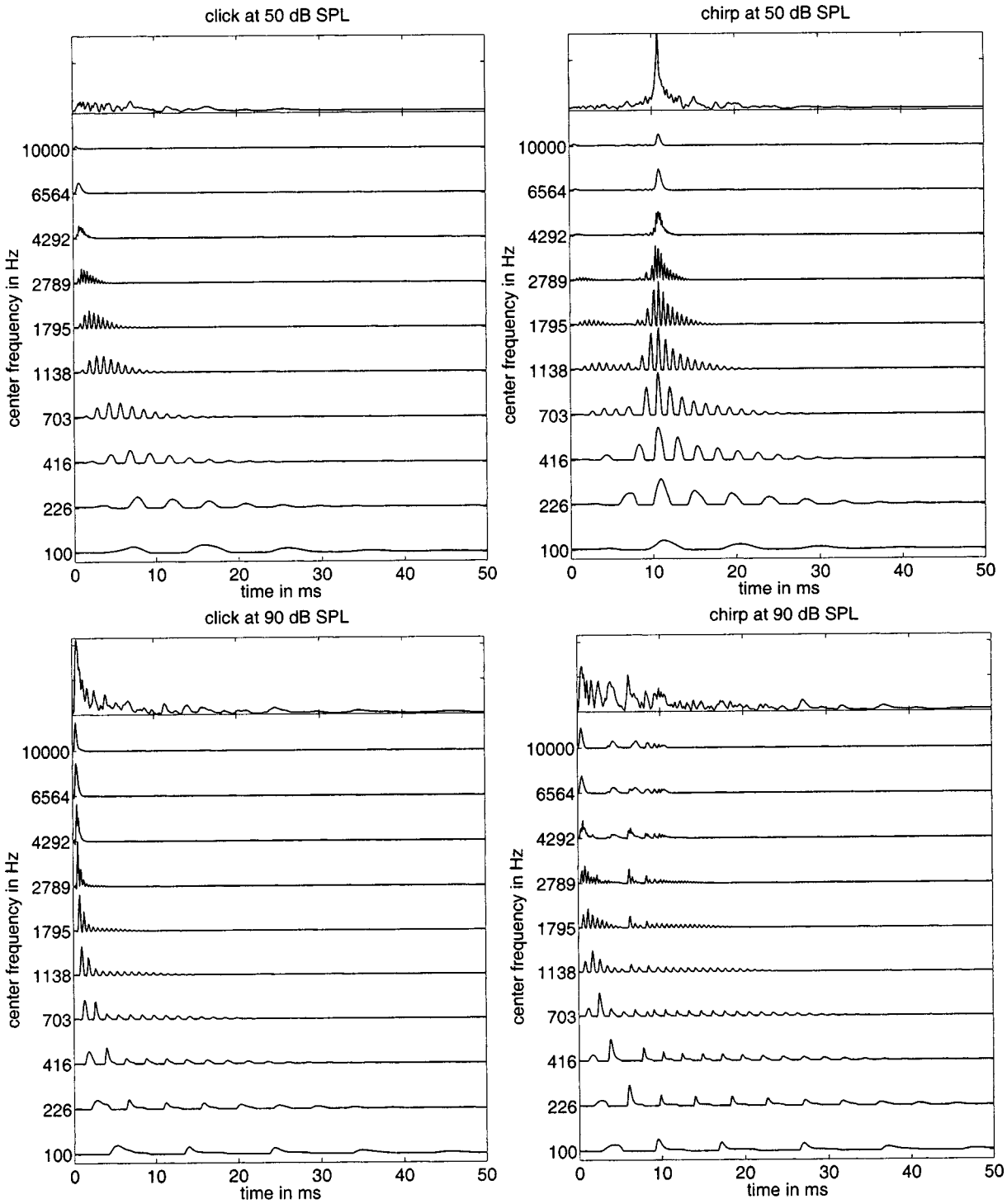


FIG. 5. Auditory-nerve rate functions in response to click (left panels) and chirp stimulation (right panels). Two level conditions are shown. The upper panels show activity for low-level stimulation of 50 dB SPL. The lower panels show corresponding activity for high-level stimulation of 90 dB SPL. The amplitude scale is the same for click and chirp but differs for the two levels.

tween data and model predictions concerning the details of the patterns and the exact shape of the response peaks, the key observations from the data are clearly also reflected in the simulations. First, for the lower stimulation levels, the peaks in the patterns reflect responses to the rising chirp sequences in the stimulus while the response to the falling chirp is smaller in amplitude. The amplitude and (relative) latencies of the peaks correspond to those obtained with the single chirp from Fig. 4. Second, the response peaks become somewhat less distinctive at high levels, a characteristic that has also been observed for single chirp presentation (see Fig. 4). These findings further demonstrate that wave V does not necessarily reflect an event evoked by stimulus on- or offset, or other discontinuities in the stimulus waveform. It is the amount of neural synchrony after cochlear processing that determines the shape of the response pattern.

Figure 8 (left panel) shows the results for the combination stimulus (that comprises the rising chirp embedded in

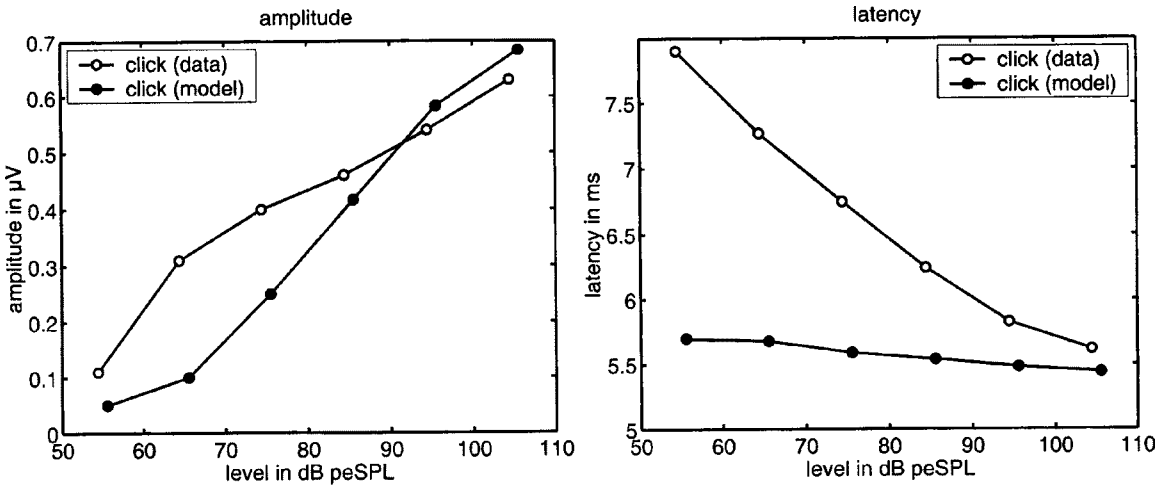


FIG. 6. The left panel shows wave-V amplitude of the simulated (filled symbols) and the measured data (open symbols, mean of nine subjects) as a function of the stimulation level. The right panel shows the corresponding values for wave-V latency.

two tones). The model responses are shown in the left panel while corresponding experimental data for one representative subject are shown in the right panel, taken from Junius *et al.* (2000). As in the previous stimulus condition, data and prediction differ in the details of the response patterns such as the width of the response peaks. Also, the model shows a stronger onset response to the stimulus than do the experimental data. The main response characteristics in the data, however, are reflected in the simulations. First, at low stimulation levels up to 60 dB SPL, data and predictions show a (wave-V) peak in response to the (rising) chirp sections, embedded in the stimulus, whereas no responses were evoked by the low-frequency tone at the beginning and the high-frequency tone at the end of the stimulus sequence. Second, the response pattern changes considerably at higher levels. For levels of 70 dB SPL and higher, the low-frequency tone generates a periodic response that corresponds to the ''classical'' FFR patterns to tones. The chirp-evoked response amplitude is considerably lower for these higher levels than for the lower levels, consistent with the observation from the previous experiments. Third, as expected, the high-frequency (8-kHz) tone does not produce any periodic response at any stimulation level since, at 8 kHz, the ability of the system to phase-lock to the fine structure is strongly reduced.

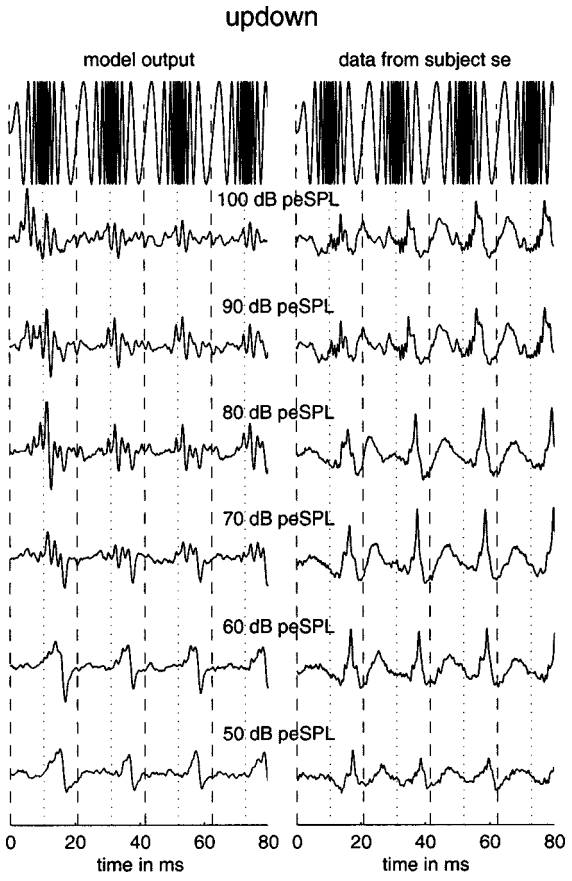


FIG. 7. Left panel: Simulated responses to the up–down chirp series. The transitions between the stimulus components were continuous. Right panel: Measured responses to the same stimuli for one representative subject, taken from Junius *et al.* (2000). In this case, only one realization (buffer) per level condition is plotted for simulation and data. The dashed and dotted vertical lines indicate the time points of the stimulus sequence where the up-chirps starts and ends, respectively.

These modeling results suggest that the same mechanisms are responsible for the generation of ABR wave V and FFR. This is consistent with experimental studies where a latency analysis suggested a common source for these two evoked responses (e.g., Smith *et al.*, 1975). To illustrate effects of temporal dispersion on FFR generation within the framework of the present model, the contributions of neural activity from different frequency regions are investigated in the following.

C. Frequency following responses

Figure 9 shows simulated responses to a 50-ms 300-Hz tone (with 10-ms Hanning windows at the on- and offset). The left panel represents simulations obtained by including channels tuned only to frequencies in the range from 0.1–1.4 kHz. The same frequency spacing was used as in the previous simulations. The response patterns do not show a clear periodic response to the tone, even for the highest stimulation levels. The right panel of Fig. 9 shows simulations obtained by including channels limited to CFs in the range

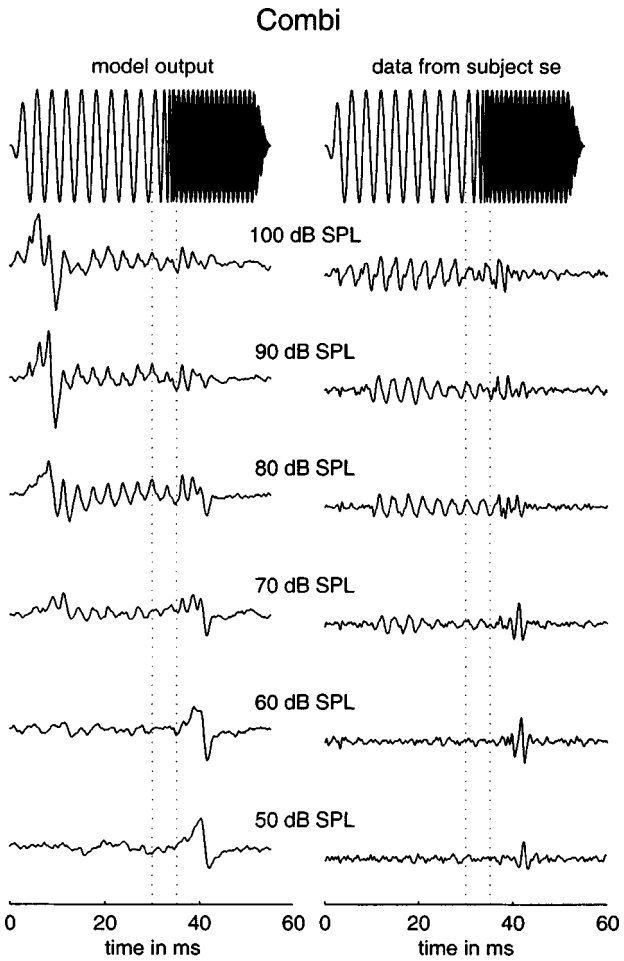


FIG. 8. Left panel: Simulated responses to the combination stimulus consisting of a 30-ms 320-Hz tone, a 5.2-ms 320–8000-Hz chirp, and a 30-ms 8000-Hz tone. As in the previous condition, the transitions between the stimulus components were continuous. Right panel: Measured responses to the same stimuli for the same representative subject as in Fig. 7, taken from Junius *et al.* (2000). The dotted vertical lines indicate the beginning and the end of the embedded chirp in the stimulus.

from 1.5 to 10 kHz. The number of channels used in these simulations was the same as for the simulations at the low frequencies. A clear periodic response can be observed at high levels, while there is no response at low levels. The responses are similar to those obtained in the original analysis with channels in the range from 0.1–10 kHz (not shown).

These results demonstrate that, within the modeling framework presented here, FFR to low-frequency tones mainly arise from the synchronized neural activities generated in the mid- and high-frequency channels. At high stimulation levels these regions are excited due to upward spread of excitation. Activity is well synchronized in these regions, while it is not in the low-frequency region due to larger travel-time differences. This is illustrated in Fig. 10 for a 90 dB SPL tone of 300 Hz. The left panel shows simulated outputs of the AN model for channels in the low-frequency region (0.1–1.5 kHz). The right panel shows corresponding output activities for channels in the high-frequency region (1.5–10 kHz). The top curve in each panel represents the overall activity summed across the frequency channels. For the low-frequency region (left), the summed pattern is smaller in amplitude and less periodic than for the high-frequency region (right). The reason for this is that the AN model incorporates neural properties of AN fibers that have been first described by Kiang and Moxon (1974): Tuning curves for AN fibers show significant low-frequency tails such that even fibers with high CFs respond to tones throughout a wide range of low and mid frequencies. For low-frequency stimuli high-CF units all tend to respond in phase and low-frequency tones at high intensities stimulate almost the entire cochlea in phase.

Figure 11 shows simulated (left panel) and measured (right panel) response patterns obtained for the linearly rising chirp at 75 dB SPL, as used in Junius *et al.* (2000). As for the tone stimulation from the previous experiments, the high-level stimulation leads to a predicted response pattern whose spectro-temporal properties correspond to those of the input stimulus. The trajectory of the frequency change of the response roughly follows that of the stimulus. This also corresponds to experimental results obtained by Krishnan and Parkinson (2000) who investigated FFR to similar tonal sweeps. The interpretation of the modeling results is the same as that of the ''classical'' FFR to tones: The response to tonal sweeps represents synchronized neural activity stemming mainly from fibers tuned to mid- and high-frequency regions while neural activity at low frequencies does not contribute effectively to the overall response due to phase cancellation as a consequence of cochlear dispersion.

V. DISCUSSION

A. Interpretation of ABR and FFR

The present study investigated the role of (nonlinear) cochlear processing for the formation of ABR and FFR. It was shown that a realistic simulation of the level-dependent signal processing in the cochlea is essential for the interpretation of ABR and FFR. First, the model reproduces the experimental findings for ABR to clicks and chirps that, at low stimulation levels, the chirp leads to a larger response than the click due to better synchronization, while at high levels, the chirp response is smaller than that of the click due to effects of cochlear spread of excitation when the earlier low-frequency energy in the chirp stimulates basal regions and produces a response. Second, within the model, FFR to low-frequency stimuli represent synchronized neural activity mainly from mid- and high-frequency channels. Due to the high velocity of the traveling wave, basal places have close phase relationships and, thus, the synchronized hair-cell activity in the basal region of the cochlear partition essentially initiates the FFR. This is consistent with the results from the study by Janssen *et al.* (1991) who investigated the mechanisms for FFR generation by studying responses to a 500-Hz tone (at 96 dB SPL) in the presence of high-pass masking noise with cutoff frequencies in the range from 0.5 to 4 kHz. FFR amplitude was found to be reduced for noise cutoff frequencies in the mid-frequency range, and was absent for cutoff frequencies of 0.75 and 0.5 kHz.

Janssen *et al.* (1991) also presented model predictions where neural synchrony after peripheral processing was calculated using Møller's transfer function of the middle ear (1963), de Boer's linear basilar-membrane model (1980), a

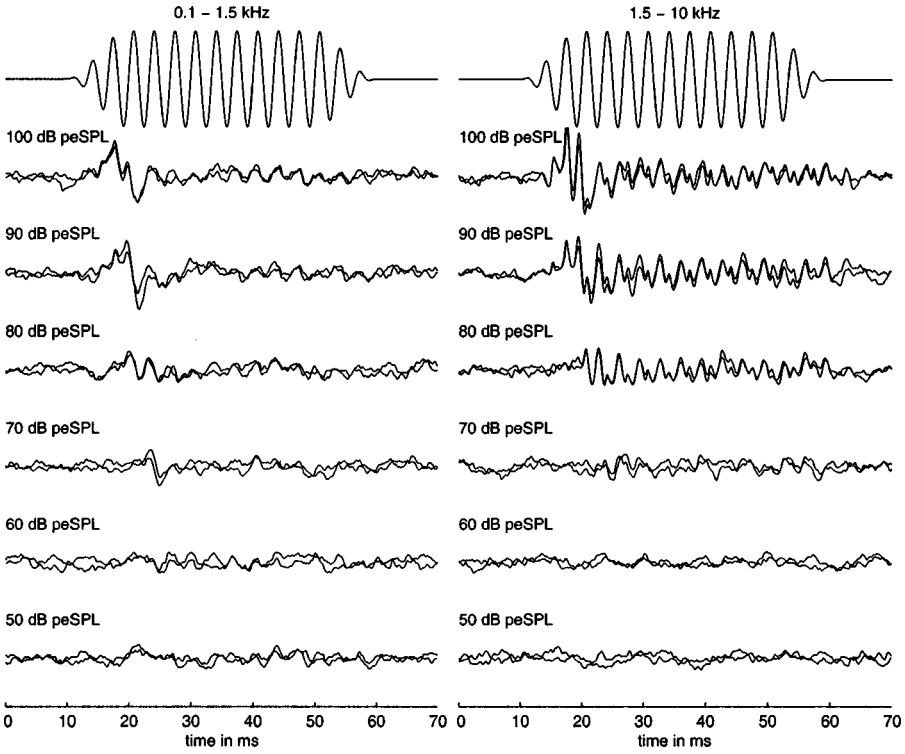


FIG. 9. Simulated responses to a 300-Hz tone. Left panel: Only CFs in the low-frequency range from 0.1 to 1.5 kHz were considered in the analysis. Right panel: Only CFs from 1.5 to 10 kHz were considered. The spectral spacing of the channels was the same as in all other simulations of the present study.

neural transduction process in cochlear hair cells as described in Russell and Cody (1985), and a neural spike generator according to Jones *et al.* (1985). While the modeling framework and some of the key results of Janssen's and the present study are similar, the interpretation of the results was clearly different in some aspects. Janssen *et al.* argued that FFR generation can be thought of as the superposition of transient ABR to each cycle in the stimulus. They compared FFR to a 16-ms 500-Hz Gaussian tone pulse, consisting of eight periods, with the linear superposition of the ABR to isolated 2-ms cycles of the tone pulse presented separately, and found similar wave patterns for these two conditions. Thus, FFR was interpreted as a composite of ABR elicited by multiple slope stimuli. While such a linear behavior may be approximately observed at very high levels (such as the 95 dB SPL in the Janssen *et al.* study) where the BM acts

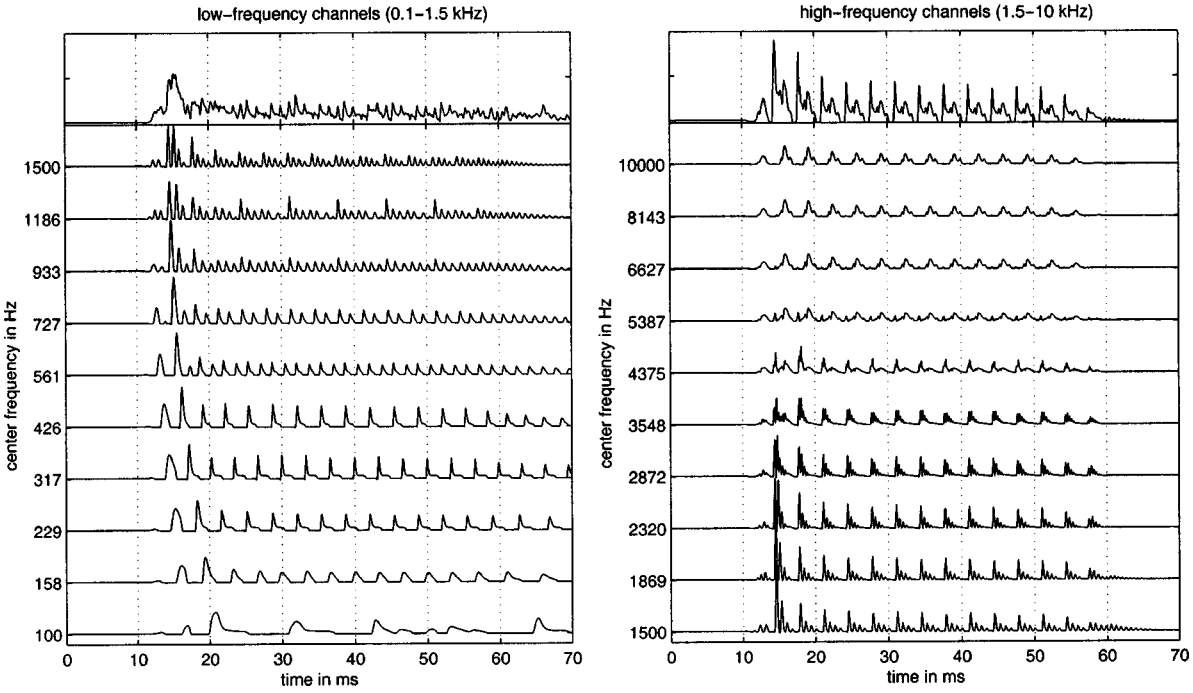


FIG. 10. Simulated auditory-nerve responses to a 300-Hz tone at 90 dB SPL, obtained with the Heinz *et al.* model (2001). The left panel shows the outputs for the low-frequency channels in the range from 0.1 to 1.5 kHz. The right panel shows corresponding responses in the frequency region from 1.5 to 10 kHz. On the top of each panel, the summed activity across all frequency channels is shown.

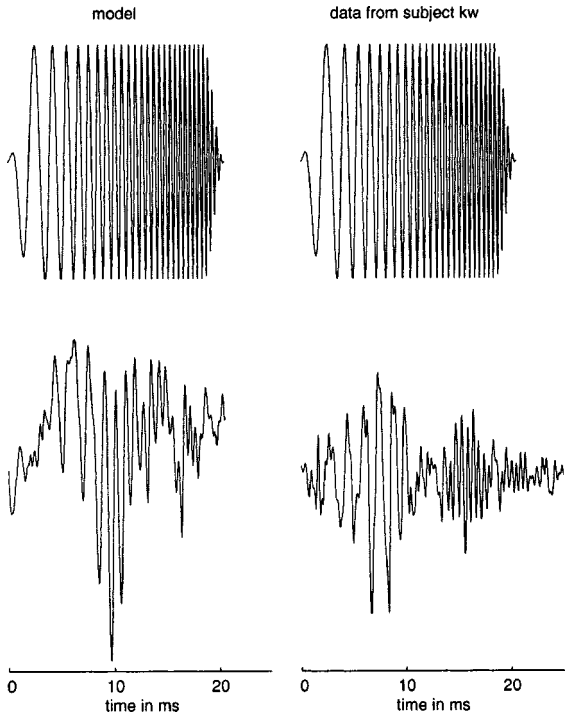


FIG. 11. Simulated responses (left panel) to a 20-ms chirp whose instantaneous frequency rises linearly from 300 to 800 Hz. The corresponding FFR data (right panel) for one representative subject are taken from Junius *et al.* (2000). The stimulation level was 75 dB SPL.

roughly linearly, this cannot be expected at levels in the BM compression range (of about 40–80 dB SPL). Also, the decomposition of the stimulus into separate parts introduces spectral splatter into the signal which makes the interpretation of the results complicated (even though Janssen *et al.* reduced the amount of spectral splatter by smoothing the on- and offsets of each cycle). Within the modeling framework of the present study, FFR generation directly results from the frequency-selective properties of the peripheral processing stages that change with level in connection with the level-dependent cochlear compression. Since the model combines level-dependent signal processing with the concept of the unitary response, it may be considered as the more general approach for describing FFR generation.

The interpretation of the FFR might have important consequences for clinical applications. If the FFR essentially reflects neural activity stemming from mid- and high-frequency channels, they cannot represent a neural correlate of low-frequency hearing, as has been often suggested. Indeed, hearing threshold at low frequencies does not seem to be relevant for FFR generation. Instead, the modeling results suggest that the *absence* of FFR to low-frequency tones could only be explained if high-frequency regions on the BM do not contribute to the response such as, e.g., in the case of a severe high-frequency hearing loss or dead region. Corresponding experiments are currently under way.

B. Limitations of the model

While the key observations in the data were reflected in the simulations some clear discrepancies between data and simulations were found in the details of the response patterns. One difference between experimental data and model predictions is the shape of the peaks obtained in some of the ABR patterns, particularly for the chirp at high stimulation levels. One possible explanation could be that the chirp was generated from de Boer's (1980) cochlea model while the filtering stage in the model is based on the gammatone filterbank. Thus, the time course of the chirp will not exactly match the inverse delay-line characteristic of the filterbank. However, additional simulations, obtained with modified chirps compensated for the frequency delays within the gammatone-filter based modeling framework, did *not* lead to a markedly sharper response in the predictions (not shown). Thus, it seems that other aspects of cochlear processing and/or later processing in the brainstem (see below) that are not included in the model may be responsible for the observed discrepancies.

Another apparent problem with the simulations is the model's failure to correctly describe the level dependence of the response latency (see Fig. 6, right panel). While the experimental data generally show a latency shift of wave V of about 2.3 ms for a level change of 50 dB, the model predicts a latency shift of less than 0.3 ms for the same level change. The latencies of the ABR can principally be separated into mechanical and neural components. The mechanical component is due to mechanical BM travel time, and varies with intensity and frequency in an orderly manner, while the remaining neural component is generally assumed to be independent of both intensity and frequency (e.g., Neely *et al.*, 1988). This is also the case in the framework of the ABR model of the present study: The neural latencies within the model were assumed to be constant while the mechanical latencies change systematically with intensity and frequency. The mechanical latencies agree reasonably well with the values obtained in physiological BM and AN impulse response data (e.g., de Boer and Nuttall, 1997, 2000). At present, it seems unclear, at least to this author, what mechanisms are responsible for the relatively large latencies obtained in the low-level ABR data.

A principal limitation of the present model might be that the frequency-selective processing is realized functionally as a bank of bandpass filters. Such a model ignores cochlear hydrodynamics which form the biophysical basis for traveling-wave generation. While the filterbank model used in the present study may effectively describe correctly many of the important characteristics of the peripheral auditory signal processing, such as level-dependent frequency tuning and other aspects associated with cochlear nonlinearity (Heinz *et al.*, 2001), other aspects may not be reflected. For example, the model does not describe the phase response of the auditory filters correctly (Shera, 2001a, b) since the dispersive properties of wave propagation are not accounted for. However, it is not clear in advance if deviations from the exact filter phase response will necessarily be critical for potential generation since it is the number of units activated synchronously *across* channels and not large within-channel activities that mainly determine the potential amplitude. As long as the filter bandwidth is estimated realistically by the filterbank approach, the corresponding build-up and decay time constants, and thus the resulting travel-time differences

across frequency, may be described appropriately (Goldstein *et al.*, 1971). However, these aspects need to be clarified in future modeling studies.

All aspects mentioned so far considered limitations due to cochlear processing. Another possible limitation within the framework of the model may be, of course, the specific choice of the unitary response. This was obtained by deconvolving the click ABR data (at a high stimulation level) with the calculated neural excitation function at the output of the AN. This linear system's approach might be unrealistic and the shape of the unitary response function might change with level and CF. The width of the unitary response (components) essentially represents the temporal window over which the rate functions are integrated. The wider the window is, the smoother the representation after the convolution. However, as was shown by Elberling (1976), the exact *shape* of the unitary response does not seem to play a critical role. Elberling calculated the ''excitation function'' by deconvolving the measured CAP with some assumed function for the unitary response, and demonstrated that the deconvolution is sensitive in only a minor degree to differences in the waveform of the unitary response. He tested different functions for the unitary response which had been suggested by Elberling (1974), Teas *et al.* (1962), de Boer (1975) and Biondi *et al.* (1975) and differed somewhat in shape and the amount of ''after-oscillations'' (Elberling, 1976). Similar observations were made in the present study; additional simulations (not shown) were run where several hypothetical unitary response functions for the different components were tested. While the (relative) amplitude and width of the different unitary response components were critical for the resulting potential waveform, the exact shape of the components did not strongly influence the results.

Another limitation of the model is that the assumed signal processing in the brainstem above the level of the AN may be oversimplified. Specifically, the assumption that the rate functions in the main contributing neural sites in the brainstem are essentially the same as in the AN may be too strong. For example, it has been shown recently that neural synchronization in the AVCN can be *enhanced* compared with AN fibers due to the convergence of inputs from two or more AN fibers onto an AVCN cell and postsynaptic cells that require coincident input spikes before they fire (Joris *et al.*, 1994). Also, even though the human ABR may be largely generated by brainstem cells in the spherical cell pathway (Melcher and Kiang, 1996), there is probably also some contribution from other cell types such as globular and multipolar cells. There is still some controversy about the exact generating sites of the ABR peaks beyond wave I. For example, the wave IV/V complex in the ABR has also been associated with the lateral lemniscus, and not with the MSO (e.g., Özdamar, 1990; Ponton *et al.*, 1996), which seems also more consistent with results of a spatio-temporal dipole model on ABR generation (Scherg and von Cramon, 1985).

The whole modeling approach should therefore be considered as a rough approximation of the real neural mechanisms involved in the generation of brainstem potentials. The assumption that all nonlinearity is restricted to BM and AN processing and that the remaining processing is essentially linear is certainly very strong given the high complexity of the neural processing at the brainstem centers. However, it appears that this description may represent a simple and very effective approximation. The model represents a powerful framework that can be tested for any stimulus of interest and for any simulated amount of hearing loss. The unitary response could be calculated for any electrode configuration and, if desired, for any subject individually. The deconvolution technique has already been successfully applied to middle-latency responses (Gutschalk *et al.*, 1999; Rupp *et al.*, 2002).

C. Future investigations

So far, transient stimuli as well as tones and tonal sweeps were considered. A following step would be to consider ''steady-state'' responses (SSR) obtained with temporally fluctuating stimuli such as complex tones or amplitude-modulated tones or noises. Such responses are assumed to be generated by units in the auditory brainstem and in the primary auditory cortex (e.g., Kuwada *et al.*, 1986). Thus the corresponding unitary response would have to be extended by a middle-latency component. To what extent such a convolution approach can successfully be applied to middle-latency responses (MLR), to transients as well as to amplitude modulation following responses (AMFR), remains to be tested. Indeed, MEG recordings obtained with chirps (as those used in the present study) and clicks showed very similar differences in response amplitude and latency for the $N_{19}P_{30}$ complex as observed for wave V at brainstem level (Rupp *et al.*, 2002), demonstrating that temporal dispersion due to cochlear processing is reflected in neural activity up to at least the level of the primary auditory cortex (Rupp *et al.*, 2002).

VI. SUMMARY AND CONCLUSIONS

A model for ABR and FFR generation was presented which is based on the concept that evoked potentials recorded at remote electrodes can be given by convolution of an elementary unit waveform (unitary response) with the instantaneous discharge rate function for the corresponding unit (Goldstein and Kiang, 1958). In the present study, the nonlinear computational AN model developed by Heinz *et al.* (2001) was used to calculate the instantaneous discharge rate functions, and the summed excitation across frequency was convolved with a unitary response which was assumed to reflect contributions from different cell populations within the auditory brainstem.

Predicted potential patterns for clicks, chirps, long-duration stimuli comprising the chirp, tones, and slowly varying tonal sweeps were compared with corresponding experimental data. Some of the main characteristics and key observations from the data were reflected in the simulations, such as (*i*) level-dependent differences in the ABR to clicks and chirps, (*ii*) level-dependent FFR patterns to tones and tonal sweeps, and (*iii*) level-dependent transitions in the patterns for composite stimuli such as the up–down chirp series and the combination stimulus the latter of which comprised tone sequences as well as the rising chirp.

The results demonstrate the importance of considering the effects of the BM traveling wave and AN processing for the formation of ABR and FFR. In particular, the results support the hypothesis that FFRs to low-frequency tones represent synchronized activity mainly stemming from basal high-frequency activity and not from activity of units tuned to frequencies around the signal frequency.

Several clear discrepancies between model predictions and experimental data were observed. The model only crudely explained the wave shape of the experimental data, and it fails to describe the correct latency behavior as a function of level. One possible reason for these differences could be that aspects of cochlear processing associated with its dispersive character are not accounted for accurately by the model. Another reason for the discrepancies might be that the assumed signal processing in the brainstem above the level of the AN was oversimplified. Further modeling efforts are needed to improve our understanding of the mechanisms underlying the generation ABR and FFR.

Nonetheless, the present model might serve as a useful tool since it can be applied to any stimulus configuration of interest. Also, the model may be applied to any form of simulated cochlear hearing loss in order to understand the effects of hearing impairment on evoked potential generation. Furthermore, it should be possible to extend the modeling framework to middle-latency responses and envelope following responses. Such an extension of the model is currently under investigation.

ACKNOWLEDGMENTS

I thank Oliver Wegner for numerous discussions and his help in preparing the figures. I also thank Michael Heinz, Laurel Carney, Helmut Riedel, Stephan Ewert, and John Culling for discussions and for critical reading of an earlier version of this paper. I would also like to thank the two anonymous reviewers for very helpful suggestions and constructive criticism of the earlier version of the paper. This study was supported by the Deutsche Forschungsgemeinschaft (DFG) and the Max Kade Foundation.

[1]Melcher and Kiang (1996) identified specific cellular generators of the click-evoked ABR in cats. From their lesion studies, they concluded that (1) the earliest extrema in the cat ABR are generated by spiral ganglion cells in the AN, (2) P2 is mainly generated by cochlear nucleus (CN) globular cells, (3) P3 is partly related by CN sperical cells and partly by cells receiving inputs from globular cells, and (4) P4 is predominantly generated by medial superior olive (MSO) principal cells. Thus, the ABR in cats mainly reflects cellular activity in *two parallel* pathways, one originating with globular cells and the other with spherical cells. Since the globular pathway is poorly represented in humans, Melcher and Kiang (1996) suggested that the human ABR is largely generated by brainstem cells in the *spherical* pathway. It has been suggested by Fullerton *et al.* (1987) that the first three peaks (1–3 in cats and I–III in humans) have similar generators, and that P4 for cat has the same generators as wave V in the human ABR. Thus, peak III in humans is argued to be mainly generated by the spherical bushy cells of the AVCN (and not by the globular bushy cells or the multipolar stellate cells of this nucleus). That peak V is predominantly generated by MSO cells is consistent with the observation that an abnormally small peak V can be found in humans with lesions of the lateral lemniscus (Voordecker *et al.*, 1988), and with the results of ABR abnormalities in patients with identified multiple sclerosis (Levine *et al.*, 1993).

[2]Only the time interval of the click response from −1 to 7 ms was considered in the deconvolution analysis, such that the first minimum after peak V was included but no later activity. Tikhonov regularization was applied (Tikhonov, 1963; Hansen, 1992) to achieve a stable and smooth solution for the inverse problem inherent in deconvolution. The extraction of an appropriate and objective regularization parameter was based on the generalized cross-correlation function (GCV). All calculations were done in MATLAB. The analysis tools for regularization problems, including the GCV function to extract the optimal parameter, were provided by Hansen (1998).

Biondi, E., Dacquino, G., and Grandiori, F. (**1975**). ''Compound action potentials and single acoustic nerve fibers activity generation: An equivalent neuron approach,'' Int. J. Bio-Med. Comput. **6**, 157.

Bourk, T. R. (**1976**). ''Electrical responses of neural units in the anteriorventral cochlear nucleus of the cat,'' Ph.D. thesis, Massachusetts Institute of Technology.

Buchwald, J. S., and Huang, C.-M. (**1975**). ''Far-field acoustic response: origins in the cat,'' Science **189**, 382–384.

Cant, N. B., and Casseday, J. H. (**1986**). ''Projections from the anteroventral cochlear nucleus to the lateral and medial superior olivary nuclei,'' J. Comp. Neurol. **247**, 457–476.

Carney, L. H. (**1993**). ''A model for the response of low-frequency auditory-nerve fibers in cat,'' J. Acoust. Soc. Am. **93**, 401–417.

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

Dau, T., Wegner, O., Mellert, V., and Kollmeier, B. (**2000**). ''Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion,'' J. Acoust. Soc. Am. **107**, 1530–1540.

de Boer, E. (**1975**). ''Synthetic whole-nerve action potentials for the cat,'' J. Acoust. Soc. Am. **58**, 1030–1045.

de Boer, E. (**1980**). ''Auditory physics. Physical principles in hearing theory I,'' Phys. Rep. **62**, 87–174.

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.

de Boer, E., and Nuttall, A. L. (**2000**). ''The mechanical waveform of the basilar membrane. III. Intensity effects,'' J. Acoust. Soc. Am. **107**, 1497–1507.

Delgutte, B. (**1990**). ''Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level,'' Hear. Res. **49**, 225–246.

Eggermont, J. J. (**1967**). ''Electrophysiology,'' in *Handbook of Sensory Physiology*, Volume V/3 edited by W. D. Keidel and W. D. Neff (Springer Verlag, New York), pp. 625–705.

Elberling, C. (**1974**). ''Simulation of cochlear action potentials recorded from the ear canal in man,'' in *Proc. Symposium on Electrocochleography*, edited by R. Ruben, C. Elberling, and J. Salomon (University Park, Baltimore).

Elberling, C. (**1976**). ''Modelling action potentials,'' Rev. Laryngol. **97**, 527–537.

Fullerton, B. C., Levine, R. A., Hosford-Dunn, H. L., and Kiang, N. Y. S. (**1987**). ''Comparison of cat and human brain-stem auditory evoked potentials,'' Electroencephalogr. Clin. Neurophysiol. **66**, 547–570.

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

Goldberg, J. M. (**1975**). ''Physiological studies of the auditory nuclei of the pons,'' in *Handbook of Sensory Physiology. Auditory System*, edited by W. D. Keideland and W. D. Neff (Springer, Berlin), Vol. V, part 2, pp. 109–144.

Goldstein, J. L., Baer, T., and Kiang, N. Y. S. (**1971**). ''A theoretical treatment of latency, group delay, and tuning characteristics for auditory-nerve responses to clicks and tones,'' in *Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore, MD).

Goldstein, M. H., and Kiang, N. Y. S. (**1958**). ''Synchrony of neural activity in electric responses evoked by transient acoustic stimuli,'' J. Acoust. Soc. Am. **30**, 107–114.

Granzow, M., Riedel, H., and Kollmeier, B. (**2001**). ''Single-sweep-based methods to improve the quality of auditory brain stem responses Part I: Optimized linear filtering,'' Z. Audiol. **40**, 32–44.

Greenwood, D. D. (**1990**). ''A cochlear frequency position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Gutschalk, A., Mase, R., Roth, R., Ille, N., Rupp, A., Hahnel, S., Picton, T. W., and Scherg, M. (**1999**). ''Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex,'' Clin. Neurophysiol. **110**, 856–868.

Hansen, P. C. H. (**1992**). *Rank-Deficient and Discrete Ill-posed Problems. Numerical Aspects of Linear Inversion* (SIAM, Philadelphia).
Hansen, P. C. H. (**1998**). ''Regularization tools. A Matlab package for analysis and solution of discrete ill-posed problems,'' http://www.imm.dtu.dk/pch.
Heinz, M. G. (**2000**). ''Quantifying the effects of the cochlear amplifier on temporal and average-rate information in the auditory nerve,'' Ph.D. thesis, Massachusetts Institute of Technology.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001**). ''Auditory-nerve model for predicting performance limits of normal and impaired listeners,'' ARLO **5**(3), 91–96.
Janssen, T., Steinhoff, H.-J., and Bøhnke, F. (**1991**). ''Zum Entstehungsmechanismus der Frequenfolgepotentiale,'' Otorhinolaryngol. Nova. **1**, 16–25.
Jewett, D. L. (**1970**). ''Volume-conducted potentials in response to auditory stimuli as detected by averaging in the cat,'' Electroencephalogr. Clin. Neurophysiol. **28**, 609–618.
Johnson, D. (**1978**). ''The relationship of post-stimulus time and interval histograms to the timing characteristics of spike trains,'' Biophys. J. **22**, 413–430.
Jones, K., Tubis, A., and Burns, E. M. (**1985**). ''On the extraction of the signal excitation function from a non-Poisson cochlear neural spike train,'' J. Acoust. Soc. Am. **78**, 90–94.
Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. T. (**1994**). ''Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency,'' J. Neurophysiol. **71**, 1022–1036.
Junius, D., Wegner, O., and Dau, T. (**2000**). ''Human auditory brainstem potentials using optimized stimuli to compensate for basilar-membrane dispersion,'' in *23rd Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, FL.
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., Moxon, E. C., and Kahn, A. R. (**1976**). ''The relationship of gross potentials from the cochlea to single unit activity in the auditory nerve,'' in *Electrocochleography*, edited by R. Ruben, C. Elberling, and J. Salomon (University Park, Baltimore).
Kimura, J., Mitsodome, A., Beck, D. O., Yamaha, T., and Dickins, Q. S. (**1984**). ''Stationary peaks from a moving source in far-field recording,'' Electroencephalogr. Clin. Neurophysiol. **58**, 351–361.
Krishnan, A., and Parkinson, J. (**2000**). ''Human frequency-following response: Representation of tonal sweeps,'' Audiol. Neuro-Otol. **5**, 312–321.
Kuwada, S., Batra, R., and Maher, V. L. (**1986**). ''Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones,'' Hear. Res. **21**, 179–192.
Levine, R. A., Gardner, J. C., Fullerton, B. C., Stufflebeam, S. M., Carlisle, E. W., Furst, M., Rosen, B. R. (**1993**). ''Effects of multiple sclerosis brainstem lesions on sound lateralization and brainstem auditory evoked potentials,'' Hear. Res. **68**, 73–88.
Marsh, J. T., Brown, W. S., and Smith, J. C. (**1975**). ''Far-field recorded frequency-following responses: correlates of low pitch auditory perception in humans,'' Electroencephalogr. Clin. Neurophysiol. **38**, 113–119.
Melcher, J. R. (**1995**). ''Contributions from AVCN cells to evoked potentials depend strongly on fiber diameter: Results from a physico-mathematical model,'' presented at the 18th Meeting of the Association for Research in Otolaryngology.
Melcher, J. R., and Kiang, N. Y. S. (**1996**). ''Generators of the brainstem auditory evoked potential in cat III: identified cell populations,'' Hear. Res. **93**, 52–71.
Moushegian, G., Rupert, A. L., and Stillman, R. D. (**1973**). ''Scalp-recorded early responses in man to frequencies in the speech range,'' Electroencephalogr. Clin. Neurophysiol. **35**, 665–667.
Møller, A. R. (**1963**). ''Transfer function of the middle ear,'' J. Acoust. Soc. Am. **35**, 1526–1543.
Moore, J. K. (**1987a**). ''The human auditory brain stem: A comparative view,'' Hear. Res. **29**, 1–32.
Moore, J. K. (**1987b**). ''The human auditory brain stem as a generator of auditory evoked potentials,'' Hear. Res. **29**, 33–43.
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.
Osen, K. K. (**1969**). ''The intrinsic organization of the cochlear nuclei in the cat,'' Acta Oto-Laryngol. **67**, 352–359.
Özdamar, Ö., and Delgado, R. E. (**1990**). ''Fiber tract model of auditory brainstem response generation using traveling dipoles,'' in *Auditory Evoked Magnetic Fields and Electric Potentials. Advances in Audiology*, edited by F. Grandori, M. Hoke, and G. L. Romani (Karger, Basel).
Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). ''Outer hair receptor currents and sensorineural hearing loss,'' Hear. Res. **42**, 47–72.
Pfeiffer, R. (**1966**). ''Classification of response patterns of spike discharges for units in the cochlear nucleus: tone-burst stimulation,'' Exp. Brain Res. **1**, 220–235.
Ponton, C. W., Moore, J. K., and Eggermont, J. J. (**1996**). ''Auditory brainstem response generation by parallel pathways: differential maturation of axonal conduction time and synaptic transmission,'' Ear Hear. **17**, 402–417.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **43**, 1120–1228.
Rhode, W. S., and Greenberg, S. (**1994**). ''Encoding of amplitude modulation in the cochlear nucleus of the cat,'' J. Neurophysiol. **71**, 1797–1825.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the bailar membrane,'' J. Neurosci. **11**, 1057–1067.
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.
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.
Rupp, A., Uppenkamp, S., Gutschalk, A., Beucker, R., Patterson, R. D., Dau, T., and Scherg, M. (**2002**). ''The representation of peripheral neural activity in the middle-latency evoked field of primary auditory cortex in humans,'' Hear. Res. **174**, 19–31.
Russell, I. J., and Cody, A. R. (**1985**). ''Transduction in cochlear hair cells,'' in *Peripheral Auditory Mechanism*, edited by J. B. Allen (Springer, Berlin).
Sachs, M. B., and Kiang, N. Y. S. (**1968**). ''Two-tone inhibition in auditory-nerve fibers,'' J. Acoust. Soc. Am. **43**, 1120–1228.
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. **81**, 680–691.
Shera, C. A. (**2001a**). ''Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion,'' J. Acoust. Soc. Am. **109**, 2023–2034.
Shera, C. A. (**2001b**). ''Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics,'' J. Acoust. Soc. Am. **110**, 332–348.
Scherg, M., and Von Cramon, D. (**1985**). ''A new interpretation of the generators of BAEP waves I–V: results of a spatio-temporal dipole model,'' Electroencephalogr. Clin. Neurophysiol. **62**, 290–299.
Sewell, W. F. (**1984**). ''The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cat,'' Hear. Res. **14**, 305–314.
Shore, S. E., and Nuttall, A. L. (**1985**). ''High synchrony compound action potentials evoked by rising frequency-swept tonebursts,'' J. Acoust. Soc. Am. **78**, 1286–1295.
Smith, J. C., Marsh, J. T., and Brown, W. S. (**1975**). ''Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources,'' Electroencephalogr. Clin. Neurophysiol. **39**, 465–472.
Stegeman, D. F., Oosterom, A. Van, and Colon, E. J. (**1987**). ''Far-field evoked potential components induced by a propagating generator: computational evidence,'' Electroencephalogr. Clin. Neurophysiol. **67**, 176–187.
Stillman, R. D., Moushegian, G., and Rupert, A. L. (**1976**). ''Early tone-evoked responses in normal and hearing impaired subjects,'' Audiology **39**, 10–22.
Teas, D. C., Eldredge, D. H., and Davis, H. (**1962**). ''Cochlear responses to acoustic transients: An interpretation of whole nerve action potentials,'' J. Acoust. Soc. Am. **34**, 1438–1459.
Tikhonov, A. N. (**1963**). ''Solution of incorrectly formulated problems and the regularization method.,'' Akust. Z. **4**, 1035–1038.
Voordecker, P., Brunko, E., and de Beyl, Z. (**1988**). ''Selective unilateral absence of attenuation of wave V of brains-stem auditory evoked potentials with intrinsic brain-stem lesions,'' Arch. Neurol. **45**, 1272–1276.
Wang, B. (**1979**). ''The relation between the compound action potential and unit discharges of the auditory nerves,'' Ph.D. thesis, Massachusetts Institute of Technology.
Warr, W. B. (**1966**). ''Fiber degeneration following lesions in the anterior ventral nucleus of the cat,'' Exp. Neurol. **14**, 453–474.

Warr, W. B. (**1982**). ''Parallel ascending pathways from the cochlear nucleus: Neuroanatomical evidence of functional specialization,'' in *Contributions to Sensory Physiology*, edited by W. D. Neff (New York, Academic), Vol. 7, pp. 1–38.
Wegner, O., and Dau, T. (**2002**). ''Frequency specificity of chirp-evoked auditory brain-stem responses,'' J. Acoust. Soc. Am. **111**, 1318–1329.
Yin, T. C. T., and Chan, J. C. K. (**1990**). ''Interaural time sensitivity in medial superior olive of cat,'' J. Neurophysiol. **64**, 465–488.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). ''A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression,'' J. Acoust. Soc. Am. **109**, 648–670.

Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing

Enrique A. Lopez-Poveda[a)]
Centro Regional de Investigaciones Biomédicas, Facultad de Medicina, Universidad de Castilla-La Mancha, Campus Universitario, 02071 Albacete, Spain

Christopher J. Plack and Ray Meddis
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

(Received 20 April 2002; accepted for publication 11 November 2002)

Cochlear nonlinearity was estimated over a wide range of center frequencies and levels in listeners with normal hearing, using a forward-masking method. For a fixed low-level probe, the masker level required to mask the probe was measured as a function of the masker-probe interval, to produce a temporal masking curve (TMC). TMCs were measured for probe frequencies of 500, 1000, 2000, 4000, and 8000 Hz, and for masker frequencies 0.5, 0.7, 0.9, 1.0 (on frequency), 1.1, and 1.6 times the probe frequency. Across the range of probe frequencies, the TMCs for on-frequency maskers showed two or three segments with clearly distinct slopes. If it is assumed that the rate of decay of the internal effect of the masker is constant across level and frequency, the variations in the slopes of the TMCs can be attributed to variations in cochlear compression. Compression-ratio estimates for on-frequency maskers were between 3:1 and 5:1 across the range of probe frequencies. Compression did not decrease at low frequencies. The slopes of the TMCs for the lowest frequency probe (500 Hz) did not change with masker frequency. This suggests that compression extends over a wide range of stimulus frequencies relative to characteristic frequency in the apical region of the cochlea. [DOI: 10.1121/1.1534838]

PACS numbers: 43.66.Dc, 43.66.Mk [MRL]

I. INTRODUCTION

The mammalian cochlear response is nonlinear in healthy animals (Rhode, 1971; Sellick *et al.*, 1982; Robles *et al.*, 1986). An increase in the magnitude of stimulation does not always produce a proportional increase in the velocity or displacement of basilar membrane (BM) vibration. It is generally accepted that for *high* characteristic frequencies[1] (CFs) the response is nonlinear for frequencies close to CF, but linear for frequencies an octave below CF (Robles *et al.*, 1986).

Using physiological techniques, cochlear responses have been measured in *animals* in terms of BM input/output (IO) functions for a wide range of CFs, stimulation frequencies, and levels (e.g., Sellick *et al.*, 1982; Robles *et al.*, 1986; Rhode and Cooper, 1996; Recio and Rhode, 2000; Rhode and Recio, 2000). The aim of the present study was to use psychophysical techniques to estimate the characteristics of the *human* cochlear response over a similar range of parameters.

The nonlinear properties of the human cochlear response can be inferred from threshold measurements of masked probe tones (for a review see Moore, 1997). A number of studies (e.g., Oxenham and Plack, 1997; Rosen *et al.*, 1998; Baker *et al.*, 1998; Glasberg *et al.*, 1999; Hicks and Bacon, 1999; Plack and Oxenham, 2000; Wojtczak *et al.*, 2001; Nelson *et al.*, 2001; Moore *et al.*, 2002) have characterized cochlear nonlinearity in normal-hearing listeners using this approach. In the present study, a revised version of the method of Nelson *et al.* (2001) was used.

The method developed by Nelson *et al.* consists of measuring the level of a pure-tone forward masker required to just mask a pure-tone probe as a function of the masker-probe time interval. The level of the probe is fixed just above absolute threshold. It is thought that the masker level at threshold depends on two variables. First, it depends on the masker-probe interval: the amount of masking decreases as the masker-probe interval increases (Zwislocki *et al.*, 1959; Duifhuis, 1973; Moore and Glasberg, 1983; Nelson and Freyman, 1987). Second, it depends on the relative excitation produced by the masker and the probe at the place on the BM tuned close to the probe frequency (Oxenham and Moore, 1995; Oxenham *et al.*, 1997; Oxenham and Plack, 1997; Nelson *et al.*, 2001). Because the probe level is fixed at all times, the method is assumed to measure the masker level (input) required to generate a fixed level of excitation after decaying during the masker-probe interval. This is the reason that the resulting functions are referred to as *iso-response* temporal masking curves (TMCs).

Obviously, higher masker levels are required as the masker-probe interval increases. However, the slope of the TMC depends on the masker frequency. It has been argued (Nelson *et al.*, 2001) that this is because on-frequency maskers are subject to cochlear compression while others are processed more linearly. Therefore, the slope of the TMCs reflects the amount of compression for a given masker. Nelson *et al.* showed this behavior for a probe frequency of 1 kHz and a wide range of masker frequencies. By assuming that the internal effect of the masker decays at the same rate

a)Electronic mail: enrique.lopezpoveda@uclm.es

regardless of masker frequency, and that maskers well below the probe frequency yield a *linear* cochlear response, they derived human cochlear IO curves at CF~1000 Hz by plotting the masker levels for the low-frequency masker (a linear reference) as a function of the masker levels for other masker frequencies.

This approach has some advantages over previous methods (e.g., Oxenham and Plack, 1997; Rosen *et al.*, 1998; Baker *et al.*, 1998; Plack and Oxenham, 2000). Fixing the probe level almost guarantees that the region of the cochlea under study is the same for different masker (input) levels. Furthermore, fixing the probe level just above threshold ensures that the CF of the cochlear region under study is close to the probe frequency. In other words, the effects of ''off-frequency listening'' are minimized.[2]

In the present study, TMCs were measured for probe frequencies from 500 to 8000 Hz, and for a range of masker frequencies at each probe frequency. It will be argued that for low probe frequencies, cochlear responses are compressed for maskers well below the probe frequency. This undermines the assumptions of the method developed by Nelson *et al*. for deriving cochlear IO curves from TMCs. An alternative method is suggested based on the more limited assumption that the response to below-CF tones is linear at *high CFs only* (see also Plack and Drga, submitted). It has been suggested by physiological results in the chinchilla (Rhode and Cooper, 1996) and guinea pig (Cooper and Yates, 1994), and by masking studies in humans (Hicks and Bacon, 1999; Plack and Oxenham, 2000), that compression is reduced at low CFs. The new method allowed a test of this hypothesis. Finally, a control experiment is reported that tested the effects of probe and masker ramp durations on the form of the TMCs.

II. METHOD

A. Stimuli

TMCs were measured for probe frequencies (f_p) of 500, 1000, 2000, 4000, and 8000 Hz, and for masker frequencies (f_m) of 0.5, 0.7, 0.9, 1.0, 1.1, and $1.6\times f_p$. For any given pair (f_m, f_p), masked thresholds were measured for masker-probe intervals (Δt) ranging from 10 to 100 ms in steps of 10 ms. Δt was defined as the duration of zero-amplitude points between the masker offset and the probe onset. The sinusoidal maskers were gated with 4-ms raised-cosine onset and offset ramps and had a total duration of 108 ms. The sinusoidal probes had a total duration of 8 ms and were gated with 4-ms raised-cosine ramps (no steady-state portion). For each f_p, the level of the probe was kept constant at 14 dB above the listener's absolute threshold for the probe.

Stimuli were generated digitally on a Silicon Graphics O2 workstation at a sampling rate of 32 kHz, with 16-bit resolution. They were played monaurally via the workstation headphone connection through a pair of circumaural Sennheiser HD-580 headphones. Listeners sat in an EYMASA CI-40 single-walled sound-attenuating booth. The booth was placed in a quiet environment to further reduce background noise. The sound pressure levels (SPLs) reported below are nominal electrical levels without allowing for the earphone diffuse-field response.

B. Procedure

The procedure was similar to that used by Plack and Oxenham (1998). Masked thresholds were measured using a two-interval, two-alternative forced-choice paradigm. In one interval, the masker tone was presented alone. In the other interval, the masker was presented followed by the probe. The two intervals were presented to the listener in random order, but each of them coincided in time with the highlighting of a window on the workstation monitor. Listeners were asked to select the interval containing the probe by pressing ''1'' or ''2'' on the numerical keyboard of the workstation, depending on whether the probe was judged to accompany the first or the second light, respectively. Visual feedback was immediately provided to the listener by means of a green or red highlighted window on the monitor, indicating correct and incorrect answers, respectively.

The initial masker level was 35 dB SPL. A two-up, one-down adaptive rule was used to estimate the 71% correct point on the psychometric function (Levitt, 1971). The level of the masker was increased and decreased by 4 dB for the first four turnpoints, and by 2 dB thereafter. Sixteen turnpoints were recorded in each experimental block and the threshold estimate was taken as the mean of the masker levels at the last 12 turnpoints. For masker levels below approximately 90 dB SPL at least three estimates were made for each condition, and the results were averaged. In some cases, it was difficult to make three measurements for masker levels above 90 dB SPL because clipping often occurred during the adaptive procedure and listeners were instructed to stop the experiment at the first sign of clipping. When this occurred with one of the estimates, the two remaining estimates were averaged. It follows that the reported masker levels above 90 dB SPL are likely to be underestimates of the true threshold.

C. Listeners

Data were collected for the left ear of three listeners (CMR, ALN, and ELP, aged 22, 26, and 31, respectively) with normal hearing.[3] Listener ELP was one of the authors, but, like the other two listeners, had no previous experience on the task. Absolute thresholds were measured for tones of the same frequencies and durations as the probes and maskers used in the forward-masking experiment. Each threshold was measured at least three times and the results (see Fig. 1) were averaged. Listeners were given at least 10 h of practice on the forward-masking task before data collection began.

III. RESULTS

A. Temporal masking curves

Results are shown in Fig. 2. Each column corresponds to a different listener (or the mean). Each row corresponds to a different probe frequency (from 500 Hz in the top row to 8000 Hz in the bottom row). As explained above, data points

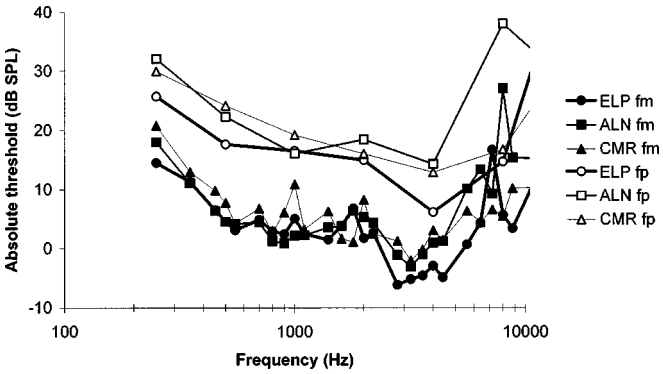


FIG. 1. Absolute hearing thresholds for the three listeners (ELP, CMR, and ALN). Open symbols show the thresholds for the 8-ms probe tones (f_p). Filled symbols represent the thresholds for the 108-ms masker tones (f_m). Data points represent the mean of three measurements. Note the high thresholds of ELP for the 7200-Hz masker and of ALN for the 8000-Hz probe and masker. These coincide with deep notches in the listeners' headphone-related frequency responses[3] (not shown).

below around 90 dB SPL are based on the average of three measurements, whereas those above 90 dB SPL are sometimes based on the average of two measurements only. Standard deviations across measurements are not shown in order to avoid clutter. These were variable across conditions and listeners, ranging from 0.0 dB (listener ELP, $f_p=500$ Hz, $f_m=800$ Hz, $\Delta t=10$ ms) to 23.0 dB (listener ELP, $f_p=8000$ Hz, $f_m=8000$, $\Delta t=50$ ms), with a mean and a standard deviation across listeners and conditions of 3.1 and 2.3 dB, respectively. The right-most column in Fig. 2 shows average TMCs across the three listeners.

Masker levels increase as Δt increases. This is reasonable since the recovery from masking is greater for longer masker-probe intervals (Zwislocki *et al.*, 1959; Duifhuis, 1973; Moore and Glasberg, 1983; Nelson and Freyman, 1987). However, for *high* probe frequencies, the rate of increase is markedly different for maskers an octave below the probe frequency (unfilled diamonds) than for masker frequencies equal to the probe frequency (filled circles). For $f_m=0.5f_p$, the TMC shows a single slope. As f_m approaches f_p, however, the TMC can be described in most cases as a two-sloped function, with a steeper slope at short to moderate Δt, and a shallower slope at long Δt. In some cases (e.g., ELP, $f_m=f_p=8000$ Hz), the TMC shows a *three*-slope pattern with a shallow slope at short Δt, followed by a steeper slope at moderate Δt, followed by a shallow slope again at longer Δt.

B. Interpretation of the TMCs

The interpretation of the TMCs depends upon two assumptions: (1) that the internal representation of the masker decays with time at the same rate for all masker frequencies and (2) that the residual excitation at the time of the probe at masked threshold is the same for all maskers. Similar assumptions were made by Nelson *et al.* (2001).

The first assumption (uniform rates of decay across masker frequency) makes it possible to identify nonlinear increases of excitation strength with masker level. To understand how this is possible, it helps to consider that the decay is exponential. An exponential decay is consistent with previous studies of recovery from forward masking (e.g., Duifhuis, 1973; Widin and Viemeister, 1979) and is also supported by the present data. The *y*-axis in Fig. 2 is logarithmic and, therefore, a straight-line is consistent with a simple exponential decay. Some straight lines are, indeed, evident. For example, the combination of a 2000-Hz masker and a 4000-Hz probe produces a straight line. Other examples can also be seen, particularly for high probe frequencies paired with low masker frequencies. In these cases, it is assumed that the masker excitation increases as a simple linear function of masker level. These instances can be used as linear reference functions (see below). Therefore, when the slope of the TMC becomes steeper than the linear reference function, this is an indication that the masker is subject to compression. For example, at high frequencies the on-frequency TMC slopes are generally steeper than the TMC slope for a masker an octave below the probe. This suggests that the on-frequency masker is being compressed.

The second assumption allows the reconstruction of the shape of the cochlear IO functions from the TMCs. The method consists of plotting the masker level of a linear-reference masker against the level for the masker of interest (Nelson *et al.*, 2001; Plack and Drga, submitted), where each pair of levels has the same masker-probe interval. The resulting curve reveals the cochlear IO function by compensating for the decay of internal masker excitation. Note that the function describing the decay of internal masker excitation cancels out in this process if it is the same for all masker frequencies. Hence, its actual form (whether exponential or otherwise) is irrelevant.

C. The choice of the best linear reference

The choice of a linear reference is critical if valid estimates of cochlear compression are to be made from the derived cochlear IO functions. A careful examination of Fig. 2 shows that the slope of TMCs for $f_m=0.5f_p$ is steeper for $f_p=500$ Hz than for $f_p=8000$ Hz. Furthermore, the former is closer to the steeper portion of the TMC for maskers at the probe frequency. To make this observation clearer, straight lines (dashed lines in Fig. 2) were fit by a method of least squares to the TMCs for $f_m=0.5f_p$, and their slopes were plotted as a function of f_p. The results are shown in Fig. 3. The slopes of the TMCs for $f_m=0.5f_p$ are much higher for lower f_p's, decrease with increasing f_p up to 2000 Hz, and then remain relatively constant.

Given that the shape of the TMC may be influenced both by the decay of the internal masker effect with time *and* by cochlear compression, this observation may be interpreted in two ways. It may mean that the rate of decay of the masker effect is *faster* for lower probe frequencies (that is, the first assumption would be incorrect). This explanation, however, is unlikely. It would imply that the temporal resolution of the auditory system improves at low frequencies. Shailer and Moore (1987) have shown that this is not the case (see also Moore *et al.*, 1993). They studied the detection of gaps in sinusoids and concluded that it varies little for frequencies between 200 and 2000 Hz and, if anything, becomes poorer

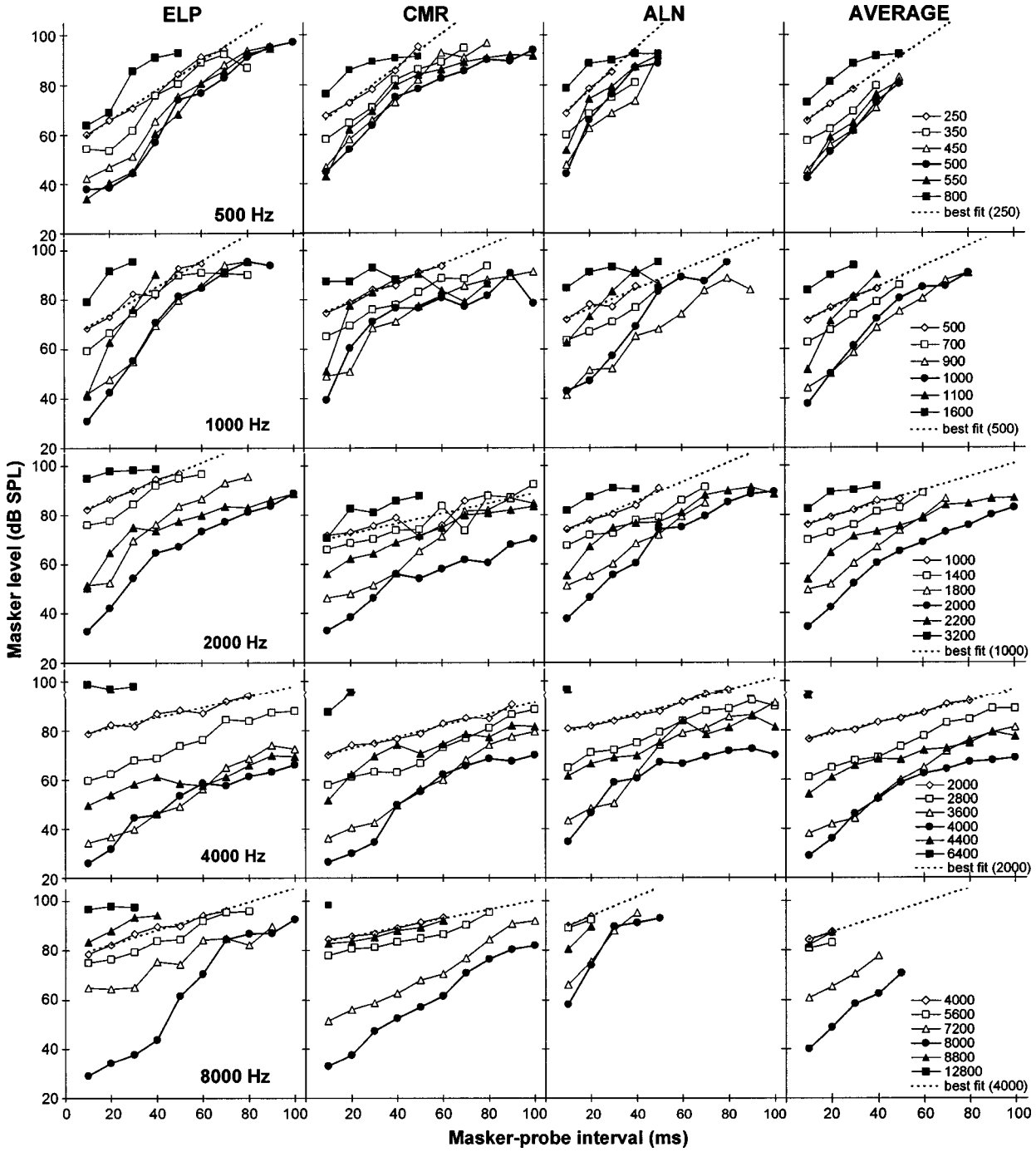


FIG. 2. Iso-response temporal masking curves for each listener (columns) at the five probe frequencies (f_p) tested (rows). The right-most column shows the results averaged across the three listeners. Insets in the panels of the right column show the masker frequencies (f_m). Open symbols represent conditions where $f_m < f_p$. Filled symbols represent conditions where $f_m \geq f_p$. The dotted lines represent the best-fit (by least squares) straight lines for the condition $f_m = 0.5 f_p$, the slopes of which are plotted in Fig. 3.

at low frequencies. Therefore the second, and most likely, interpretation of the data in Fig. 3 is that the human cochlear response for low probe frequencies is *still compressive* as the stimulus frequency is moved *below CF*. In other words, not only do the present data provide evidence for substantial compression at low CFs, but they also support the physiological finding that compression is not frequency dependent at low CFs (Rhode and Cooper, 1996).

As a result of this analysis, the individual (or average) TMCs for $f_p = 4000$ Hz, $f_m = 2000$ Hz, were chosen as the optimum linear references to derive the cochlear IO curves for each listener (or for the average). Furthermore, these linear references were *fixed* across probe frequencies. There are several reasons for this choice: First, the TMCs in question appear as shallow straight lines, suggesting no deviation from linearity across level; second, the slope of the TMC for this condition is the least variable across listeners (see Fig. 3); third, a large number of data points are available for every listener; and finally, the available physiological data (Rhode and Recio, 2000) at reasonably close CFs (5500 Hz)

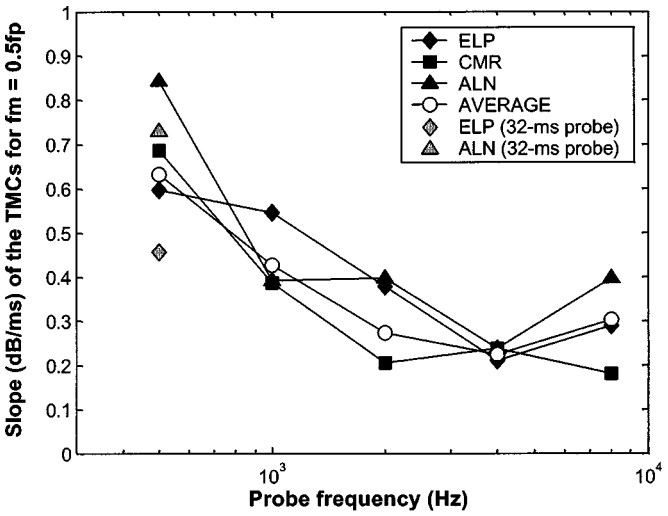


FIG. 3. Slopes of straight-line best fits to the TMCs (Fig. 2, dotted lines) for masker frequencies an octave below the probe frequency. Note that the slope decreases with increasing probe frequency up to 2000 Hz, and then remains approximately constant. This suggests that the cochlear response to low-frequency tones may be compressed at low CFs (see text for details). Filled and open symbols represent the slopes of below-f_p TMCs for a short, 8-ms probe with 4-ms onset/offset ramps. Gray symbols (at 500 Hz only) represent the slopes of below-f_p TMCs for a longer, 32-ms probe with 16-ms ramps (see Fig. 5 and main text in Sec. III E).

suggest that a 2000-Hz stimulus frequency will produce a linear response at the 4000-Hz place.

For convenience, instead of using the original data as the linear reference, a smoothed version was used. This was obtained by reading off a new masker level for each Δt from the regression lines fit (dashed lines in Fig. 2) to the TMCs in question. Therefore, for each listener, cochlear IO functions were derived by plotting their individual, smoothed, linear reference against the masker levels for any other TMC.

D. Derived cochlear IO curves

Figure 4 shows the resulting cochlear IO curves (note that the *y*-axis scale is different for different panels). To ease the physiological interpretation, f_p and f_m have been equated to CF and stimulus frequency, respectively, in the discussion below.[4]

1. Compression at CF

All of the IO curves for tones at CF show shallow slopes (<1 dB/dB) for a range of input levels. This suggests that compression at CF occurs *across the range* of CFs tested. To facilitate a quantitative analysis, the curves at CF are considered as two- or three-stage functions showing two or three segments, L1, L2, and L3, with markedly different slopes at low, moderate, and high input levels, respectively. The limits of these segments are depicted (after visual inspection) by the vertical thin line in each panel of Fig. 4 (note that L1 and/or L3 might not be present in some curves). Table I shows the slopes in every segment for each CF, for each listener, and for the average data across listeners.

The slope of segment L2, where compression is most obvious, is approximately constant at 0.2–0.3 dB/dB across CF. Although larger variability must be acknowledged when looking at the values for individual listeners (varying from 0.15 dB/dB at 8000 Hz for ALN to 0.38 dB/dB at 8000 Hz for CMR), the most common value is also within the range 0.2–0.3 dB/dB. This suggests compression ratio (inverse of the slope) estimates of 3:1 to 5:1 across the CF range tested. Remarkably, compression does *not* decrease for lower CFs, as has been suggested previously (Hicks and Bacon, 1999; Plack and Oxenham, 2000).

The slopes of L1 and L3 are less than one in most cases, suggesting that the cochlear response may be compressive also for low and high input levels. However, the values are always larger than the slope of L2. This is consistent with other studies that have reported less compression, approaching linearity, for very low (e.g., Nelson *et al.*, 2001, p. 2054) and very high signal levels (e.g., Plack and Oxenham, 1998; Nelson *et al.*, 2001). Indeed, the slope of L3 is close to unity for CFs of 4000 and 8000 Hz. For one case only, its value exceeds unity considerably (1.46 dB/dB), but this corresponds to a condition (ALN at 8000 Hz) for which only two data points are available (see Fig. 4).

2. Compression below CF

Figure 4 shows *compressive* IO curves for *tones below CF at low CFs*. The slopes of straight lines fit to the IO curves for 0.5CF tones are <1.0 for CFs$\leqslant$2000 Hz (see Table I). Overall, there is a trend for the below-CF slopes to increase with CF until they approach unity at 4000 Hz, suggesting that the response to below-CF tones becomes linear for high CFs. The slopes of the IO curves for stimulus frequencies of 0.5CF at CF$=$4000 Hz are necessarily very close[5] to unity because these curves were used as the basis for the linear reference for deriving all other IO curves.

However, it is noteworthy that the slopes of the 0.5CF curves at CF$=$8000 Hz differ from unity. They are lower for two listeners (ELP and ALN) and for the average, but higher for listener CMR. The deviation from unity, and the observed variability, may be the result of slope estimates that are based on considerably fewer points[6] than at 4000 Hz, particularly for ALN and the average data sets.

E. Detection mediated by spectral splatter

The spectral splatter produced by a short probe may improve the detectability of the probe in some circumstances. For a given probe duration, the effects would be expected to be greatest at low frequencies, where cochlear frequency selectivity is greatest (i.e., absolute filter bandwidths are narrower), and hence where the spread of excitation produced by the splatter would be most detectable. The probe used in the current experiments was relatively short (8 ms). It could be argued, therefore, that the detection of spectral splatter may have had an influence on the masker levels at threshold at low frequencies.

Furthermore, the detection of the probe may be also affected by the spectral splatter caused by an abrupt masker offset. A remote-frequency masker may be more effective with a short decay ramp because the spectral splatter caused by its abrupt offset may reach the place on the BM tuned to the probe frequency. In the current experiments, the masker decay ramps were relatively short (4 ms). Therefore, it could

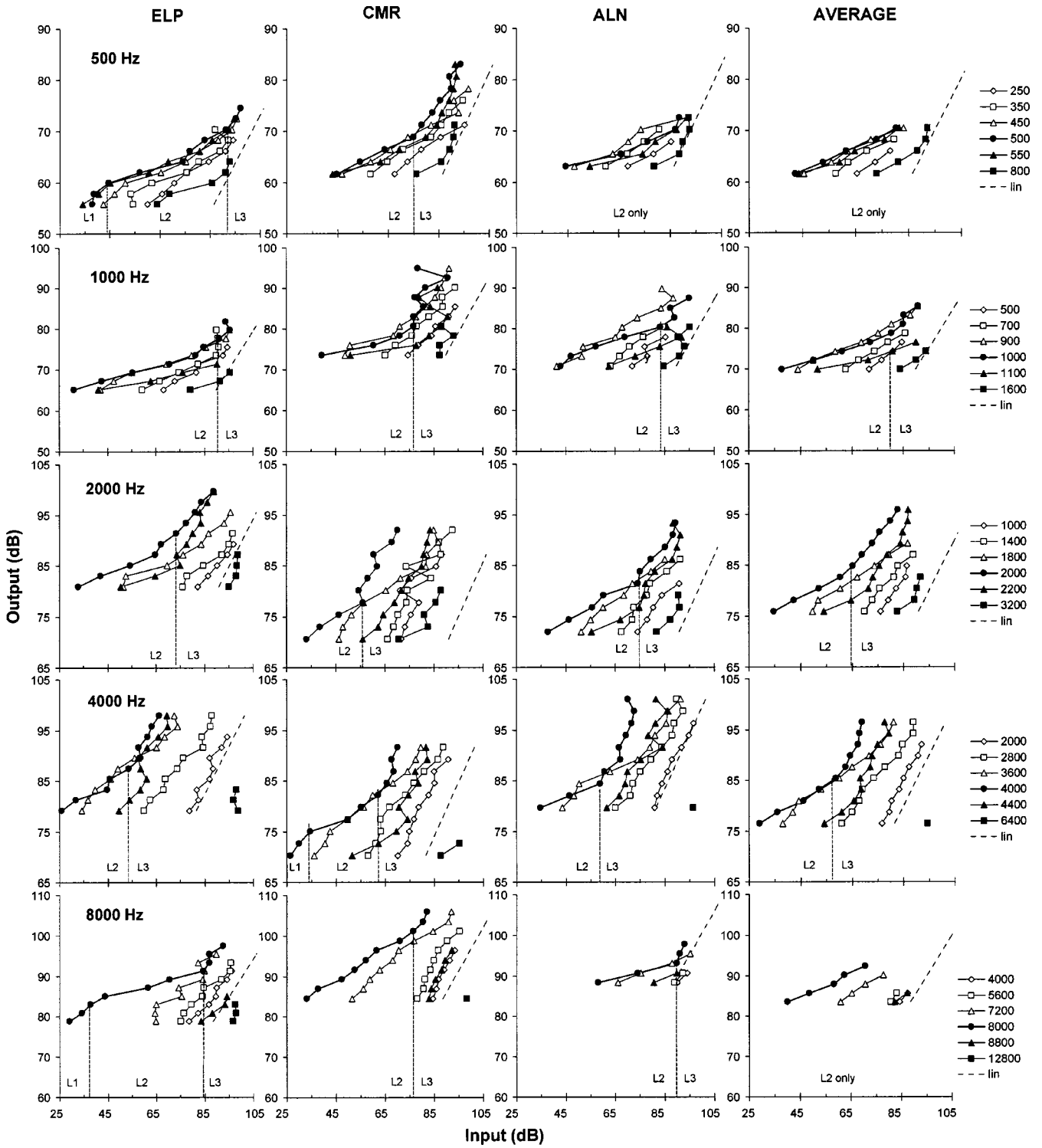


FIG. 4. Cochlear input/output curves derived from the TMCs in Fig. 2. Each row corresponds to a different CF or probe frequency in Fig. 2 (in bold on the left column panels). Legends on the right show the stimulus frequencies (f), corresponding to the masker frequencies in Fig. 2. Open symbols represent conditions where $f<$CF. Filled symbols represent conditions where $f\geqslant$CF. Dashed lines illustrate linear growth. The thin vertical lines only apply to the on-CF curves, and delimit segments (L1, L2, and L3) with clearly different slopes after visual inspection. The curves were derived assuming that the TMC for $f_p=4000$ Hz, and $f_m=2000$ Hz reflects a cochlear linear response (see text for details). Under this assumption, responses at CF are compressed over the whole range of CFs. Moreover, the degree of compression over segment L2 varies little across CFs (see Table I). At low CFs, compression extends to tones an octave below CF.

be argued that the reported levels for *remote* maskers are lower than would have been obtained if longer ramps had been used. The effect would not occur for on-frequency maskers, and would be less important for off-frequency maskers at high probe frequencies, where the frequency differences between the masker and the probe were large.

To investigate these possibilities, TMCs were measured for two listeners (ELP and ALN), for $f_p=500$ Hz, and for two masker frequencies (f_p, and $0.5f_p$). This time, however, the total duration of the probe was 32 ms (16-ms ramps, no

TABLE I. Slopes (dB/dB) of the cochlear IO curves of Fig. 4. Slopes are given for IO curves corresponding to stimulus frequencies (f) of CF and 0.5CF. For the IO curves at CF, two or three slopes are given (CF/L1, CF/L2, CF/L3) corresponding to each of the characteristic segments depicted in Fig. 4. N/A: segment not observed, or insufficient data points for a good slope estimate.

	CF (Hz)				
f (Hz)	500	1000	2000	4000	8000
			Listener ELP		
CF/L1	0.54	N/A	N/A	N/A	0.47
CF/L2	0.22	0.19	0.24	0.29	0.17
CF/L3	0.69	0.54	0.56	0.85	0.72
0.5CF	0.35	0.37	0.55	0.93	0.71
			Listener CMR		
CF/L1	N/A	N/A	N/A	0.57	N/A
CF/L2	0.24	0.22	0.30	0.25	0.38
CF/L3	0.75	0.38	0.81	1.15	0.80
0.5CF	0.34	0.61	0.71	0.97	1.29
			Listener ALN		
CF/L1	N/A	N/A	N/A	N/A	N/A
CF/L2	0.20	0.22	0.27	0.20	0.15
CF/L3	N/A	0.56	0.60	1.05	1.46
0.5CF	0.28	0.51	0.58	0.98	0.60
			Average responses		
CF/L1	N/A	N/A	N/A	N/A	N/A
CF/L2	0.23	0.21	0.28	0.29	0.29
CF/L3	N/A	0.52	0.62	1.04	N/A
0.5CF	0.35	0.47	0.79	0.99	0.74

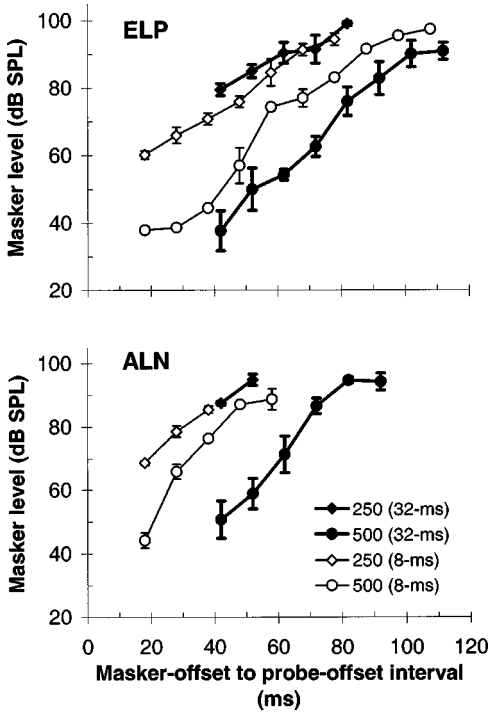


FIG. 5. TMCs for $f_p = 500$ Hz and two masker frequencies (250 and 500 Hz), for different probe durations, and different ramp durations on the masker. Each panel corresponds to a different listener (ELP and ALN). The legend informs about the masker frequency (Hz), and the total duration of the probe (ms). Open symbols represent TMCs (replotted from Fig. 2) for 8-ms probes and 108-ms maskers, both gated with 4-ms onset and offset ramps. Filled symbols represent TMCs for 32-ms probes and 132-ms maskers both gated with 16-ms onset and offset ramps. Every black datapoint is the average of at least four measurements. Error bars represent one standard deviation across trials.

steady-state portion). The masker had a total duration of 132 ms and was gated with 16-ms rise/decay ramps. The probe level was fixed at 14 dB above absolute threshold for the 32-ms probe. At least four measurements were made per condition (even for masker levels above 90 dB SPL). The average results are shown in Fig. 5. The results from the main experiment (with 8-ms probe, 4-ms ramps on the masker) are replotted from Fig. 2 for comparison. This time, however, masker level is plotted against the time interval between the masker offset and the probe offset (at the half-amplitude points). For a given value of the offset-onset interval, the duration of the offset-offset interval is different for both experiments.

Previous work (Zwislocki *et al.*, 1959) suggests that poststimulatory thresholds depend mainly on the time interval between the masker-offset and the probe-offset, rather than on the duration of the zero-amplitude gap or of the probe. Therefore, when plotted against the offset-offset interval (as in Fig. 5), the masker levels for both experiments should overlap (unless other effects, such as those described above, mediate probe detection). This is the case for the off-frequency masker, but not for the on-frequency masker. For the latter, masker levels are considerably lower for the long-ramp/long-probe condition, particularly for short to moderate offset-offset intervals.

A possible interpretation of these results is that detection of the short probe is *not* facilitated by splatter. Otherwise, masker levels for the short probe would be consistently higher both for on- and off-frequency maskers. This explanation seems reasonable, as the level of the probe was too low to cause significant splatter (Moore, 1981). If this were true, the lower masker levels for the on-frequency masker in the long-ramp/long-probe condition could be the result of ''confusion'' (Terry and Moore, 1977; Moore and Glasberg, 1982). That is, probe detection would be harder because it would be harder to distinguish the probe from a continuation of the masker as a result of using longer ramps (Moore, 1981). This confusion would not occur for the off-frequency masker because its frequency differs considerably from the probe frequency.

An alternative explanation for the results is that spectral splatter caused by the short probe *does* facilitate detection, both for on- and off-frequency maskers. The reason that it does not affect threshold for the off-frequency masker may be that the effect is cancelled out by the additional masking produced by the short *masker* ramps, as suggested above.

In any case, the most important feature of the data in Fig. 5 is, perhaps, that the shapes of the TMCs from both experiments are similar. The slopes of the TMCs for the 250-Hz masker are *slightly* shallower for the 32-ms probe than for the 8-ms probe (see Fig. 3). This suggests that the steepness of the $0.5f_p$ TMCs at 500 Hz may be attributed in part to using short probes and/or short ramps. However, Fig. 3 also shows that the slopes of the TMCs for the longer probe are still considerably greater than the slopes for the $0.5f_p$ TMCs at 4000-Hz, which are assumed to reflect a linear cochlear response. Therefore, it can be reasonably con-

cluded that compression of below-CF tones occurs at low CFs, although the slopes for the 0.5CF tones at 500 Hz given in Table I may be slight overestimates of the amount of compression.

IV. DISCUSSION

The aim of this paper was to compare the characteristics of the human cochlear response with measurements made physiologically in other mammals. In particular, the aim was to study compression as a function of CF, over the range of CFs from 500 to 8000 Hz.

A. Cochlear compression across CFs

The results presented here suggest that the response of the human cochlea to tones at CF is compressed over the studied frequency range (Fig. 4, right-most column), and that, on average, the amount of compression at moderate levels varies between 3:1 and 5:1 across CFs (see slopes CF/L2 in Table I). Compression at CF does not decrease for lower CFs. Additionally, the results suggest that compression spans a wider frequency range relative to CF at the low CFs (see slopes 0.5CF in Table I).

B. Assumptions and interpretations

These conclusions are based on the assumption that the response to a 2000-Hz tone is linear at a CF of 4000 Hz, but allows for the possibility that the off-frequency response may be compressive for lower CFs. Support in favor of this assumption comes from the data in Figs. 2 and 3, and from recent data on TMCs and forward-masking growth with level (Plack and Drga, submitted). The present assumption is also supported by BM responses to tones well below CF in chinchilla, which appear to be compressive for CFs around 400–800 Hz (Rhode and Cooper, 1996, Fig. 7), but linear from 5500 to 14000 Hz (Rhode and Recio, 2000).

The choice of the linear reference is critical when cochlear compression at CF is estimated by comparison of *on*-CF, and *below*-CF responses. Different assumptions lead to different conclusions. For instance, Plack and Oxenham (2000) suggested that, in contrast to the present results, compression on the human BM increases from 1.3:1 at 500 Hz, to 2.8:1 at 4000 Hz, or 2.4:1 at 8000 Hz. However, they assumed that linear responses to below-CF tones occur for *any* CF. Plack and Oxenham acknowledged that their results, and those of Hicks and Bacon (1999), are consistent with high compression at low CFs, *if* the compression does not vary with frequency in the apical region of the cochlea. The present results suggest that their estimates of compression for tones at CF should be regarded as *relative* to the compression for the below-CF tones. Estimates of relative compression from the present data can be derived from the values in Table I as the ratio of the slopes of derived IO curves for 0.5CF and CF/L2. The resulting values (based on the average responses) range from 1.5:1 at 500 Hz, to 3.4:1 at 4000 Hz or 2.55:1 at 8000 Hz. These estimates closely match those reported by Plack and Oxenham (2000).

C. Comparison with other studies of auditory nonlinearity

The present results are consistent with other psychophysical studies where no specific assumptions were made about the linearity of the response for tones below CF. For instance, Duifhuis (1980) reported slightly larger amounts of two-tone suppression for a 200-Hz suppressor and a 500-Hz suppressee than for an 800-Hz suppressor and 2000-Hz suppressee (see Fig. 12 in Duifhuis, 1980). Suppression is likely to be evidence of compression (Rhode and Cooper, 1993). The fact that low-frequency suppressor tones produce similar amounts of suppression on probe tones of 500 and 2000 Hz suggests similar amounts of compression at CFs of 500 and 2000 Hz. Plack and Drga (submitted) reported similar TMCs to those presented here at probe frequencies of 250, 500, and 4000 Hz and reached the same conclusions. In addition, they showed that the growth of forward masking with masker level, another estimate of compression, does not vary between 250 and 4000 Hz. Oxenham and Dau (2001) reported large effects of the relative phase of harmonics on the amount of masking produced by a complex tone. If a system is compressive, then the response to peaky waveforms is less than that to flat waveforms, for the same input rms level. An effect on masking of harmonic phase, which alters the envelope of the waveform, is taken as evidence for auditory compression. Oxenham and Dau found large phase effects for signal frequencies as low as 125 Hz. A final result in support of the present findings is that loudness growth with level, which may be related to cochlear compression (Schlauch *et al.*, 1998), hardly varies across the range of CFs studied here (see Moore, 1997 for a review; see also Plack and Drga, submitted).

In contrast to the results in humans, compression may decrease for low CFs in other mammals. In *chinchilla*, BM responses for tones at CF appear more linear at CFs $\sim$400–800 Hz (Rhode and Cooper, 1996) than at CFs between 4000 and 14000 Hz (Rhode and Recio, 2000). As for *guinea pig*, Cooper and Yates (1994) derived cochlear IO functions over a wide range of CFs from auditory-nerve fiber responses. For each fiber, they plotted the response rate for a tone well below CF against the response rate for a tone of the same level but at CF. Their results show a distinct variation in the degree of compression along the length of the guinea pig cochlea. Their compression ratio estimates vary from 2:1 for CFs$<$4000 Hz to as much as 7:1 for CFs$>$4000 Hz. Again, they assumed linear cochlear responses to tones well below CF for *all* fibers. This assumption may be justified in their case (guinea pig), because of ''...the relative stability of the below-CF (auditory-nerve fiber) rate-level slopes with CF'' (Cooper and Yates, 1994, p. 230), as shown in their Fig. 5A. However, the TMC data in Fig. 2 suggests that the same is not true for humans, as the slopes of the TMCs for maskers well below CF do vary across CFs (Fig. 3). Therefore, while it may be justified to conclude that compression is reduced for low CFs in guinea pigs, the same may not be true for humans.

D. The source of compression

Despite the focus on cochlear processing in the current discussion, it could well be argued that the compression at low CFs inferred from Fig. 2 does not originate in the cochlea. It could reflect, instead, other nonlinear processes in the auditory receptor, such as the saturation of the receptor potential of inner-hair cells, or of auditory-nerve fiber discharge rates, which need not be frequency-specific relative to CF. If these nonlinear processes are different for different CFs, they might account for the observed decrease in the slope of the TMCs for $0.5f_p$ maskers with increasing f_p illustrated in Fig. 3. The issue may be resolved by studies on listeners with sensorineural hearing loss at low frequencies. For example, if the TMCs for these listeners were shallower than those for normal-hearing listeners, then that would be good evidence that the compression is cochlear in origin.

V. CONCLUSIONS

The main conclusions of the present study can be summarized as follows:

(i) Human cochlear responses to tones at CF are compressed over the CF range from 500 to 8000 Hz. On average, the estimated compression for moderate input levels ranged from 3:1 to 5:1. Compression does *not* decrease for lower CFs, as has previously been suggested.

(ii) Compression extends over a wider range of stimulus frequencies at low CFs than at high CFs. The estimated compression to tones *an octave below* CF decreased with increasing CF, from 2.8:1 at CF $=500$ Hz to approximately 1:1 at CF$=4000$ Hz.

ACKNOWLEDGMENTS

This work was supported by the Consejería de Sanidad of the Junta de Comunidades de Castilla-La Mancha (ref. 01044), and by EPSRC Project Grant No. GR/R65794/01. We thank Consuelo Martínez Redondo and Alberto López-Nájera for gathering some of the data presented in this report. We also thank José Luis Blanco and Almudena Eustaquio-Martin for technical support. We are indebted to Brian C. J. Moore and an anonymous reviewer for their excellent and helpful reviews of earlier versions of this paper.

[1]It is a physiological property of the mammalian cochlea (at least in nonhuman mammals) that the frequency of a pure-tone stimulus required to yield a maximal response at a given BM site changes with stimulation level (Johnstone *et al.*, 1986; Sellick *et al.*, 1982). In the present report, the term ''characteristic frequency'' (CF) refers to the frequency of a pure-tone stimulus that yields a maximum response at stimulation levels close to absolute hearing threshold.

[2]The use of a fixed, low-level probe confines the spread of the probe's excitation pattern and, hence, reduces off-frequency listening. However, off-frequency listening is not fully eliminated (Johnson-Davis and Patterson, 1979; O'Loughlin and Moore, 1981). For this reason, care is taken to distinguish between f_p and CF in the text. f_p is used when describing the psychophysical data, whereas the term CF is used when discussing the data in terms of the physiological behavior; for instance, when commenting on the cochlear IO functions derived from TMCs.

[3]The threshold of listener ALN at 8000 Hz was 27 dB SPL; that is, 11 dB above the normal audibility threshold for a circumaural headphone according to ANSI 3.6-1996. Hence, this listener could be argued to be at the limit of normal hearing at this frequency. However, this high threshold corresponded to a sharp notch in the listener's headphone frequency response (not shown). Therefore, the threshold is possibly the result of sound cancellation in the external ear (Lopez-Poveda and Meddis, 1996) and not of cochlear damage.

[4]Equating CF to f_p is not strictly correct because, as explained in footnote 2, off-frequency listening may occur.

[5]IO curves were actually derived by plotting a *smoothed* version of the low-frequency data at CF$=4000$ Hz (the linear reference) as a function of the original data. The linear reference was obtained by linear regression of the original data. Therefore, the abscissa and ordinate values are not identical. That is the reason that the slopes differ slightly from unity.

[6]Data points for longer Δt were not collected because of clipping problems.

Baker, R. J., Rosen, S., and Darling, A. (**1998**). ''An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London).

Cooper, N. P., and Yates, G. K. (**1994**). ''Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve fibres: Variations with characteristic frequency,'' Hear. Res. **78**, 221–234.

Duifhuis, H. (**1973**). ''Consequences of peripheral frequency selectivity for nonsimultaneous masking,'' J. Acoust. Soc. Am. **54**, 1471–1488.

Duifhuis, H. (**1980**). ''Level effects in psychophysical two-tone suppression,'' J. Acoust. Soc. Am. **67**, 914–927.

Glasberg, B. R., Moore, B. C. J., and Stone, M. A. (**1999**). ''Modeling changes in frequency selectivity with level,'' in *Psychophysics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohman, and B. Kollmeier (World Scientific, Singapore).

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.

Johnson-Davis, D., and Patterson, R. D. (**1979**). ''Psychophysical tuning curves: Restricting the listening band to the signal region,'' J. Acoust. Soc. Am. **65**, 765–770.

Johnstone, B. M., Patuzzi, R., and Yates, G. K. (**1986**). ''Basilar membrane measurements and the travelling wave,'' Hear. Res. **22**, 147–153.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lopez-Poveda, E. A., and Meddis, R. (**1996**). ''A physical model of sound diffraction and reflections in the human concha,'' J. Acoust. Soc. Am. **100**, 3248–3259.

Moore, B. C. J. (**1981**). ''Interactions of masker bandwidth with signal duration and delay in forward masking,'' J. Acoust. Soc. Am. **70**, 62–68.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, London).

Moore, B. C. J., and Glasberg, B. R. (**1982**). ''Contralateral and ipsilateral cueing in forward masking,'' J. Acoust. Soc. Am. **71**, 942–945.

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., Alcántara, J. I., and Glasberg, B. R. (**2002**). ''Behavioural measurement of level-dependent shifts in the vibration pattern on the basilar membrane,'' Hear. Res. **163**, 101–110.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1993**). ''Detection of temporal gaps in sinusoids: Effects of frequency and level,'' J. Acoust. Soc. Am. **93**, 1563–1570.

Nelson, D. A., and Freyman, R. L. (**1987**). ''Temporal resolution in sensorineural hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 709–720.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). ''A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 2045–2064.

O'Loughlin, B. J., and Moore, B. C. J. (**1981**). ''Off-frequency listening: Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking,'' J. Acoust. Soc. Am. **69**, 1119–1125.

Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Oxenham, A. J., and Dau, T. (**2001**). ''Towards a measure of auditory filter phase response,'' J. Acoust. Soc. Am. **110**, 3169–3178.
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.
Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Plack, C. J., and Oxenham, A. J. (**2000**). ''Basilar-membrane nonlinearity estimated by pulsation threshold,'' J. Acoust. Soc. Am. **107**, 501–507.
Plack, C. J., and Drga, V. (**in press**). ''Psychophysical evidence for auditory compression at low characteristic frequencies,'' J. Acoust. Soc. Am.
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. Suppl. 2 **49**, 1218.
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 cochlea *in vivo*,'' Aud. Neurosci. **3**, 101–121.
Rhode, W. S., and Recio, A. (**2000**). ''Study of mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**, 3317–3332.
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.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.
Schlauch, R. S., DiGiovanni, J. J., and Reis, D. T. (**1998**). ''Basilar membrane nonlinearity and loudness,'' J. Acoust. Soc. Am. **103**, 2010–2020.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Shailer, M. J., and Moore, B. C. J. (**1987**). ''Gap detection and the auditory filter: phase effects using sinusoidal stimuli,'' J. Acoust. Soc. Am. **81**, 1110–1117.
Terry, M., and Moore, B. C. J. (**1977**). '''Suppression' effects in forward masking,'' J. Acoust. Soc. Am. **62**, 781–784.
Widin, G. P., and Viemeister, N. F. (**1979**). ''Intensive and temporal effects in pure-tone forward masking,'' J. Acoust. Soc. Am. **66**, 388–395.
Wojtczak, M., Schroder, A. C., Kong, Y.-Y., and Nelson, D. A. (**2001**). ''The effect of basilar-membrane nonlinearity on the shapes of masking period patterns in normal and impaired hearing,'' J. Acoust. Soc. Am. **109**, 1571–1586.
Zwislocki, J., Pirodda, E., and Rubin, H. (**1959**). ''On some poststimulatory effects at threshold of audibility,'' J. Acoust. Soc. Am. **31**, 9–14.

Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners

Peggy B. Nelson[a)]
Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

Su-Hyun Jin
Department of Communication Disorders, Department of Otolaryngology, University of Minnesota, Minneapolis, Minnesota 55455

Arlene Earley Carney
Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

David A. Nelson
Department of Otolaryngology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 15 March 2002; accepted for publication 4 November 2002)

Many competing noises in real environments are modulated or fluctuating in level. Listeners with normal hearing are able to take advantage of temporal gaps in fluctuating maskers. Listeners with sensorineural hearing loss show less benefit from modulated maskers. Cochlear implant users may be more adversely affected by modulated maskers because of their limited spectral resolution and by their reliance on envelope-based signal-processing strategies of implant processors. The current study evaluated cochlear implant users' ability to understand sentences in the presence of modulated speech-shaped noise. Normal-hearing listeners served as a comparison group. Listeners repeated IEEE sentences in quiet, steady noise, and modulated noise maskers. Maskers were presented at varying signal-to-noise ratios (SNRs) at six modulation rates varying from 1 to 32 Hz. Results suggested that normal-hearing listeners obtain significant release from masking from modulated maskers, especially at 8-Hz masker modulation frequency. In contrast, cochlear implant users experience very little release from masking from modulated maskers. The data suggest, in fact, that they may show negative effects of modulated maskers at syllabic modulation rates (2–4 Hz). Similar patterns of results were obtained from implant listeners using three different devices with different speech-processor strategies. The lack of release from masking occurs in implant listeners independent of their device characteristics, and may be attributable to the nature of implant processing strategies and/or the lack of spectral detail in processed stimuli. [DOI: 10.1121/1.1531983]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ts, 43.64.Me [CWT]

I. INTRODUCTION

Many natural background noises are temporally fluctuating, such as clattering dishes or background conversations. Listeners with normal hearing sensitivity take advantage of gaps in these fluctuating or modulated maskers. They are able to ''listen in the dips'' of the modulated masker to extract information about the speech signal. These extracted pieces of the message, then, are often sufficient to provide full understanding of the message. This improvement in speech recognition provided by modulated maskers compared to steady maskers is referred to as a ''release from masking.'' The amount of release from masking in normal-hearing listeners ranges in published reports from less than 5 dB to as much as 20 dB, depending on the stimuli and the temporal characteristics of the maskers (e.g., Bacon *et al.*, 1998). For most speech stimuli, the optimal masker modulation rates for observing masking release fall between 10 and 32 Hz (e.g., Gustafson and Arlinger, 1994). At slower modulation rates, whole syllables or words may occasionally be masked by a cycle of noise. At faster modulation rates, forward masking may perceptually fill the nominal silent interval, resulting in performance similar to that of a continuous masker.

Listeners with hearing loss are less able than normal listeners to obtain release from modulated maskers (e.g., Festen and Plomp, 1990; Takahashi and Bacon, 1992; Eisenberg *et al.*, 1995; Bacon *et al.*, 1998). Eisenberg and colleagues tested listeners with normal hearing and listeners with hearing loss for their understanding of consonants in steady and fluctuating noise. Listeners with normal hearing were tested with shaped noise designed to simulate the hearing sensitivity of the impaired listeners. Their results suggested that listeners with true hearing loss obtained far less release from modulated maskers than did normal-hearing listeners with or without simulated hearing losses. Amplification restored some, but not all, of the expected release from masking for impaired listeners. Eisenberg and colleagues concluded that audibility alone cannot explain the additional masking experienced by listeners with sensorineural hearing loss.

In contrast, Trine (1995) hypothesized that the primary problem for listeners with hearing loss was, in fact, reduced

[a)]Electronic mail: nelso477@umn.edu

audibility of signals that occurs in the dips of the fluctuating maskers. He noted a high negative correlation between masking release and the degree of hearing loss of his impaired listeners. He also noted that when amplification was provided to the impaired listeners, especially in the high-frequency region, the amount of release from masking increased, and approached that obtained by normal-hearing listeners. He postulated that if it were possible to amplify all signals, such that the temporal dips in the modulated maskers resulted in full audibility of the signals during that cycle, then impaired listeners might obtain normal release from masking.

Subsequently, Bacon *et al.* (1998) reported that some listeners with hearing loss obtained less release from temporally fluctuating maskers than did normal-hearing listeners with and without simulated hearing loss. They evaluated listeners' understanding of sentences in speech-shaped noise that was modulated by the envelope of one of the following: steady-state noise, multitalker babble, single-talker babble, and a 10-Hz square wave with 100% modulation depth. They observed that for normal-hearing listeners, the square-wave modulation provided the greatest release from masking. In addition, they found that the impaired listeners obtained significantly less release from masking than did their normal-hearing counterparts. Noise-masked normal-hearing listeners obtained somewhat less masking release than they had with full access to the signals. However, six of the 11 impaired listeners obtained significantly less release from masking than did their counterparts with simulated hearing loss. They concluded that audibility accounts for some loss of masking release, but additional factors, such as excessive forward masking in impaired ears, may account for the additional loss of masking release.

More recently, Dubno, Horowitz, and Ahlstrom (2002) suggested that audibility explained only a small percentage of the variability in older and younger listeners' identification of consonants in modulated noise. Older and younger listeners with normal or near-normal hearing sensitivity were matched for their thresholds using threshold-matching noise. Significant differences in consonant identification between groups were found. They further noted a significant correlation between forward masking and masking release in older listeners with near-normal hearing sensitivity, suggesting that factors other than audibility can affect masking release.

Kwon and Turner (2001) investigated consonant identification in normal-hearing listeners' understanding of implant simulations. They suggested that two opposing factors may influence hearing-impaired listeners' understanding of speech in modulated noise. First, these listeners may benefit from the same release from masking that is observed in normal-hearing subjects. As a result, their performance may improve when noise is modulated rather than constant. Second, listeners with hearing loss may be negatively affected by modulation masking because listeners with reduced spectral resolution rely on natural amplitude modulations for speech recognition (e.g., Hedrick and Jesteadt, 1996; Hedrick and Carney, 1997). The modulated noise may actually interfere with the acoustic envelope cues, at syllabic or segmental levels, that are used by the listener with hearing loss. Kwon and Turner (2001) evaluated the effects of modulated noise on understanding spectrally impoverished signals (12-band noise simulations). Their listeners apparently experienced a mix of masking release and modulation masking. When the signals and/or the maskers were bandlimited, they found that midfrequency modulated maskers provided the listeners some masker release, resulting in improved consonant recognition when compared to unmodulated maskers. In contrast, a modulated high-frequency masker sometimes caused reduced consonant identification when compared to an unmodulated masker. They concluded that high-frequency modulated maskers can cause some interference in consonant recognition that may offset any benefit provided by the masking release.

Listeners with cochlear implants have well-documented difficulties understanding speech in steady noise (e.g., Fu *et al.*, 1998). Most realistic noise, however, is fluctuating in nature, and a listeners' ability to ''listen in the dips'' is important for communication in these realistic environments. It is not known whether listeners with cochlear implants obtain release from masking when listening in fluctuating noise. If Kwon and Turner's (2001) hypothesis is true, that modulated maskers can cause both masking release and modulation masking (interference), then listeners with cochlear implants may not benefit from masker temporal fluctuations. Instead, implant listeners who use speech processors with envelope-extracting processor algorithms may be adversely affected by fluctuating maskers like individual competing talkers. A lack of masking release might cause additional difficulty in day-to-day situations.

The current experiment was designed to evaluate the ability of cochlear implant listeners to take advantage of temporal gaps in background noise. Listeners with implants were compared to listeners with normal hearing sensitivity for the understanding of sentences in background noise, when the noise was either steady or square-wave modulated across a range of modulation frequencies.

II. METHODS

A. Subjects

Subjects were eight young adult listeners with normal hearing sensitivity who listened to typical full-spectrum speech (normal group), eight additional young adult listeners with normal hearing sensitivity who listened to implant simulations (simulation group), and nine adult listeners with hearing loss who were cochlear implant users (implant group). Characteristics of listeners in the implant group are shown in Table I. All implant users were postlingually deafened. Their mean age was 49 years (range: 34 to 64 years), and their average length of deafness prior to implantation was 16 years (range 1 to 44 years). All listeners had worn their implants for more than 2 years (mean: 5 years, range 2 to 11 years) and derived significant benefit from their devices. As shown in Table I, three listeners used the Nucleus 22 device with a spectral-peak (SPEAK) speech-processing strategy, three used the Clarion 1.2 device with a continuous interleaved sampling (CIS) strategy, and three used the Clarion HiFocus device with a CIS strategy. Listeners in the

TABLE I. Summary of subject characteristics.

Listener	CI/Processor	Age at test	Age at onset of deafness	Age at implantation
N12	Nucleus 22/SPEAK	53	32	42
N14	Nucleus 22/SPEAK	58	49	50
N32	Nucleus 22/SPEAK	34	5	29
C02	Clarion 1.2/CIS	42	18	37
C03	Clarion 1.2/CIS	53	22	49
C05	Clarion 1.2/CIS	47	42	43
C14	Clarion HiFocus/CIS	64	16	60
C15	Clarion HiFocus/CIS	42	33	40
C16	Clarion HiFocus/CIS	48	29	43

implant group used their own speech processors with typical sensitivity and volume settings, and no noise reduction. At the beginning of each session, the users set the sensitivity and/or volume controls while listening to practice lists, and they were instructed not to change the settings during the test session.

B. Stimuli

Speech stimuli consisted of IEEE (1969) sentence materials spoken by five male and five female talkers. Stimuli were recorded on digital audio tape at 44 kHz. They were digitized, downsampled to 20k samples per second, and normalized for long-term rms amplitude using COOLEDIT PRO©. Sentences contained an average of five key words. Blocks of ten sentences were presented, each block containing one sentence spoken by each talker, in random order.

Noise stimuli were generated in real time using the Tucker-Davis waveform generator (TDT WG1). The noise was passed through a Rane 30-band equalizer so that the spectrum of the resulting noise matched the long-term spectrum of the IEEE sentences. Noise stimuli were presented either continuously (steady), or gated with 2-ms cosine-squared ramps. Gating was implemented with 50% duty cycles and 100% modulation depths. Six gate frequencies ranged from 1 to 32 Hz, resulting in noise bursts that ranged in duration from 16 ms (32 Hz) to 500 ms (1 Hz). Signal-to-noise ratios (SNRs) were computed based on the long-term rms of the noise and the speech. SNRs were +16, +8, 0, −8, or −16 dB, depending upon the listener and the condition.

Sentences from two talkers were modified to create four-channel simulations of implant processing. Sentences were filtered into four narrow bands (after Shannon *et al.*, 1995): 100–300, 300–500, 500–1700, and 1700–6000 Hz. The envelope of each filter output was extracted and narrow-band noises of the same frequency region were modulated by the respective envelope (low-pass filtered at 500 Hz).

C. Test procedures

Listeners were seated in the center of a sound-treated chamber. Speech signals were delivered diotically through two Bose 301 speakers at an overall level of 65 dBA. Speech stimuli were presented in blocks of ten sentences, using all ten talkers in random order for each list. All SNR and gating conditions were randomized prior to the beginning of each subject's testing. The listeners responded verbally to each sentence, and the experimenter scored the key words correct for each sentence, circling the correct answers on an answer form. Each listener's results (percent-correct key words) for each condition were later entered into computer files.

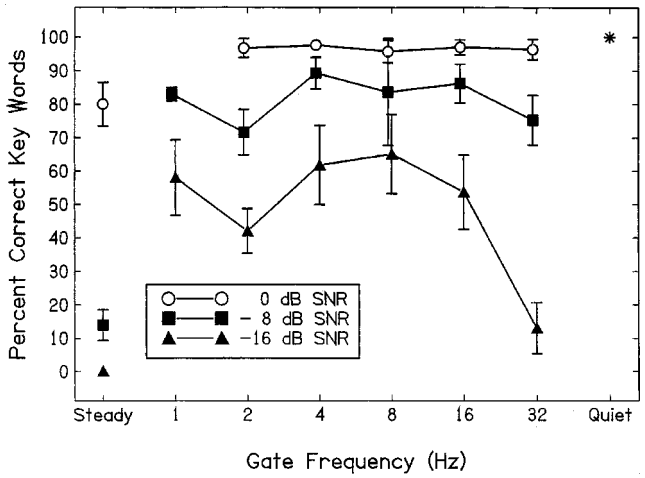


FIG. 1. Average percent-correct key word identifications are shown as a function of noise gate frequency for normal-hearing listeners at SNRs of 0, −8, and −16 dB. Error bars indicate one standard deviation from the mean.

Key word identification was evaluated in steady and gated noise. On each trial, the masking noise started initially, with the sentence beginning after a random delay that ranged from 10 to 100 ms. The noise was either steady or gated. The level of the noise varied depending upon the condition being tested. Listeners in the implant and simulation groups heard the noise at +8 and +16 dB SNR. (Listeners in the simulation group also heard the noise at 0 dB SNR. Pilot testing with three high-performing implant listeners indicated that performance was near 0% for all gate conditions at 0 dB SNR and lower.) Listeners in the normal group heard the noise at 0, −8, and −16 dB SNR. All listeners also completed two blocks of sentences in quiet.

III. RESULTS

A. Normal group

Results from listeners with normal hearing sensitivity are shown in Fig. 1. These listeners were able to repeat nearly 100% of the key words in quiet. When the SNR was 0 dB, they obtained scores of approximately 80% correct for steady noise, and near 100% for all gated noise conditions. When the SNR was −8 dB, their performance in steady noise was only 10% correct key words, while in gated noise their mean performance ranged from 70% to 90% correct. When the SNR was −16 dB, they scored 0% correct in steady noise, with average gated noise performance ranging between 15% and 65% correct. Performance was dependent upon the gate frequency. Although there was considerable variability among listeners in the normal group, release from masking was maximal for gate frequencies between 1 and 16 Hz, and was reduced at gate frequencies at or above 16 Hz.

Figure 2 shows the normal group's masking release, or improvement in scores for gated vs steady noise, for the three SNR conditions. For SNR of 0 dB, improvement from gating was approximately 20% for all gate frequencies and

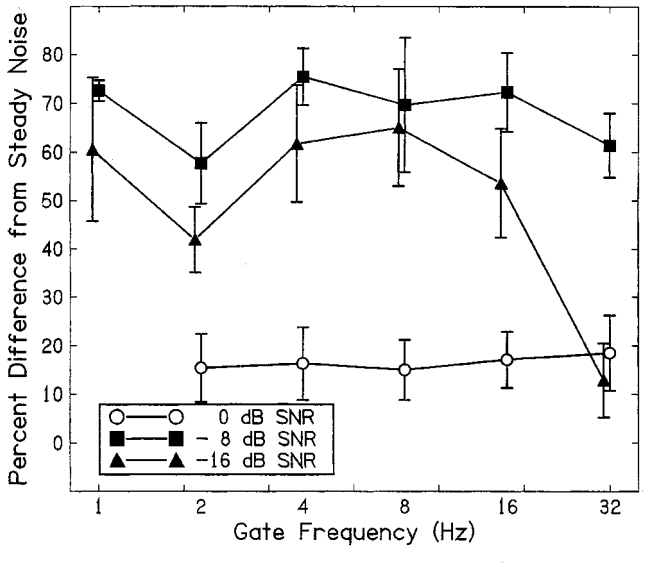


FIG. 2. Average percent improvement from steady noise is shown as a function of noise gate frequency for normal-hearing listeners, at SNRs of 0, −8, and −16 dB. The amount of release from masking is especially large for the SNR of −8 dB.

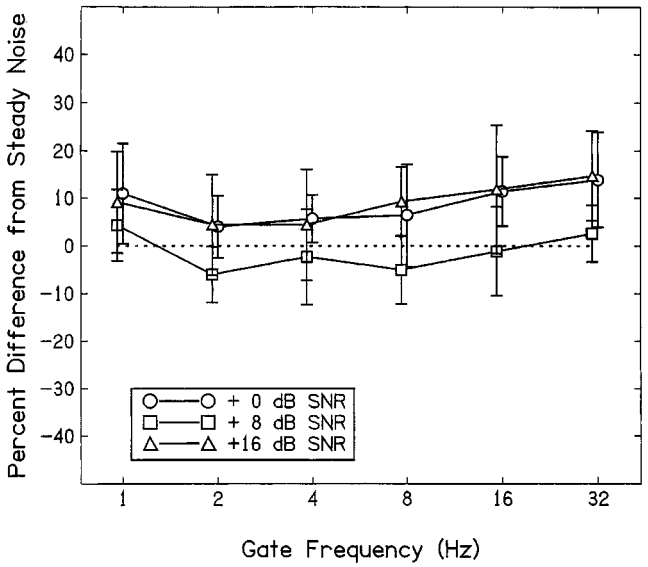


FIG. 4. Average percent improvement from steady noise is shown as a function of noise gate frequency for simulation listeners, at SNRs of 0, +8, and +16 dB. Little release from masking is seen for simulation listeners.

was limited by a ceiling effect for the gated conditions (see Fig. 1). The maximum release from masking occurred at the SNR of −8 dB, with improvement ranging from 60%–80%. Masking release at −8 dB SNR was relatively independent of gate frequency, with a possible minimum at 2 Hz. For −16 dB SNR, release from masking ranged from 10%–60% and was strongly affected by gate frequency. Normal-hearing listeners' release showed the same apparent minimum at 2 Hz and was reduced for very fast (32 Hz) gate frequencies.

B. Simulation group

Results from listeners in the simulation group are shown in Fig. 3. The stimuli for these listeners were 4-band modulated noise replicas of the IEEE sentences from one talker. Their mean key word identification score in quiet was approximately 55%, indicating that these listeners showed performance somewhat typical of implant listeners. Mean scores in steady noise dropped to approximately 30% correct for the SNR of +16 dB, 25% for SNR of +8 dB, and 5% correct for SNR of 0 dB, suggesting that all levels of noise had a significant negative effect on word understanding. When the noise was gated, mean scores for an SNR of +16 dB were near 40% correct for all gate frequencies, still poorer than the mean score correct in quiet and only slightly better than their performance in steady noise. Mean scores in gated noise at +8 and 0 dB SNR showed a similar pattern; scores in gated noise were very close to those in steady noise and were independent of gate frequency.

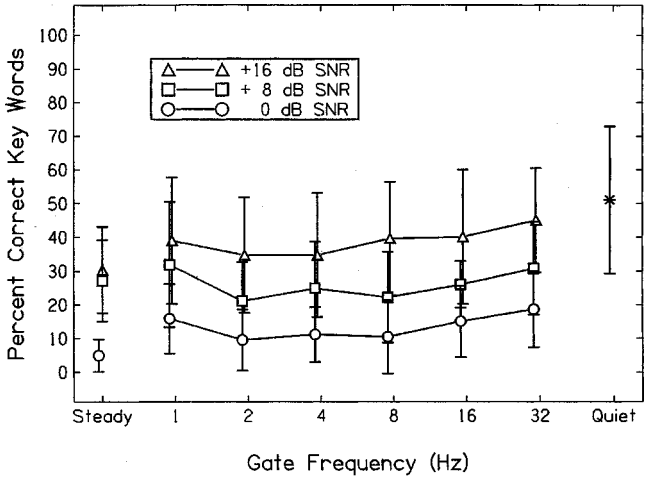


FIG. 3. Average percent-correct key word identification is shown as a function of noise gate frequency for simulation group normal listeners for SNRs of 0, +8, and +16 dB.

The data from the simulation group are replotted in Fig. 4 showing release from masking, or percent improvement in scores for gated versus steady noise for their three SNR conditions. For conditions with an SNR of +16 and 0 dB, the gated noise provided a slight benefit over the steady noise, except perhaps for the fastest gate rates. For a +8-dB SNR, no masking release was observed. No effect of gate frequency was seen. An analysis of variance (ANOVA) indicated that there was a significant effect of SNR [$F(1,10) = 18.91$, $p < 0.01$], but no significant effect of gate frequency [$F(1,10) = 1.43$, $p > 0.05$].

C. Implant group

Results from listeners in the implant group are shown in Fig. 5. Their mean key word identification score in quiet was 80%, indicating that these listeners were successful implant users. Mean scores in steady noise dropped to 60% correct for an SNR of +16 dB, and to 35% correct for an SNR of +8 dB, suggesting that both levels of noise had a significant negative effect on word understanding. When the noise was gated, mean scores for an SNR of +16 dB ranged from 55% to 65% correct, still significantly poorer than the mean 80% correct in quiet and not different from their performance in steady noise.

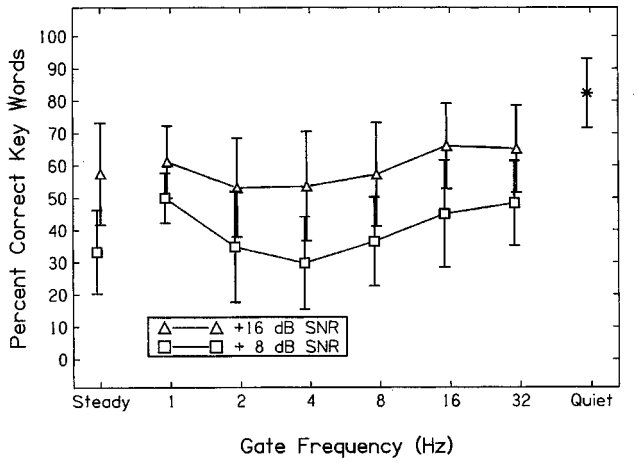


FIG. 5. Average percent-correct key word identification is shown as a function of noise gate frequency for listeners with cochlear implants for SNRs of +8 and +16 dB.

The data from the implant group are replotted in Fig. 6 showing release from masking, or percent improvement in scores for gated versus steady noise for their two SNR conditions. For conditions with an SNR of +16 dB, the gated noise provided little benefit over the steady noise, except perhaps for the fastest gate rates. For the +8-dB SNR condition, performance was slightly better for the slowest and fastest gate frequencies than for the steady noise, with a minimum in masking release seen at 2-, 4-, and 8-Hz gate frequencies.

The amount of masking release obtained by the different listener groups was compared using analysis of variance. To determine whether normal listeners had significantly more masking release than the other groups, results had to be compared at different SNRs because normal listeners were tested at SNRs that were different from the other two groups. Masking release results (pooled across gate frequencies between 2 and 16 Hz) were compared at −8-dB SNR for the normal group, +8-dB SNR for the simulation group, and +8-dB SNR for the implant group. Results indicated that normal listeners obtained significantly more masking release than did implant and simulation listeners [$F(2,18)=46.4$, $p<0.0001$]. More detailed comparisons were possible between simulation and implant groups. Repeated measures analysis of variance was applied to the masking release data for the two groups (simulation and implant), two SNRs (+8 and +16 dB), and six gate frequencies. No significant difference between groups was noted [$F(1,12)=0.3$, $p>0.05$], suggesting that the implant and simulation listeners had similar release from masking. A significant effect of SNR was noted [$F(1,12)=6.1$, $p<0.05$] with no significant group by SNR interaction [$F(1,12)=4.3$, $p>0.05$]. A significant effect of gate frequency was also noted [$F(4,48)=7.6$, $p<0.01$]; however, there was a significant gate frequency by group interaction [$F(4,48)=3.6$, $p<0.05$]. No higher-order interactions were significant.

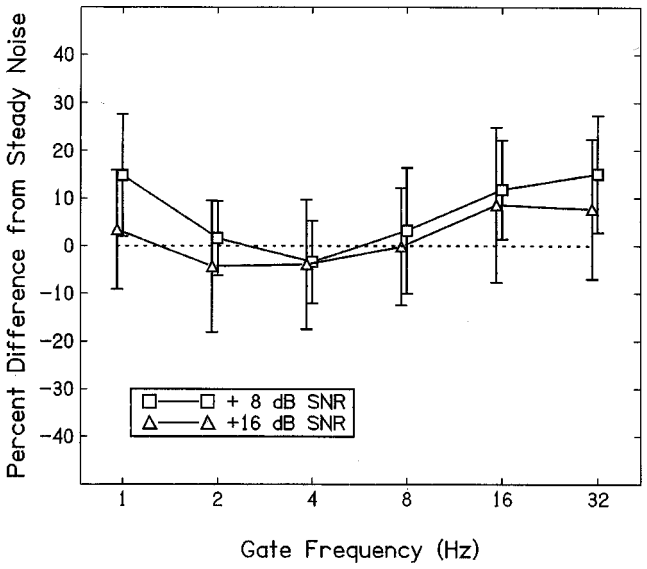


FIG. 6. Average percent improvement from steady noise is shown as a function of noise gate frequency for listeners with cochlear implants, at SNRs of +8 and +16 dB. Little release from masking is seen for cochlear implant users.

As noted in the previous section, for the simulation group, no significant effect of gate frequency was found. Analysis of the implant group indicated that the effect of gate frequency on masking release approached, but did not reach significance [$F(1,10)=4.68$, $p=0.056$] across both SNRs. Multiple regression analysis indicated that gate frequency accounted for 32% of the overall variance in implant listeners' performance, while SNR accounted for 16%. Both gate frequency and SNR accounted for 47% of the variance in implant listeners' masking release. When the gated noise was presented at +8-dB SNR, mean scores ranged from 35% to 50%. Some improvement over steady noise was seen at the slowest (1 Hz) and fastest (16 and 32 Hz) gate frequencies, but performance remained low at moderate gate frequencies (2 to 8 Hz). Paired *t*-tests for the data from the +8-dB SNR condition indicated that mean performance in 1-Hz gated noise (500-ms alternating cycles of noise and silence) was significantly better than performance in steady noise ($t[6]=-2.72$, $p=0.017$). Performance in 16-Hz ($t[7]=-3.7$, df=7, $p=0.0037$) and 32-Hz ($t[7]=-3.26$, $p=0.007$) gated noises were significantly better than performance in steady noise. When corrected for multiple comparisons, these individual comparisons retain their significance.

Figure 7 shows mean data for the subgroups of implant listeners with different devices. This figure shows the improvement in performance for listeners divided by implant type for gated vs steady noise at an SNR of +8 dB. Clearly, although there were overall performance differences between listeners, the trend was that all listeners obtained minimal to no benefit of gated noise over steady noise for all devices, with an apparent minimum in performance at gate frequencies around 4 Hz. This suggests that specific characteristics of a given implant device (Nucleus 22 vs Clarion 1.2) or speech processing strategy (SPEAK vs CIS) were not primarily responsible for implant listeners' failure to demonstrate release from masking. Examination of individual data functions revealed that only one implant listener did *not* show the characteristic minimum performance near the 4-Hz gate frequency. That listener showed a relatively flat performance function for 1–8-Hz gate frequencies, with increased masking release at 16 to 32-Hz gate frequencies. All other implant

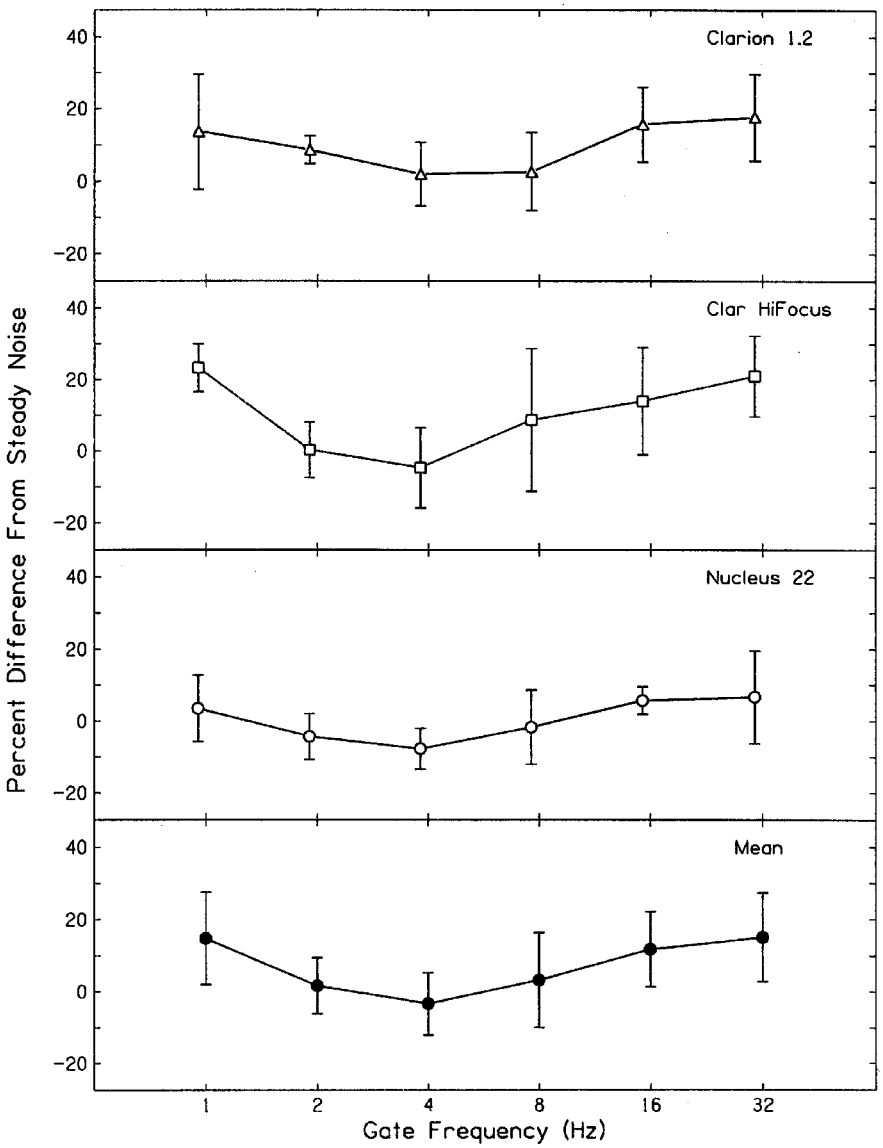


FIG. 7. Average percent improvement from steady noise is shown as a function of noise gate frequency for listeners with different implant devices. Users with Clarion 1.2 processors using CIS strategy are shown with open triangles; users with Clarion Hi-Focus processors are shown with open squares; users with Nucleus N22 processors using the SPEAK strategy are shown with open circle symbols. The overall mean performance for all implant users is shown with filled circles. No meaningful differences between processor types can be seen.

listeners showed a minimum in the performance function at 2 or 4 Hz, with improved performance at slower and faster gate frequencies.

IV. GENERAL DISCUSSION

A. Normal group

As expected, listeners in the normal group obtained significant release from masking from the gated maskers. The greatest amount of masker release was obtained for SNRs of −8 dB, at which the speech signals were an approximate 65 dB A and the noise was 73 dB A. At those levels, the words were very difficult to hear in steady noise, and mean performance was approximately 10% correct. When the noise was gated, however, performance improved considerably to mean levels of approximately 80% correct for gate frequencies of 4 Hz or higher. Presumably, some minimal amount of speech information was audible in the presence of the steady noise because performance was better than chance (mean scores were approximately 10% correct). When parts of the signals were made fully audible during silent intervals in the gated noise, performance improved considerably. Clearly, listeners were using bits of information to fill in the message, and as a result they were understanding a majority of the key words. The IEEE sentences that were used in this investigation have been shown to have relatively small linguistic context effects (Nittrouer and Boothroyd, 1990). Nevertheless, the partial acoustic and linguistic cues obtained by the normal group were used to understand most of the key words.

The results were somewhat different for the normal group at an SNR of −16 dB, when the speech and noise signals were at 65 and 81 dBA, respectively. At these levels, none of the key words was identifiable in steady noise. Gated noise maskers again provided listeners with significant masker release, but in this case, the amount of release was related to the masker's gate frequency. At the slowest gate frequencies (1–2 Hz, corresponding to alternating 500- or 250-ms periods of noise and silence) approximately 40%–50% of the key words were identified. In this condition, whole words and syllables were presumably completely inaudible, and listeners were unable to extract more than 50% of the information. However, at 4- and 8-Hz gate frequencies (125- and 62-ms periods), parts of many syllables and words were probably audible. Listeners used these parts to identify

approximately 60% of the key words. At faster gate frequencies, the periods of silence were only approximately 30 and 15 ms in duration. Because the noise levels significantly exceeded the level of the speech signals, we presume that some forward masking occurred, at least partially obscuring the speech during these short silent intervals. At the gate frequency of 32 Hz, performance was greatly reduced to a mean score of 15%.

B. Implant and simulation groups

Listeners with cochlear implants and normal-hearing listeners responding to implant simulations were much more affected by background noise than were normal-hearing listeners. Initial pilot results had suggested that none of the best implant users could understand any key words at a 0-dB SNR for either gated or continuous noise. As a result, implant and simulation listeners were tested at SNRs that were different from those used with the normal group. Even though the implant listeners demonstrated very good performance in quiet (around 80% for difficult stimuli), they were greatly affected by noise. Even at the favorable SNR of +16 dB, performance dropped by more than 20%. At an SNR of +8 dB, an SNR typical of many environmental situations, key word identification dropped by about 50%. These results indicate that when context is low and speakers unfamiliar, even low levels of background noise affect implant listeners substantially.

Noise affected simulation listeners in a similar way. Simulation group listeners understood approximately 55% of words in quiet, somewhat poorer than the implant group results, but typical of some implant listener performance. At the favorable SNR of +16 dB, their performance also dropped by more than 20%, indicating a significant effect of the steady background noise.

It seems likely that noise (modulated or steady) disrupts the ideal amplitude envelope cues that are coded by the implant processors. Even when the noise occurred 16 dB below the speech signal, one can imagine that the random envelope of the noise could disrupt natural envelope cues extracted by the implant processor. This may explain the significant drop in performance from quiet to steady noise, even at an SNR of +16 dB. However, this does not explain a lack of the ability to use intervals of quiet speech within gated noise to extract some key words.

Interestingly, the implant and simulation group listeners did not show significant masking release from temporal gaps in noise. Simulation group listeners showed very little masking release (10%) for SNRs of +16 and 0 dB, and no masking release for SNRs of +8 dB. No effect of gate frequency was seen, suggesting that listeners responding to four-channel implant simulations do not take advantage of temporal gaps in noise, even when that gap is as long as 500 ms. For implant listeners at an SNR of +16 dB, there was no difference in performance between steady and gated noises at any gate frequency. For an SNR of +8 dB (a condition quite typical of conversational settings), there seems to be some slight release from masking at extremely slow (1 Hz) and fast (16 and 32 Hz) modulation rates, with a minimum in performance between 2 and 8 Hz.

We do not attribute the lack of masking release to either a lack of audibility in the ''dips'' nor to forward masking. At the SNRs used by the implant and simulation groups the signal level greatly exceeded the level of the maskers. Implant listeners set the sensitivity of their devices so that the quiet sentences were at a comfortable and audible level. When noise was introduced, it was always at a level 8 or 16 dB below the level of the speech. Also, because the simulation group showed a lack of masking release similar to that of the implant group, inaudibility cannot be the primary cause. Clearly those normal-hearing listeners had full access to the sentence information in the temporal dips in noise. Thus, we do not expect that an inability to repeat key words was due to a lack of audibility of the quiet stimuli. Similarly, because of the low-level noise we did not expect, nor did we see, any decrement in performance at the fastest gate frequencies (like that observed at −16 dB SNR and 32 Hz for the normal group) that might be attributed to forward masking.

We had presumed, however, that 250-ms silent intervals (the 2-Hz gating condition) would be sufficient for at least some implant listeners to identify some key words. Because initial pilot data had shown no masker release even at 2 Hz, the 1-Hz condition was added. Based on the results for the 1-Hz condition, it seems that most implant listeners were able to take advantage of 500-ms silent intervals to identify some key words, at least for the 8-dB SNR condition. Eight of nine individual implant users showed some release from masking at 1-Hz gate frequency. It was surprising that one remaining implant listener and all simulation listeners did not show significant word understanding with silent intervals as long as 500 ms in the noise. None of the implant or simulation group listeners could take advantage of 250-ms silent intervals to identify at least some key words. In fact, performance was the same for steady noise and for maskers with 2-, 4-, and 8-Hz gate frequencies.

One logical explanation for this effect is that gated maskers at those syllabic-like rates were actually a distraction or interference, rather than a benefit to the implant listener. In fact, some implant group users reported anecdotally that the gated noise mixed with the sentences sounded like additional syllables, perhaps in another language. When the gated noise was presented alone, one listener described the noise appropriately as bursts of noise at slow modulation rates, and as ''fluttering'' noise at faster rates. This confirmed that gaps in the noise were perceived by the listeners. However, when the noise was mixed with the speech at moderate modulation rates, he reported that he heard it as a strange competing talker. This would support the Kwon and Turner (2001) hypothesis that gated maskers can provide some release (seen here at 1-, 16-, and 32-Hz gating for 8-dB SNR) as well as some interference (seen here at 2-, 4-, and 8-Hz gating for 8-dB SNR maskers).

There seems to be no significant performance difference between users of different implant devices or speech-processing algorithms, at least among the pulsatile strategies evaluated here (CIS and SPEAK). Also, there was very little difference in performance between the implant and simulation groups. Thus, the specific processing characteristics of

the implant devices such as the processing algorithm, the number of electrodes stimulated, the automatic gain control, or the range of acoustic amplitudes encoded (input dynamic range), do not seem to account for the lack of masking release. Listeners' performance was not apparently restricted by implant processing hardware. The implant processor was providing them with the temporal envelope information at sufficiently high (at least 250 Hz) rate (Kwon, 2002).

Listeners with both devices (Clarion and Nucleus) showed the minimum in performance at gate frequencies between 2 and 8 Hz. The lack of release from masking is apparently, then, not related to characteristics of the implant devices themselves.

In addition, it seems unlikely that these results can be explained on the basis of abnormal forward masking of the implant users. Previous studies (e.g., Nelson and Donaldson, 2001) have suggested that most cochlear implant users demonstrate rapid-recovery time constants of less than 7 ms. That rapid recovery should allow implant users to take advantage of temporal gaps in the noise that were as long and 250 and 500 ms in some conditions. Performance functions from the implant group (seen in Figs. 3 and 4) do not show the characteristic shape seen in the data from the normal group listeners (Figs. 1 and 2). While listeners from the normal group show decreased release from masking as modulation rate increases, the listeners from the implant group do not. The decrease in benefit from modulation noise at rapid modulation rates (temporal gaps $<$30 ms) can be attributed to forward masking perceptually filling the gaps for the normal-hearing listeners. Forward masking, then, cannot explain the relatively flat functions of the implant group.

It may be more likely that the implant and simulation listeners receive such an impoverished spectral code that they are unable to integrate the speech information into a well-defined auditory image, or to segregate the speech signal from the background noise. Because of the limited acoustic cues available to the listener, a longer temporal ''glimpse'' of the quiet signal (longer than 500 ms) is needed before the speech stream can be integrated and whole words extracted. Additional testing of auditory stream segregation by implant users is warranted and is underway. In addition, further evaluation of the role of spectral resolution (increased numbers of spectral channels) in masking release is underway and will be reported in a companion paper.

These results suggest that listeners with cochlear implants are likely to be extremely disrupted in acoustic situations with a single competing talker, even when the level of the talker's voice is significantly less than the target signal. Similarly, they may be quite affected by a reverberant room where the envelope cues are disrupted by echoes. Further study is needed to explain and understand these results.

V. CONCLUSIONS

Although normal-hearing listeners are able to obtain release from masking from modulated noise, listeners with cochlear implants cannot. Implant and simulation listeners are significantly affected by background noise, even at very favorable signal-to-noise ratios. When noise is modulated, even with 250-ms silent intervals, implant and simulation listeners are unable to take advantage of a silent gap to extract meaningful words. Performance of implant users seems poorest at modulation frequencies between 2 and 8 Hz, encompassing rates corresponding to syllables and words. These results imply that modulation interference, or masking, may be responsible for the lack of masking release in the implant group listeners. Performance does not seem to vary with implant device or processing strategy, and may be due to a disruption in the envelope cues extracted by the devices and used by the listeners. The lack of masking release, then, may be attributable to general characteristics of the implant processing, including the lack of spectral information in the processed signal. Implant listeners may have noticeable difficulty in situations with fluctuating noise, such as in restaurants or with single competing talkers.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. DC00110, by the Lions 5M International Hearing Foundation, and by the University of Minnesota. The authors wish to thank Gail Donaldson and Benjamin Munson for their assistance on early versions of this manuscript.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). ''The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds,'' J. Speech Hear. Res. **41**, 549–563.

Dubno, J., Horwitz, A., and Ahlstrom, J. (**2002**). ''Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing,'' J. Acoust. Soc. Am. **111**, 2897–2907.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). ''Speech recognition in amplitude modulated noise of listeners with normal and impaired hearing,'' J. Speech Hear. Res. **38**, 222–233.

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuation noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). ''Masking of speech by amplitude-modulated noise,'' J. Acoust. Soc. Am. **95**, 518–529.

Hedrick, M. S., and Carney, A. E. (**1997**). ''Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants,'' J. Speech Lang. Hear. Res. **40**, 1445–1457.

Hedrick, M. S., and Jesteadt, W. (**1996**). ''Effect of relative amplitude, presentation level, and vowel duration on perception of voiceless stop consonants by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **100**, 3398–3407.

IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' IEEE Trans. Audio Electroacoust. **17**(3), 225–246.

Kwon, B. J. (**2002**). Personal communication.

Kwon, B. J., and Turner, C. W. (**2001**). ''Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?,'' J. Acoust. Soc. Am. **110**, 1130–1140.

Nelson, D. A., and Donaldson, G. S. (**2001**). ''Psychophysical recovery from single-pulse forward masking in electric hearing,'' J. Acoust. Soc. Am. **109**, 2921–2933.

Nittrouer, S., and Boothroyd, A. (**1990**). ''Context effects in phoneme and word recognition by young children and older adults,'' J. Acoust. Soc. Am. **87**, 2705–2715.

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

Takahashi, G. A., and Bacon, S. P. (**1992**). ''Modulation detection, modulation masking, and speech understanding in noise in the elderly,'' J. Speech Hear. Res. **35**, 1410–1421.

Trine, T. D. (**1995**). ''Speech recognition in modulated noise and temporal resolution: Effects of listening bandwidth,'' Unpublished doctoral dissertation, University of Minnesota, Twin Cities.

Cochlear toughening, protection, and potentiation of noise-induced trauma by non-Gaussian noise

Roger P. Hamernik,[a] Wei Qiu, and Bob Davis
Auditory Research Laboratory, State University of New York, 107 Beaumont Hall, Plattsburgh, New York 12901

(Received 28 June 2002; revised 31 October 2002; accepted 4 November 2002)

An interrupted noise exposure of sufficient intensity, presented on a daily repeating cycle, produces a threshold shift (TS) following the first day of exposure. TSs measured on subsequent days of the exposure sequence have been shown to decrease relative to the initial TS. This reduction of TS, despite the continuing daily exposure regime, has been called a cochlear toughening effect and the exposures referred to as toughening exposures. Four groups of chinchillas were exposed to one of four different noises presented on an interrupted (6 h/day for 20 days) or noninterrupted (24 h/day for 5 days) schedule. The exposures had equivalent total energy, an overall level of 100 dB(A) SPL, and approximately the same flat, broadband long-term spectrum. The noises differed primarily in their temporal structures; two were Gaussian and two were non-Gausssian, nonstationary. Brainstem auditory evoked potentials were used to estimate hearing thresholds and surface preparation histology was used to determine sensory cell loss. The experimental results presented here show that: (1) Exposures to interrupted high-level, non-Gaussian signals produce a toughening effect comparable to that produced by an equivalent interrupted Gaussian noise. (2) Toughening, whether produced by Gaussian or non-Gaussian noise, results in reduced trauma compared to the equivalent uninterrupted noise, and (3) that both continuous and interrupted non-Gaussian exposures produce more trauma than do energy and spectrally equivalent Gaussian noises. Over the course of the 20-day exposure, the pattern of TS following each day's exposure could exhibit a variety of configurations. These results do not support the equal energy hypothesis as a unifying principal for estimating the potential of a noise exposure to produce hearing loss. [DOI: 10.1121/1.1531981]

PACS numbers: 43.66.Ed, 43.50.Pn, 43.50.Qp [NFV]

I. INTRODUCTION

The role of temporal variables in a noise exposure paradigm has taken on a new level of interest since (a) Miller *et al.* (1963) and Clark *et al.* (1987) showed that threshold shifts (TS) following a noise exposure could be modulated by a cyclic pattern of exposure and (b) the discovery of the motor process associated with the outer hair cell (OHC) system (see Brownell, 1990 for a review). The reduction in TS following daily repeated exposures to the same noise has been called toughening (TS_R). That is, threshold shifts following each day's exposure are reduced despite the continuing exposure. Toughening, while dependent on the level and spectrum of the stimulus Subramaniam *et al.*, 1991), has been shown to occur as a result of interrupted exposures to continuous octave bands of noise (Boettcher *et al.*, 1992; Subramaniam *et al.*, 1992) as well as from broadband impact noise exposures with peak SPLs over 125 dB (Hamernik and Ahroon, 1998). The ability of the auditory system to produce a TS_R seems to be dependent on an intact OHC system and is not affected by large inner hair cell (IHC) losses (Abroon and Hamernik, 2000; Hamernik *et al.*, 1998).

Another demonstration of the effects that the temporal variables of a noise exposure can have on the cochlea is provided by experiments showing that for the same exposure energy and spectrum, noise-induced trauma is a function of the kurtosis statistic (β), where kurtosis is defined as the ratio of the fourth-order central moment to the squared second-order central moment of the amplitude distribution. That is, non-Gaussian continuous noise exposures are more traumatic than are Gaussian exposures having equivalent energy and spectra. The increased hazard increases as β increases (Lei *et al.*, 1994; Hamernik and Qiu, 2001). Temporal effects, such as those described above, affect the applicability of current standards for predicting the hazard posed by excessive noise exposures such as the ISO-1999 (ISO, 1990) document. This standard incorporates an energy-based evaluation of an exposure. An energy metric is, however, insensitive to temporal factors.

The experimental results presented here show that: (1) Exposures to interrupted high-level, non-Gaussian signals produce a toughening effect comparable to that produced by an equivalent interrupted Gaussian noise. (2) Toughening, whether produced by Gaussian or non-Gaussian noise, results in reduced trauma compared to the equivalent uninterrupted noise, and (3) both continuous and interrupted non-Gaussian exposure produce more trauma than do energy and spectrally equivalent Gaussian noises.

II. METHODS

Thirty-six chinchillas (between 1- and 2-years old) were used as subjects. Each animal was made monaural by the

[a]Electronic mail: roger.hamernik@plattsburgh.edu

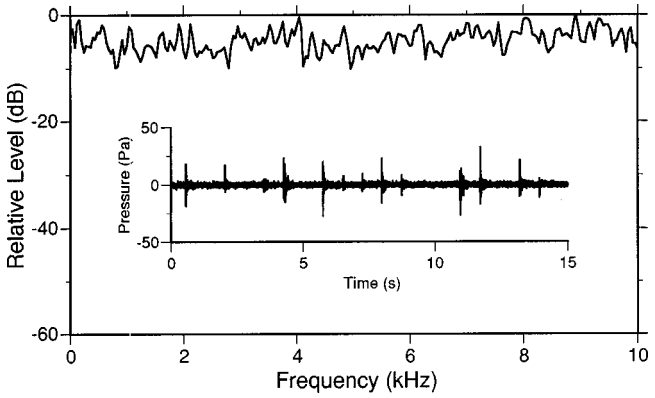


FIG. 1. The average spectrum obtained from eight 40-s samples of the digitized noise waveform. The spectrum was approximately the same for each of the four noise exposures. The inset shows a 15-s sample of the pressure-time waveform of a non-Gaussian exposure. Peak SPLs and inter-impact intervals were randomly varied.

surgical destruction, under anesthesia, of the left cochlea. During this procedure a bipolar electrode was implanted, under stereotaxic control, into the left inferior colliculus and the electrode plug cemented to the skull for the recording of auditory evoked potentials (AEP). The AEP was used to estimate pure-tone thresholds, and surface preparations of the organ of Corti were used to estimate the IHC, OHC populations. Additional details of the experimental methods, beyond those presented below, may be found in Ahroon *et al.* (1993).

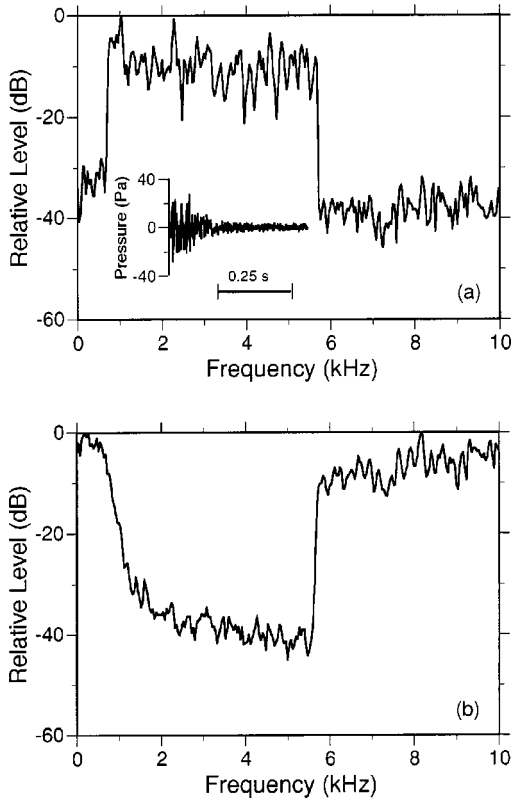


FIG. 2. (a) The spectrum of one of the impacts that was used to create the character of the non-Gaussian noise exposure. The insert shows an impact waveform. (b) The spectra of the complementary Gaussian noise that was mixed with the impact stimuli to form the non-Gaussian noise with kurtosis $\beta=33$.

TABLE I. Octave band SPL (dB) averaged over eight 40-s samples of the digitized waveform for the four exposure conditions.

Octave band cf (kHz)	G-5/20d	NG-5d	NG-20d
0.50	89	86	85
1.00	89	95	95
2.00	88	94	94
4.00	91	96	97
8.00	99	93	92
16.00	98	93	92
Mean L_{eq}	103	101.5	101.3
Mean L_{eq}(A)	100	100	100
s.d.	0.04	0.65	0.77

A. Noise exposures

During the exposures the noise field was monitored with a Larson Davis 814 sound-level meter equipped with a 1/2-in. microphone. The acoustic signal produced by the Electro-Voice Xi-1152/94 speaker system was transduced by a Brüel & Kjær 1/2 inch microphone (model 4134), amplified by a Brüel & Kjær (model 2610) measuring amplifier and fed to a WINDOWS PC-based analysis system. The design and digital generation of the acoustic signal is detailed in Hsueh and Hamernik (1990, 1991).

During exposure, individual chinchillas were confined to cages (10×11×16 in.) with free access to food and water. Peak SPLs of the impact transients in the non-Gaussian conditions were randomly varied between 115 and 129 dB. The impact had a probability of occurring in a 750-ms window of 0.6. The exposure field was uniform to within 2 dB. The four groups of animals were exposed to one of the following exposure protocols:

Group G-5d ($n=16$) Continuous Gaussian noise, 24 h/day for 5 days.
Group G-20d ($n=4$) Interrupted Gaussian noise, 6 h/day for 20 days.
Group NG-5d ($n=12$) Continuous non-Gaussian noise, 24 h/day for 5 days.
Group NG-20d ($n=4$) Interrupted non-Gaussian noise, 6 h/day for 20 days.

Each exposure had in common the same flat spectrum between 0.125 and 10.0 kHz shown in Fig. 1 and was pre-

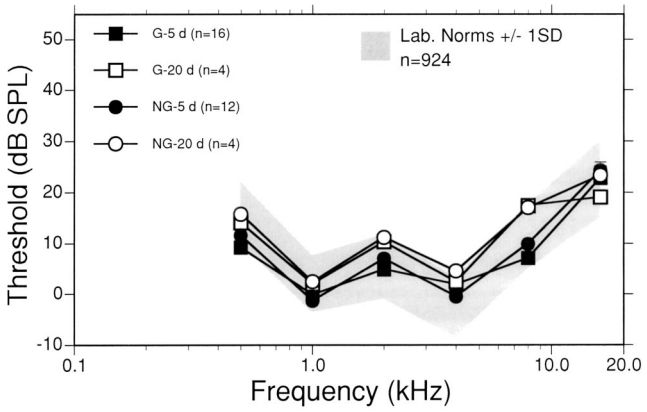


FIG. 3. The group preexposure AEP audiograms for each of the four experimental groups compared to the laboratory norm (shaded area).

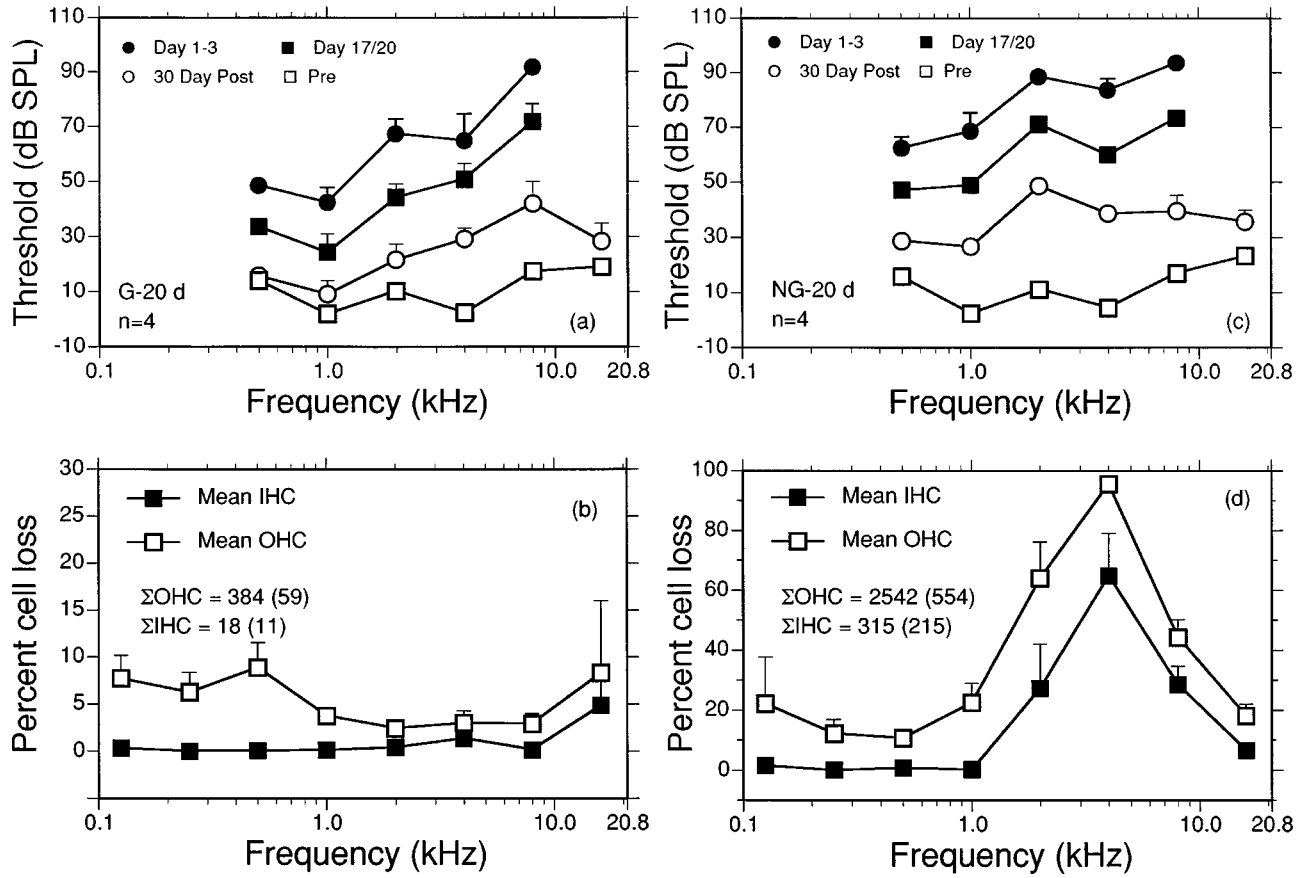


FIG. 4. (a) and (c) Group mean AEP threshold for the animals exposed to the (a) Gaussian noise 6-h/day for 20 days and (c) non-Gaussian noise 6h/day for 20 days. Preexposure threshold (□). 30-day postexposure threshold (○). The maximum threshold measured following exposure on day 1, 2, or 3 (●). The mean threshold measured following exposure on days 17 and 20 (■). (b) and (d) The group mean percent sensory cell loss in adjacent octave band lengths of the basilar membrane for animals exposed to (b) Gaussian noise 6h/day for 20 days and (d) non-Gaussian noise 6h/day for 20 days. Total mean IHC and OHC losses and the standard errors ( ) are indicated.

sented at an Leq-100 dB(A). The exposures differed only in their temporal structure, which was designed to produce two Gaussian and two non-Gaussian exposure conditions. The 5-day continuous exposures produced an asymptotic threshold shift (ATS) while the 20-day interrupted exposures produced a variety of TS patterns during the 20-day course of the exposure which resulted most often in a toughening effect, TS_R. The non-Gaussian conditions were designed in the frequency domain as described by Hsueh and Hamernik (1990, 1991) and were the result of inserting impacts, whose spectra were complementary to the background Gaussian noise, into an otherwise Gaussian signal. The impact peak levels were randomly varied between the limits indicated above and the probability of an impact occurring in a 750-ms window was set at 0.6. The inset in Fig. 1 shows a 15-s sample of the non-Gaussian, nonstationary waveform. Figure 2(a) shows one sample of the noise transients that produced the non-Gaussian signal along with its spectrum. Figure 2(b) shows the spectrum of the Gaussian component of the non-Gaussian signal. Table I presents the octave band levels of each noise exposure. Values shown are the mean values obtained from eight 40-s samples of the digitized waveform.

B. Threshold testing

AEP audiograms were measured at octave intervals from 0.5 to 0.8 or 16.0 kHz. The mean (in dB SPL) of three threshold determinations measured on different days defined each animal's pre- and 30-day postexposure audiogram. For the 20-day interrupted exposures, a complete audiogram (to 8.0 kHz) was measured following days 1, 2, 3, and 17 through 20. Between day 3 and 17 an audiogram was measured every other day. Because of the instability of TS discussed in a later section, the amount of threshold shift recovery (TS_R) at each audiometric test frequency was defined as the difference between the maximum TS measured at that frequency following days 1, 2, or 3 and the mean of the thresholds measured following exposure on days 17 through 20. TS_R is a measure of toughening, i.e., the amount that TS decreases during the 20-day interrupted exposure. A complete audiogram was measured once daily during each of the 5 exposure days of the uninterrupted exposures and the average (in dB SPL) taken over the 5 days established the mean asymptotic threshold levels and shifts.

C. Histology

Following the last AEP test protocol, each animal was euthanized under anesthesia and the right auditory bulla removed and opened to gain access to the cochlea for perfusion. Fixation solution consisting of 2.5% glutaraldehyde in veronal acetate buffer (final pH=7.3) was perfused through the cochlea. After 12 to 24 h of fixation the cochlea was

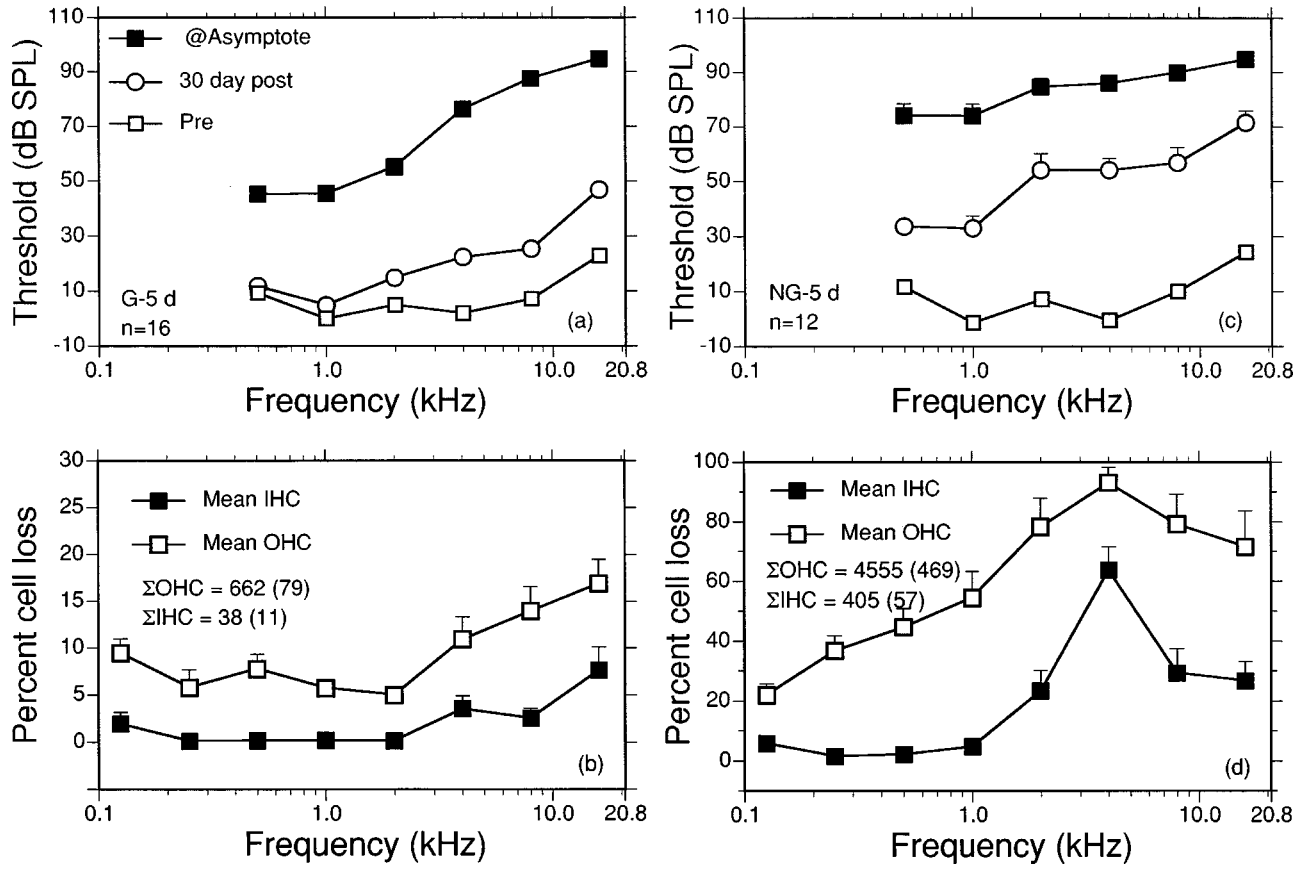


FIG. 5. (a) and (c) Group mean AEP thresholds for the animals exposed to the (a) Gaussian noise 24-h/day for 5 days and (c) non-Gaussian noise 24h/day for 5 days. Preexposure threshold (□). 30-day postexposure threshold (○). Asymptotic threshold levels (■). (b) and (d) The group mean percent sensory cell loss in adjacent octave band lengths of the basilar membrane for animals exposed to (b) Gaussian noise 24-h/day for 5 days and (d) non-Gaussian noise 24-h/day for 5 days. Total mean IHC and OHC loss and the standard errors ( ) are indicated.

postfixed in 1% OsO_4 in veronal acetate buffer. Surface preparation mounts of the entire organ of Corti were prepared and IHC, OHC populations were plotted as a function of frequency and location using the frequency-place map of Eldredge *et al.* (1981). Missing cells were identified by their characteristic phalangeal scars. For purposes of this presentation, sensory cell population data are presented as group averages (in percent missing) taken over octave-band lengths of the cochlea centered on the primary AEP test frequencies.

D. Statistical analysis

The dependent variables reported in this paper are (1) AEP thresholds and threshold shifts, before, during, and following noise exposure(s) and (2) sensory cell losses computed over octave-band lengths and the cochlea. Comparisons of groups of animals receiving different treatments were accomplished by mixed model analyses of variance with repeated measured on at least one factor (frequency). The probability of a type I error was set at 0.05 for all analyses. Statistically significant main effects of frequency are expected in most of the following analyses because of the frequency-specific nature of the chinchilla audiogram (Fay, 1988). For this reason any main effects of frequency will not be repeatedly discussed in the following presentation of results. Analysis of variance summary tables may be obtained from the authors.

III. RESULTS

The initial group mean thresholds for each of the four groups are shown in Fig. 3. In general, the group mean thresholds fall within ± one standard deviation of laboratory norms. Statistical analyses indicates that there was no significant main effect of group, but there was a significant interaction between group and frequency. The shaded region on the AEP audiograms in this figure represents the mean normative AEP audiogram (± one standard deviation) based on a population of 924 chinchillas. The bars on the data points in this and all subsequent figures represent one standard error of the mean. Where no bar is shown the standard error was less than the size of the datum symbol.

AEP thresholds prior to, during, and 30 days after exposure to the Gaussian (G-20 d) and non-Gaussian (NG-20 d) interrupted exposures are shown in Fig. 4 along with the respective group mean cochleograms. Total OHC and IHC losses along with standard errors in parentheses are also given. The ''day 1–3'' data points in Figs. 4(a) and (c) represent the maximum AEP threshold measured following day 1, 2, or 3 of the 20-day interrupted exposures, while the ''day 17/20'' data points represent the mean thresholds measured following exposure on days 17 through 20. In both of these figures the vertical distance between pairs of solid symbols at a given frequency represents the amount of toughening (TS_R) produced by the interrupted exposure at that frequency.

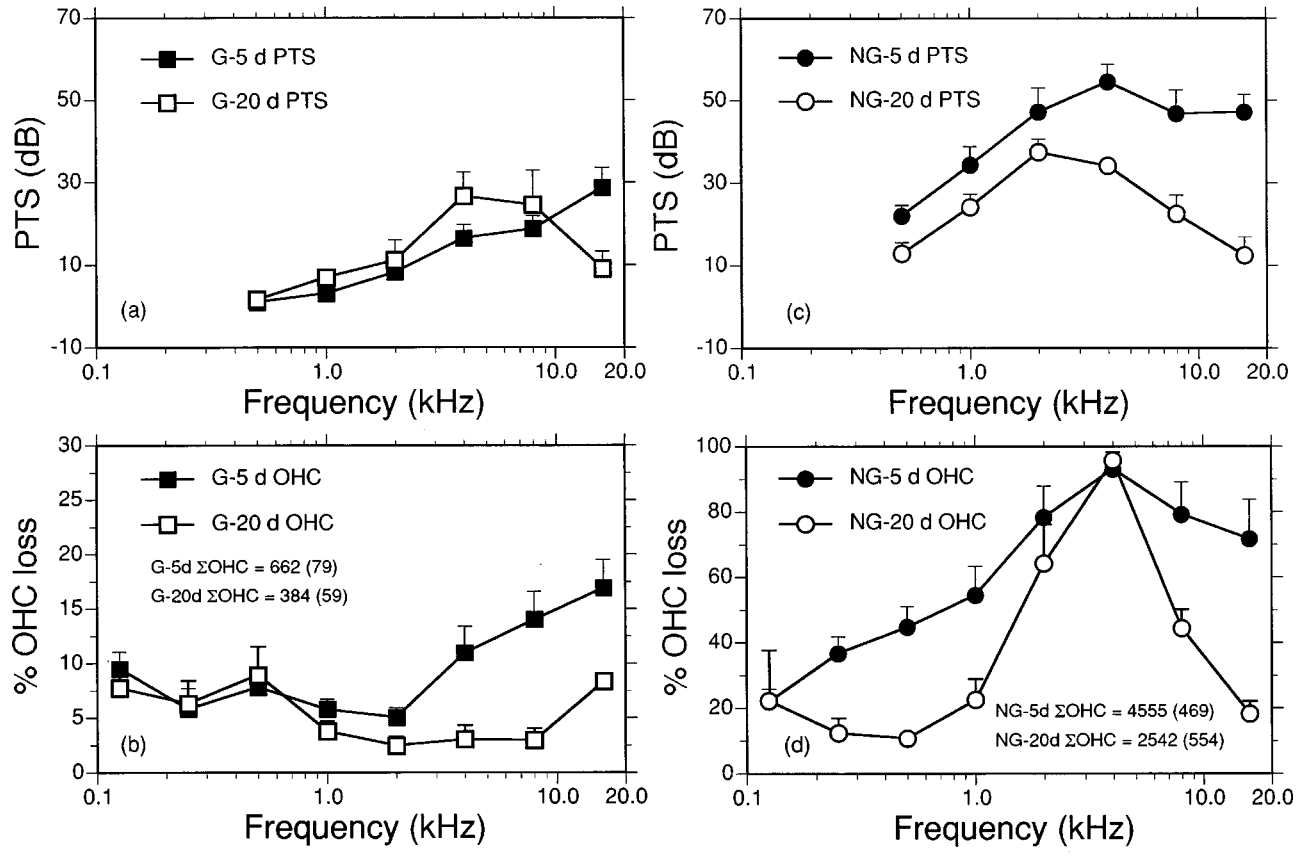


FIG. 6. (a) A comparison of the group mean PTS produced by the Gaussian noise 24-h/day for 5 days (■) and for 6-h/day for 20 days (□). (b) The group mean percent OHC loss in adjacent octave band lengths along the basilar membrane for groups shown in panel (a). (c) A comparison of the group mean PTS produced by the non-Gaussian noise 24-h/day for 5 days (●) and for 6-h/day for 20 days (○). (d) The group mean percent OHC loss in adjacent octave band lengths along the basilar membrane for groups shown in panel (c). Total mean OHC loss for each group and the standard errors ( ) are indicated.

Toughening refers to the improvement in threshold despite the continuing exposure. Note that while the interrupted non-Gaussian exposure produced greater TSs, the amount of TS_R (~15 to 20 dB) in both groups were roughly the same. Permanent effects from these two exposures, as quantified by PTS (the vertical distance between frequency-specific pairs of open symbols) and IHC and OHC loss are also presented.

Figure 5 shows a similar presentation of data from the Gaussian and non-Gaussian continuous 5-day exposures. The solid square symbols in panels (a) and (c) represent the asymptotic threshold levels measured during the continuous 5-day exposures. ATS can be estimated in these two panels by the frequency-specific vertical distance between pairs of square symbols. At and below 4 kHz the ATS produced by the non-Gaussian exposure (60 to 80 dB) was significantly greater than that produced by Gaussian exposure (35 to 80 dB). The lack of a difference above 4 kHz is probably a reflection of the upper limit of our AEP test system. The PTS and sensory cell loss data sets in Figs. 4 and 5 will be compared in the following several figures.

A comparison of the PTS produced by the Gaussian 5 and 20-day exposures and that produced by the two non-Gaussian exposures is shown in Fig. 6 along with the respective group mean OHC losses. The PTS produced by the two Gaussian exposures [panel (a)] varied from 0 to ~30 dB with greater loss at the higher frequencies. There was no statistically significant difference in the PTS produced by the 5- and 20-day Gaussian noise exposures, despite the approximately 15- to 20-dB TS_R (Fig. 4) found in the 20-day group. There was, however, a statistically significant decrease in the total as well as the frequency-specific OHC loss [panel (b)] for the interrupted 20-day exposure. For the two non-Gaussian exposures the interrupted 20-day exposure produced up to 35 dB less PTS than did the energy equivalent 5-day exposure [panel (c)] and a large statistically significant reduction in OHC loss [panel (d)]. PTS varied from 10 to ~35 dB in the non-Gaussian, 20-day interrupted group and from 20 to ~55 dB in the 5-day uninterrupted group. The profile of PTS also differed considerably between the two groups, with much less high frequency loss in the 20-day group. Note that the OHC loss profile generally reflects the PTS profile for each of the four groups shown in Fig. 6.

In Fig. 7 the PTS and OHC loss for the 20-day Gaussian and non-Gaussian exposures are compared in panels (a) and (b). The non-Gaussian exposures produced statistically significant more PTS (up to 26 dB) and OHC loss than did the energy equivalent Gaussian exposure. A similar comparison in panels (c) and (d) between the Gaussian and non-Gaussian 5-day exposures also showed large statistically significant differences in PTS (up to 35 dB) and OHC loss between the two energy equivalent exposures.

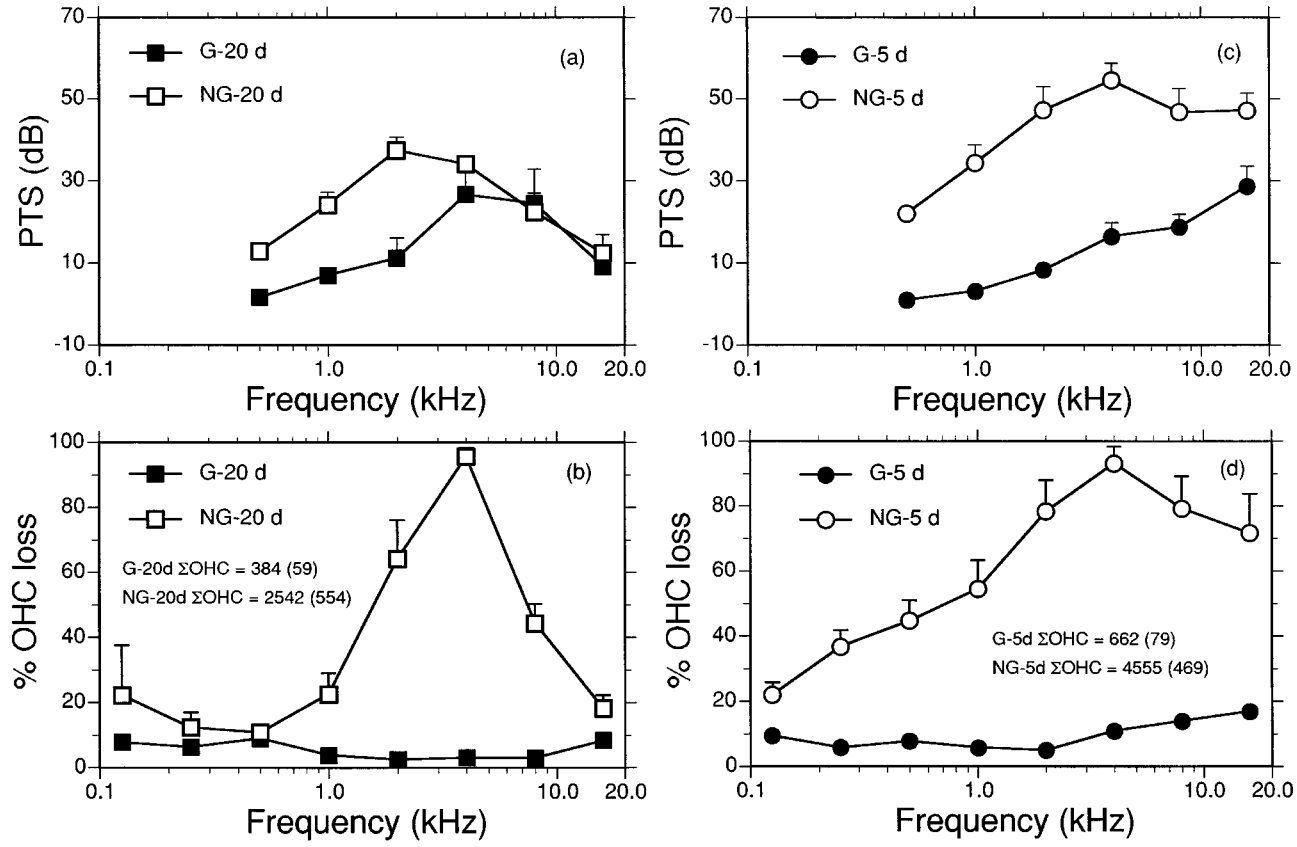


FIG. 7. (a) A comparison of the group mean PTS produced by the Gaussian noise 6-h/day for 20 days (■) and the non-Gaussian noise 6-h/day for 20 days (□). (b) The group mean percent OHC loss in adjacent octave band lengths along the basilar membrane for groups shown in panel (a). (c) A comparison of the group mean PTS produced by the Gaussian noise 24-h/day for 5 days (●) and the non-Gaussian noise 24-h/day for 5 days (○). (d) The group mean percent OHC loss in adjacent octave band lengths along the basilar membrane for groups shown in panel (c). Total mean OHC loss for each group and the standard errors ( ) are indicated.

IV. DISCUSSION

With the exception of the early work of Miller *et al.* (1963), most interrupted exposures have used octave bands of noise to study toughening effects and the potential for protection that toughening may produce for a subsequent exposure. While a number of studies have demonstrated protective effects when the subject is subsequently exposed to a traumatic noise (e.g., Campo *et al.*, 1991; Henselman *et al.*, 1994; McFadden *et al.*, 1997) others have not (Miller *et al.*, 1963; White *et al.*, 1998; Subramaniam *et al.*, 1993). Another protective effect that has received less attention is that produced by the toughening effect on the interrupted noise that produced the toughening. In this situation both the toughening effect and the recovery process that take place during the quiet periods are not easily separated, although the latter influence would be expected to reduce trauma (Ward, 1991). Clark *et al.* (1987) and Bohne *et al.* (1987) showed less PTS and hair cell loss in subjects toughened by an interrupted noise compared to a control group. Ward (1991) also showed that less trauma is produced by interrupted/intermittent exposures compared to equivalent energy controls. His experimental paradigm did not allow for any estimate of TS_R. Hamernik and Ahroon (1998) used high-level narrow-band impacts and a large sample size to study the toughening phenomena. The impacts clearly produced a TS_R. They showed, however, that the toughened subjects had approximately the same levels of PTS and sensory cell loss as control subjects exposed to the same impacts but on an uninterrupted schedule. In our present study, Fig. 4 shows that both Gaussian and non-Gaussian exposures, having the same energy and spectra, produce a clear 15- to 20-dB TS_R across the entire range of test frequencies (0.5 through 8.0 kHz). The threshold shifts, however, for the non-Gaussian exposure were greater. And, in agreement with the above studies that showed a reduction in trauma, both of the interrupted exposures showed reduced sensory cell loss. For the non-Gaussian interrupted exposure [NG-20 d] there was also a large (up to 35 dB) reduction in PTS [Fig. 6(c)] when compared with the uninterrupted exposure.

While the toughening effect is similar for both the Gaussian and non-Gaussian exposures, the level of trauma produced by both the interrupted and uninterrupted non-Gaussian exposures exceeds by a large amount (Fig. 7) the trauma produced by the respective Gaussian exposures. This increased trauma from the non-Gaussian exposures agrees with our earlier data (Lei *et al.*, 1994; Hamernik and Qiu, 2001). The large differences in PTS and sensory cell loss between the effects of Gaussian and non-Gaussian exposures are likely related to the excessive stress/strain on the epithelial tight cell junctions induced by the high-level impulsive forces and the subsequent momentum changes that are produced. Disrupting the integrity of the tight cell junctions with

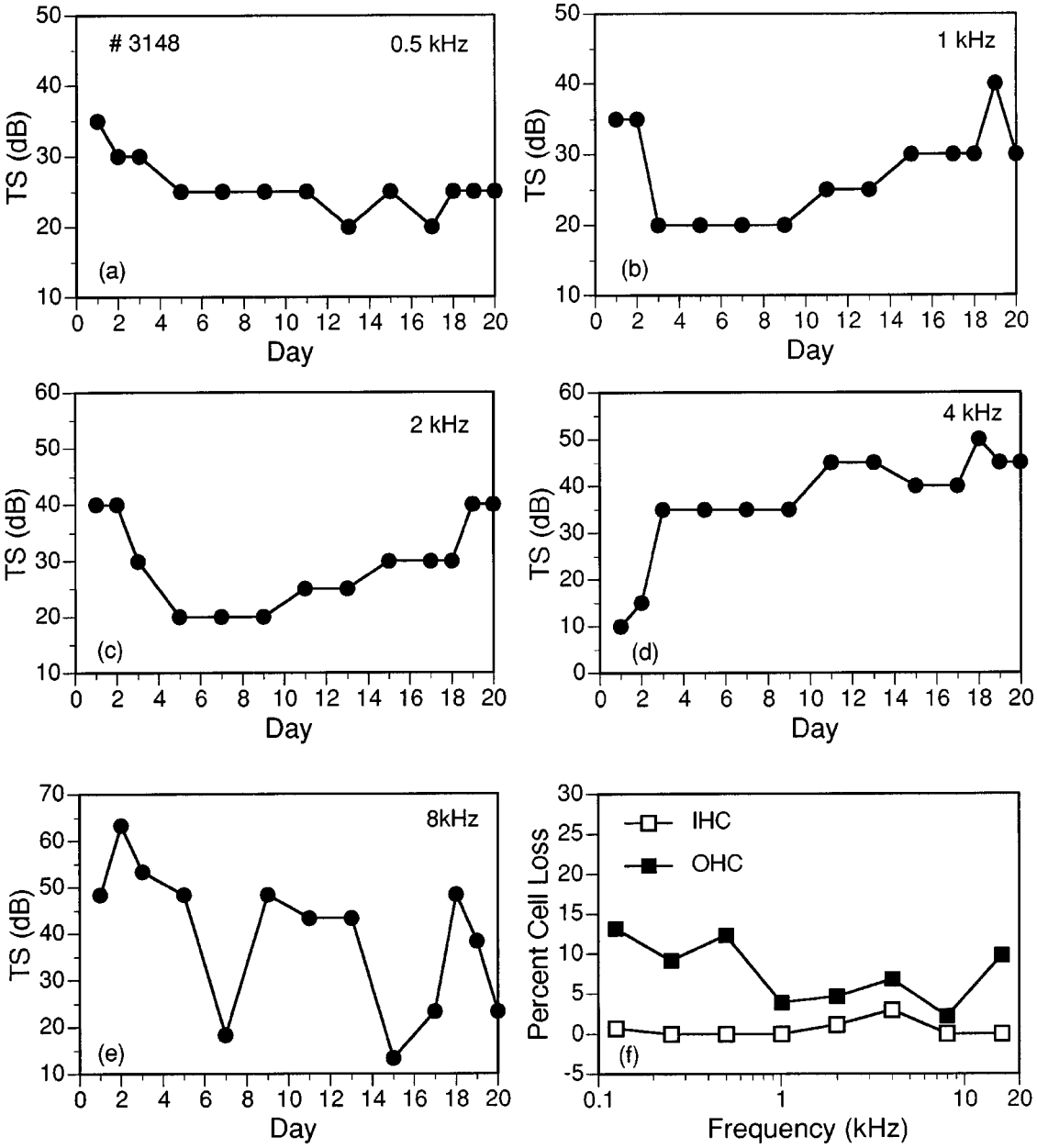


FIG. 8. Examples of an individual animal's TSs, measured at the indicated AEP test frequency [panels (a)–(e)], during the course of the 20-day exposure. The animal is #3148 from group G-20 d. The octave band percent IHC and OHC loss is shown in panel (f).

the entry to endolymph into the space below the reticular lamina is known to increase the extent of cell loss (Bohne and Rabbitt, 1983). Also, the severe tearing of the organ of Corti seen with high-level impacts (Hamernik *et al.*, 1984) probably releases a variety of free radicals into the sensory cell environment as a consequence of the processes involved in clearing the cellular debris. These highly reactive species are also known to have a detrimental effect on cochlear function (Jacono *et al.*, 1998). In addition, the induction of apoptosis may further exacerbate these effects (Hu *et al.*, 2002).

For both non-Gaussian exposures $\beta=33$ (approximately). This is an average value calculated from eight 40-s samples of the temporal waveform. The above results would not be anticipated from an application of the equal energy hypothesis, which is the basis of the current international noise standard for the protection of hearing. These results also show that the temporal structure, whether altered through intermittence or through a non-Gaussian peak distribution, can exert a strong effect on the outcome of an exposure. Much of the variability seen in the human demographic data (Mills *et al.*, 1996) may be the result of a neglect of temporal variables. The temporal variables of an exposure need to be taken into account in the formulation of damage risk criteria for noise exposure.

While most animals in the two interrupted exposure paradigms showed a TS_R at most frequencies, there were some frequencies that displayed ''nontypical'' TS configurations over the course of the exposure. One animal in particular was quite variable and showed an interesting assortment of TS functions, any one of which could be found in the other subjects at some frequency. Figure 8 shows the TS measured in this animal (#3148) during the course of the

Gaussian, 20-day exposure along with the animal's cochleogram [panel (f)]. At 0.5 kHz TS follows a pattern typical of the toughening phenomena. The animal shows a 35-dB TS following the first day of the exposure. Over the next several days TS decreases and reaches a stable value of about 25 dB, i.e., a $TS_R = 10$ dB. At 1.0 and 2.0 kHz the TS pattern is ''U'' shaped, with TS initially decreasing as much as 20 dB followed by a gradual increase to levels close to or approaching those measured following the first day's exposure. There is little or no TS_R at these frequencies. At 4.0 kHz TS is initially relatively low (~10 dB) but then increases over the course of the exposure to ~40 to 45 dB. At this frequency the cochlea has become more susceptible to the noise as the exposure continued, i.e., a $-TS_R$. Finally, at 8.0 kHz TS measures are unstable. TS fluctuates as much as 30 dB over the course of the exposure. The large TS fluctuations seen in the ''unstable'' configuration could be found at various frequencies and on different exposure days in different animals. The AEP waveforms recorded prior to a fluctuation, on the day of the TS dip and on the day following the dip, showed clear and regular intensity-dependent AEP waveforms with threshold well delineated. Clearly, a variety of different processes both protective and pathological must underlie these functions which eventually resolve and yield stable but shifted hearing thresholds for the subject.

V. ANIMAL USE

The care and use of the animals used in this study was approved by the Plattsburgh State University of New York Institutional Animal care and Use Committee. In conducting this research the investigators adhered to the Guide for Care and Use of Laboratory Animals, as promulgated by the Committee on Care and Use of Laboratory Animals in the Institute of Laboratory Resources Commission on Life Sciences, National Academy of Sciences-National Research Council, revised 1985.

ACKNOWLEDGMENTS

This work was supported by Grant No. 1-R01-OH02317 from the National Institute for Occupational Safety and Health. The able technical assistance of Ann Johnson, George A. Turrentine, Diane Fresch, Ben Moenter, and Thomas Lewis is greatly appreciated.

Ahroon, W. A., Hamernik, R. P., and Davis, R. I. (**1993**). ''Complex noise exposures: An energy analysis,'' J. Acoust. Soc. Am. **93**, 997–1006.

Ahroon, W. A., and Hamernik, R. P. (**2000**). ''The effects of interrupted noise exposures on the noise-damaged cochlea,'' Hear. Res. **143**, 103–109.

Boettcher, F. A., Sponger, V. P., and Salvi, R. J. (**1992**). ''Physiological and histological changes associated with the reduction of the threshold shift during interrupted noise exposure,'' Hear. Res. **62**, 217–236.

Bohne, B. A., and Rabbitt, K. D. (**1983**). ''Holes in the reticular lamina after noise exposure: Implication for continuing damage in the organ of Corti,'' Hear. Res. **11**, 41–53.

Bohne, B. A., Yohman, L., and Gruner, M. M. (**1987**). ''Cochlear damage following interrupted exposure to high-frequency noise,'' Hear. Res. **27**, 251–264.

Brownell, W. E. (**1990**). ''Outer hair cell electromotility and otoacoustic emissions,'' Ear Hear. **11**, 82–92.

Campo, P., Subramanian, M., and Henderson, D. (**1991**). ''The effects of 'conditioning' exposures on hearing loss from traumatic exposure,'' Hear. Res. **55**, 195–200.

Clark, W. W., Bohne, B. A., and Boettcher, F. A. (**1987**). ''Effects of periodic rest on hearing loss and cochlear damage following exposure to noise,'' J. Acoust. Soc. Am. **82**, 1253–1264.

Eldredge, D. H., Miller, J. A., and Bohne, B. A. (**1981**). ''A frequency-position map for the chinchilla cochlea,'' J. Acoust. Soc. Am. **69**, 1091–1095.

Fay, R. A. (**1988**). *Hearing in Vertebrates* (Hill-Fay, Winnetka, IL).

Hamernik, R. P., Turrentine, G., Roberto, M., Salvi, R., and Henderson, D. (**1984**). ''Anatomical correlates of impulse noise-induced mechanical damage in the cochlea,'' Hear. Res. **13**, 229–247.

Hamernik, R. P., and Ahroon, W. A. (**1998**). ''Interrupted noise exposures: Threshold shift dynamics and permanent effects,'' J. Acoust. Soc. Am. **103**, 3478–3488.

Hamernik, R. P., Ahroon, W. A., Jock, B. M., and Bennett, J. A. (**1998**). ''Noise-induced threshold shift dynamics measured with distortion-product otoacoustic emissions and auditory evoked potentials in chinchillas with inner hair cell deficient cochleas,'' Hear. Res. **118**, 73–82.

Hamernik, R. P., and Qiu, W. (**2001**). ''Energy-independent factors influencing noise-induced hearing loss in the chinchilla model,'' J. Acoust. Soc. Am. **110**, 3163–3168.

Henselman, L. W., Henderson, D., Subramaniam, M., and Sallustio, V. (**1994**). ''The effect of 'conditioning' exposures on hearing loss from impulse noise,'' Hear. Res. **78**, 1–10.

Hsueh, K. D., and Hamernik, R. P. (**1990**). ''A generalized approach to random noise synthesis: Theory and computer simulation,'' J. Acoust. Soc. Am. **87**, 1207–117.

Hsueh, K. D., and Hamernik, R. P. (**1991**). ''Performance characteristics of a phase domain approach to random noise synthesis,'' Noise Control Eng. J. **36**, 18–32.

Hu, B. H., Henderson, D., and Nicotera, T. M. (**2002**). ''Involvement of apoptosis in progression of cochlear lesion following exposure to intense noise,'' Hear. Res. **166**, 62–71.

ISO-1999 (**1990**). ''Acoustics: Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment'' (International Organization for Standardization, Geneva).

Jacono, A. A., Hu, B., Kopke, R. D., Henderson, D., Van De Water, T. R., and Steinman, H. M. (**1998**). ''Changes in cochlear antioxidant enzyme activity after sound conditioning and noise exposure in the chinchilla,'' Hear. Res. **117**, 31–38.

Lei, S. F., Ahroon, W. A., and Hamernik, R. P. (**1994**). ''The application of frequency and time domain kurtosis to the assessment of hazardous noise exposures,'' J. Acoust. Soc. Am. **96**, 1435–1444.

McFadden, S. L., Henderson, D., and Shen, Y. H. (**1997**). ''Low-frequency 'conditioning' provides long-term protection from noise-induced threshold shifts in chinchillas,'' Hear. Res. **103**, 142–150.

Miller, J. D., Watson, C. S., and Covell, W. P. (**1963**). ''Deafening effects of noise on the cat,'' Acta Oto-laryngologica Suppl. **176**, 1–81.

Mills, J. H., Lee, F. S., Dubno, J. R., and Boettcher, F. A. (**1996**). ''Interactions between age-related and noise-induced hearing loss,'' in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Axelsson, H. Borchgrevink, R. P. Hamernik, P. A. Hellstrom, D. Henderson, and R. J. Salvi (Thieme, New York), pp. 193–212.

Subramaniam, M., Campo, P., and Henderson, D. (**1991**). ''The effect of exposure level on the development of progressive resistance to noise,'' Hear. Res. **52**, 181–188.

Subramaniam, M., Henderson, D., Campo, P., and Sponger, V. (**1992**). ''The effect of 'conditioning' on hearing loss from a high frequency traumatic exposure,'' Hear. Res. **58**, 57–62.

Subramaniam, M., Henderson, D., and Sponger, V. (**1993**). ''Effect of low-frequency 'conditioning' on hearing loss from high-frequency exposure,'' J. Acoust. Soc. Am. **93**, 952–956.

Ward, W. D. (**1991**). ''The role of intermittence in PTS,'' J. Acoust. Soc. Am. **90**, 164–169.

White, D. R., Boettcher, F. A., Miles, L. R., and Gratton, M. A. (**1998**). ''Effectiveness of intermittent and continuous acoustic simulation in preventing noise-induced hearing and hair cell loss,'' J. Acoust. Soc. Am. **103**, 1566–1572.

Perception of the low pitch of frequency-shifted complexes

Geoffrey A. Moore and Brian C. J. Moore[a)]
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 2 October 2001; revised 6 March 2002; accepted 18 November 2002)

When all of the components in a harmonic complex tone are shifted in frequency by Δf, the pitch of the complex shifts roughly in proportion to Δf. For tones with a small number of components, the shift is usually somewhat larger than predicted from pitch theories, which has been attributed to the influence of combination tones [Smoorenburg, J. Acoust. Soc. Am. **48**, 924–941 (1970)]. Experiment 1 assessed whether combination tones influence the pitch of complex tones with more than five harmonics, by using noise to mask the combination tones. The matching stimulus was a harmonic complex. Test complexes were bandpass filtered with passbands centered on harmonic numbers 5 (resolved), 11 (intermediate), or 16 (unresolved) and fundamental frequencies ($F0$s) were 100, 200, or 400 Hz. For the intermediate and unresolved conditions, the matching stimuli were filtered with the same passband to minimize differences in the excitation patterns of the test and matching stimuli. For the resolved condition, the matching stimulus had a passband centered above that of the test stimulus, to avoid common partials. For resolved and intermediate conditions, pitch shifts were observed that could generally be predicted from the frequencies of the partials. The shifts were unaffected by addition of noise to mask combination tones. For the unresolved condition, no pitch shift was observed, which suggests that pitch is not based on temporal fine structure for stimuli containing only high unresolved harmonics. Experiment 2 used three-component complexes resembling those of Schouten [J. Acoust. Soc. Am. **34**, 1418–1424 (1962)]. Nominal harmonic numbers were 3, 4, 5 (resolved), 8, 9, 10 (intermediate), or 13, 14, 15 (unresolved) and $F0$s were 50, 100, 200, or 400 Hz. Clear shifts in the matches were found for all conditions, including unresolved. For the latter, subjects may have matched the ''center of gravity'' of the excitation patterns of the test and matching stimuli. [DOI: 10.1121/1.1536631]

PACS numbers: 43.66.Hg, 43.66.Ba [NFV]

I. INTRODUCTION

Schouten (1940) found that the perceived pitch of harmonic complex tones without a fundamental component was equal to the pitch of the missing fundamental component. He concluded that the higher harmonics were perceived collectively as a ''residue pitch'' determined by the periodicity of the waveform. Here, we use the more neutral term ''low pitch.'' If each component in a harmonic complex is shifted in frequency by the same amount, Δf, the perceived pitch generally shifts in the same direction, by Δp. There are two classes of model to account for this pitch shift, both of which predict that the shift should be proportional to the frequency shift of the complex and inversely proportional to the harmonic number (n) of one of the components in the unshifted complex

$$\Delta p = \Delta f / n. \tag{1}$$

For complex tones with only three harmonics, n has usually been taken as the harmonic number of the middle component (Schouten *et al.*, 1962), but for complex tones with more harmonics, n is often taken as corresponding to the lowest harmonic (Patterson, 1973; Patterson and Wightman, 1976). The pitch shift defined by Eq. (1) is called the ''first effect of pitch shift'' (Schouten *et al.*, 1962).

The first class of model assumes that low pitch is based on the time intervals between peaks in the fine structure of the waveform (on the basilar membrane) close to successive envelope maxima (de Boer, 1956; Schouten *et al.*, 1962). The time intervals may be extracted directly from interspike intervals (Cariani and Delgutte, 1996a, 1996b; Moore, 2003) or by a process of autocorrelation (Licklider, 1951; Yost, 1996; Meddis and O'Mard, 1997). The second class of model assumes that low pitch is derived from the pitches or frequencies of individual resolved partials, by the generation of subharmonics (Terhardt, 1974), by determining the fundamental frequency ($F0$) whose harmonics would provide the best match to the frequencies of the components in the test stimulus (Goldstein, 1973), or by some other pattern-matching process (Cohen *et al.*, 1995). All of these models predict results of the form described by Eq. (1).

The first systematic investigation of the phenomenon of pitch shift was carried out by de Boer (1956) using five-tone complexes. He confirmed that the pitch shift was proportional to Δf. However, the shift was sometimes bigger than predicted from the first effect of pitch shift [as defined by Eq. (1)]. This larger than expected effect is sometimes called the ''second effect of pitch shift'' (Schouten *et al.*, 1962). The deviation of the pitch shift from the ''expected value'' is reduced if n is taken as the harmonic number of the lowest component in the complex. However, even when this is done, a larger than expected effect is sometimes observed.

[a)]Electronic mail: bcjm@cus.cam.ac.uk

These findings were confirmed by Schouten *et al.* (1962) using three-tone complexes, and by Smoorenburg (1970) using two-tone complexes.

The importance of the lower harmonics in determining pitch-shift effects can be understood in terms of the concept of dominance, which was introduced by Ritsma (1967): ''If pitch information is available along a large part of the basilar membrane the ear uses only the information from a narrow band. This band is positioned at 3–5 times the pitch value. Its precise value depends somewhat on the subject'' (Ritsma, 1970). Moore *et al.* (1984, 1985) suggested that there are in fact large individual differences in which harmonics is dominant, and that for some subjects the first two harmonics are the most important. Edge components, and in particular the bottom component, may be better resolved than components within a complex (Moore *et al.*, 1984; Moore and Ohgushi, 1993) and therefore may contribute more to low pitch.

Smoorenburg (1970) suggested that combination tones may influence the low pitch of complex tones containing only high harmonics. Combination tones of the type $2f_1-f_2$ correspond to components lower than those in the presented complex, and they would therefore fall closer to the dominant region. Introducing lower components would effectively reduce the value of n in Eq. (1), thus explaining the second effect of pitch shift. One way to reduce the influence of combination tones is to use very low sound levels (Smoorenburg, 1970; Houtsma and Goldstein, 1972). For example, Smoorenburg (1970) showed that the effect of combination tones on the pitch of two-tone complexes decreased as the level of (each of) the primary tones was decreased from 55 to 25 dB SL (overall levels of about 63 to 33 dB SPL). Another method is to use two-tone complexes presented dichotically (Houtsma and Goldstein, 1972). However, for harmonic numbers at which the second effect of pitch shift becomes measurable, the low pitch of a dichotic two-component complex becomes very weak, and is hard to match consistently. A third method of reducing the influence of combination tones is to use a noise to mask those tones. Smoorenburg (1970) reported limited data on the effect of a low-pass filtered pink noise on the pitch of frequency-shifted two-tone complexes. He used a single nominal $F0$ of 200 Hz, and a single nominal harmonic number for each of two subjects (the nominal harmonic number of the lowest component was 9 for one subject and 10 for the other). He found that the noise essentially eliminated the second effect of pitch shift, and also reduced the clarity of the pitch sensation. These results are consistent with what would be expected on the basis of the noise masking combination tones.

A second effect of pitch shift was not found by Patterson (1973), or by Patterson and Wightman (1976); both these studies used complex tones with either 6 or 12 components and n was taken as corresponding to the lowest component. This may indicate that combination tones do not influence low pitch when the complex tones contain many harmonics. However, it is also possible that combination tones play some role, but that role is not sufficient to lead to a second effect of pitch shift in a 6- or 12-component stimulus. In the present study we assessed the influence of combination tones on the pitch of complex tones with at least five components, by using noise to mask the combination tones. We included a control condition where no noise was present. If combination tones do influence low pitch, then the noise should reduce the pitch-shift effect even if no second effect is observed. We used stimuli at a moderate sound level (70 dB SPL overall), so as to be representative of typical ''real-life'' listening conditions; many previous experiments have used stimuli at rather low levels, although the data of Smoorenburg (1970), described earlier, suggest that combination tones have a greater influence on pitch at higher levels. The data of Pressnitzer and Patterson (2001) indicate that for complex tones containing five or more components with a level per component of 54 dB SPL (overall level of 65 dB SPL), the level of combination tones estimated using a cancellation method (including the simple difference tone) may be only about 15 dB below the level of the primary tones, i.e., at about 40 dB.

The task used by Schouten *et al.* (1962) required listeners to adjust the $F0$ of a harmonic complex tone so that its pitch matched that of a frequency-shifted ''test'' complex tone. The matching and test stimuli had the same nominal harmonic numbers, i.e., they overlapped spectrally. As the matching tone was shifted upwards, its excitation pattern would also have shifted upwards. We hypothesized that subjects might match the excitation patterns of the test and matching stimuli rather than matching their low pitch. In experiment 1, therefore, we used multicomponent complexes that were spectrally shaped so as to reduce or eliminate these excitation-pattern cues. This was achieved by giving the stimuli a spectral envelope with a flat central section and sloping edges; the spectral envelope was kept fixed when the component frequencies were shifted. When the stimuli contained only high harmonics, this resulted in excitation patterns (Glasberg and Moore, 1990), which hardly changed with component frequency shift. Experiment 2 used stimuli like those of Schouten *et al.* (1962). By comparing the results of these experiments we hoped to elucidate the importance of excitation-pattern cues. In both experiments we set out to investigate systematically how low pitch shift is related to component frequency shift for a range of nominal $F0$s, within the range shown by Ritsma (1962) to elicit a sense of low pitch. We also used a range of harmonic numbers n in order to include stimuli that contain only low (resolved) components, intermediate components, or high (unresolved) components.

II. EXPERIMENT 1. PITCH MATCHING WITH SPECTRALLY SHAPED MULTICOMPONENT COMPLEXES

A. Method

1. Stimuli

Listeners were asked to adjust the $F0$ of a harmonic complex tone to match the perceived pitch of a frequency-shifted ''test'' complex tone. The test complex was derived from components of a 100-, 200-, or 400-Hz $F0$. We planned to use an $F0$ of 50 Hz as well, but pilot studies showed that pitch matching for an $F0$ of 50 Hz was highly erratic, so this $F0$ was omitted. For each $F0$, three fixed spectral envelopes were used for the test stimulus, each with a flat passband and

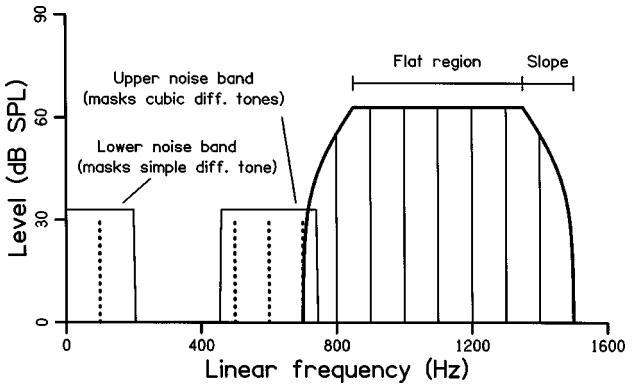


FIG. 1. Schematic diagram showing the spectral shaping applied to stimuli and the noise bands used to mask combination tones in experiment 1. The INT condition is shown for $F0 = 100$ Hz. Level is actual level in dB SPL for partials and spectrum level for noises.

sloping edges. This is illustrated schematically in Fig. 1. The passbands were characterized by the components in the passband: resolved (RES), intermediate (INT), and unresolved (UNRES). The ''matching'' complex consisted of harmonics of an $F0$ controlled by the subject. The spectral envelope of the matching stimulus was the same as for the test stimulus in the INT and UNRES conditions, but differed from that of the test stimulus in the RES condition; see below for details. The INT and UNRES conditions used spectral envelopes centered on component numbers 11 and 16, respectively. The components for condition INT are classed as being of intermediate resolvability because the ''edge'' components might have been resolved even though the components within the complex were not resolved (Moore *et al.*, 1984; Moore and Ohgushi, 1993).

For the INT and UNRES conditions, the flat region of the spectral envelope covered a frequency range equal to $5F0$. The amplitude of the spectral envelope at a given frequency F within the sloping region was determined by a frequency variable x, defined as $1-|(F-F_e)/1.5F0|$, where F_e is the frequency at the edge of the flat region. The amplitude relative to that in the flat region was set to $(10^x - 1)/9$. The amplitude was set to zero when x was less than or equal to zero. Excitation patterns calculated as described by Glasberg and Moore (1990) hardly changed when the components were frequency shifted for the INT and UNRES conditions.

To prevent listeners comparing the pitches of individual components in the RES condition, we used test and matching stimuli in different spectral regions. The test stimulus envelope was centered on component number 5 and the flat region extended over $3F0$. The sloping regions were defined in the same way as for conditions INT and UNRES. The matching stimulus was a harmonic complex comprising components 8–15, with the levels of the two lowest and two highest components being progressively reduced with a slope of -6 dB/component.

The component frequency shifts, Δf, ranged from 0% to 24% of $F0$ in steps of 8%. Pilot studies showed that pitch matching became very difficult for shifts larger than 24%. The test and matching stimuli were presented in alternation with a 500-ms interval between test and matcher and an 800-ms interval between matcher and test. The component starting phase was cosine. All stimuli had a duration of 500 ms including 10-ms raised-cosine onset/offset ramps. The overall level for each complex was 70 dB SPL. Stimuli were generated through a 16-bit D/A converter (TDT, DD1) at a 50-kHz sampling rate, attenuated (TDT, PA4), and presented via a headphone buffer (TDT, HB6) through one earpiece of a Sennheiser HD580 headset.

TABLE I. Noise band cutoff frequencies for each combination of $F0$ and resolvability.

$F0$ (Hz)	Resolvability	Lower band (Hz)	Upper band (Hz)
50	RES	0–120	None
50	INT	0–150	230–370
50	UNRES	0–150	480–620
100	RES	0–230	None
100	INT	0–200	460–740
100	UNRES	0–200	960–1240
200	RES	0–500	None
200	INT	0–300	900–1500
200	UNRES	0–300	1900–2500
400	RES	0–1000	None
400	INT	0–500	1800–3000
400	UNRES	0–500	3800–5000

Two conditions were used, one with masking noise (NOISE) and one without (QUIET). All noise was generated and filtered digitally. One noise band was shaped to cover the spectral region of the simple difference tone generated by two consecutive components of the test complex. A separate band of noise was placed below the test complex covering the spectral region of the main cubic difference tones (Fig. 1). See Table I for cutoff frequencies of the lower and upper noise bands for each combination of $F0$ and resolvability (the section showing $F0 = 50$ Hz is relevant to experiment 2). The slopes of the spectra of the noises outside the passbands were essentially infinite. The noise spectrum level within the passbands was 33 dB. This should have been sufficient to mask any combination tones whose level was 15 dB or more below the level of the primaries. The noise was presented with both the test and matching stimuli. For both, it was gated on 150 ms before and gated off 150 ms after the stimuli.

2. Subjects

Five normal-hearing subjects were tested. All were musically trained and their ages ranged from 19–31 years. All had absolute thresholds better than 20 dB HL at the standard audiometric frequencies.

3. Procedure

Subjects sat in a sound-isolated booth and responded using a three-button box. Subjects adjusted the $F0$ of the harmonic matching complex so that the pitch matched that of the test complex. The matching $F0$ starting value was randomized within a range of 6% above or below a ''target'' $F0$ calculated for each trial based on the first effect of pitch shift [Eq. (1)], taking n as the harmonic number of center component in the test complex. The subject used one button to raise the matching stimulus $F0$ and another one to lower it; the

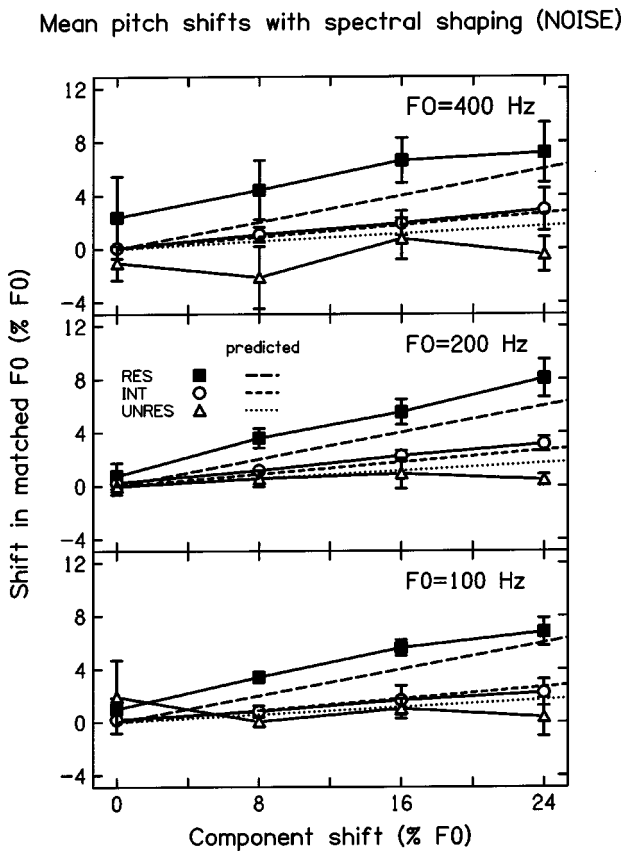


FIG. 2. Mean results for all subjects in the NOISE condition of experiment 1. In each panel the symbols represent the conditions RES (squares), INT (circles), and UNRES (triangles). Shift in $F0$ of the matching complex is plotted against component frequency shift of the test complex, Δf. Error bars show ± one s.d. of the mean across subjects for each combination of $F0$, resolvability, and component shift. Dashed and dotted lines show predictions based on n corresponding to the lowest component in the flat region of the spectral envelope.

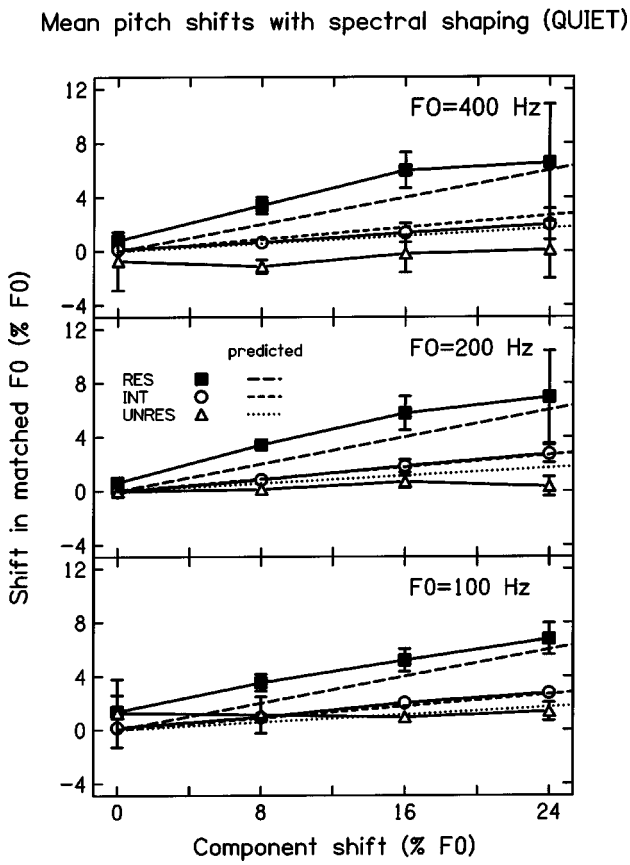


FIG. 3. As Fig. 2, but for the QUIET condition.

starting $F0$ step size was 1%, and it was reduced to 0.4% after two reversals in the direction of change and to 0.2% after three reversals. $F0$ was not changed during presentation of a matching stimulus. There was no limit on the number of reversals, and when the subject was satisfied with the pitch match (s)he pressed a third button on the button box. The computer then recorded the results and stopped the trial. Subjects were asked to listen for the ''low'' pitch in the stimuli rather than to the pitches of individual components, and were encouraged to use a ''bracketing'' strategy. At least six blocks of practice trials were given. These included one at each $F0$, for at least one each of the conditions RES, INT, UNRES.

Blocks consisted of four consecutive trials (one for each shift) for a given $F0$ and resolvability, and the order of block presentation was randomized. For each combination of $F0$, resolvability, and shift, three pitch matches were recorded in different blocks. All NOISE conditions were completed before starting on QUIET conditions.

B. Results

Results are shown in Figs. 2 and 3. The shift of the matcher $F0$ is plotted as a function of Δf expressed as a percentage of $F0$. Error bars show ± one standard deviation (s.d.) of the mean across subjects. The straight lines show predictions based on the first effect of pitch shift [Eq. (1)], taking n as corresponding to the lowest component in the flat region of the spectral envelope (component 4 for RES, 9 for INT, and 14 for UNRES). These values were chosen because the lowest-frequency high-amplitude component might be expected to be dominant (Moore *et al.*, 1984; Moore *et al.*, 1985). However, our choice of n does not imply that partials on the lower sloping portion of the spectral envelope play no role.

When subjects match the pitch of two signals falling in different spectral regions, systematic shifts may be found (Walliser, 1969; Ohgushi, 1978). Walliser (1969) found that the pitch of bandpass-filtered click trains containing harmonics above the third, was systematically lower than the pitch of a pure tone with frequency equal to $F0$. Comparable shifts can be seen as vertical offsets of the pitch shift lines for condition RES, for which test and matching stimuli fell in different spectral regions. However, it is the slopes of the pitch shift lines rather than the absolute values which are of interest here. Linear best-fitting lines were calculated separately for the NOISE and QUIET conditions, for each combination of $F0$ and resolvability. These lines were not constrained to pass through the origin.

To test the hypothesis that a pitch shift was present, t-tests were performed to examine whether the slopes of the best-fitting lines were significantly different from zero (Snedecor and Cochran, 1967; Howell, 1997). Zero slopes would be predicted if the envelope rather than temporal fine

TABLE II. Actual and predicted slopes for the NOISE condition of experiment 1. Predictions shown are based on the first effect of pitch shift with n taken for the lowest component in the flat region of the spectral envelope. An X in the last two columns indicates a line whose slope is *not* well predicted by the first effect, or by zero slope, respectively; a blank indicates a line whose slope falls within a 95% confidence interval around the comparison value.

$F0$ (Hz)	Resolvability	Actual slope	Pred. slope	Diff. from pred.	Diff. from zero
100	RES	0.245	0.25		X
100	INT	0.087	0.11		X
100	UNRES	−0.046	0.071	X	
200	RES	0.298	0.25		X
200	INT	0.122	0.11		X
200	UNRES	0.023	0.071		
400	RES	0.208	0.25		X
400	INT	0.118	0.11		X
400	UNRES	0.060	0.071		

structure determined pitch. The value of p used to calculate the critical t value was taken as 0.05/36, as 36 comparisons were being made (18 with zero and 18 with the predicted slopes; see below). Table II shows mean slopes and corresponding t-test results for the NOISE condition. For conditions RES and INT, slopes for all $F0$s were significantly different from zero. For condition UNRES, no slopes were significantly different from zero.

To test whether there was a second effect of pitch shift, the obtained slopes were compared with slopes predicted from the first effect, taking n as the rank of the lowest partial within the flat portion of the spectral envelope. Results are shown in Table II. The only slope that was significantly different from the predicted value was for condition UNRES for an $F0$ of 100 Hz. Here, the slope was less then predicted, i.e., opposite to what would be expected from the second effect of pitch shift. Thus, for conditions RES and INT, the first effect of pitch shift predicts the data well. The large errors associated with condition UNRES make it difficult to show any significant differences. However, the slope for condition UNRES with $F0=100$ Hz was significantly less than the predicted value. Coupled with the fact that no UNRES slopes were significantly different from zero, this suggests that, for condition UNRES, temporal envelope cues rather than fine structure cues were used.

Similar t-tests were performed for data from the QUIET condition (see Table III). The results were identical to those for the NOISE condition except that the slope for condition UNRES for $F0=200$ Hz, instead of $F0=100$ Hz, was significantly different from the predicted value. Slopes for the NOISE and QUIET conditions were similar for corresponding conditions and were equally well predicted. This suggests that for the RES and INT conditions, the addition of noise to mask combination tones had no effect.

C. Discussion

Previous research, reviewed in the Introduction, provided evidence that the pitch of complex tones with two or three harmonics is influenced by combination tones. Our results showed no effect of the noise on the pitch shifts for complex tones with at least five components. This suggests that the low pitch of these tones was not influenced by combination tones. This conclusion is consistent with our finding that the pitch shifts for the RES and INT conditions were well predicted from the frequency of the lowest component within the passband. It is also consistent with the findings of Patterson (1973) and Patterson and Wightman (1976) that the pitch shifts for their 6-component or 12-component complex tones could be predicted well from the frequencies of components within the complexes.

For the UNRES condition, no clear pitch shift was found. In this condition, the lowest nominal harmonic number for components in the flat region of the passband was 14. It is not surprising that no pitch shift was found for the 400-Hz $F0$, since the lowest nominal harmonic frequency in this case was 5600 Hz, which is above the range where phase locking is thought to occur (Palmer, 1995). However, for the lower $F0$s, the lowest nominal harmonic frequencies fell in the range where phase locking does occur, yet still no pitch shifts were observed. There have been few, if any, previous studies of pitch shifts using such high harmonic numbers. Schouten *et al.* (1962) reported pitch matches for three subjects using three-component stimuli whose lowest component corresponded to harmonic number 11, and Smoorenburg (1970) reported pitch matches for two-component stimuli whose lowest component corresponded to harmonic 14, although only one of his two subjects was able to make reliable matches to this stimulus. In both of these studies, pitch shifts were reported.

It is possible that, in the study of Schouten *et al.* (1962), subjects matched the spectral center of gravity or the excitation patterns of the stimuli in some conditions, rather than matching the low pitch. They were aware of the possibility that subjects might ''mistakenly match the pitch of two pure tones rather than that of the residues,'' but they did not consider the possibility that subjects matched the excitation patterns. They did conduct some limited control trials, using nonoverlapping partials for the test and matching stimuli, and found for one subject the same results as for overlapping components ''within the experimental error.'' However, such tests were not conducted for stimuli with very high harmonic numbers (in their most extreme condition, the lowest harmonic in the test stimulus had $n=10$ and the lowest harmonic in the matching stimulus had $n=9$). It remains possible, therefore, that the matches for stimuli with high

TABLE III. Actual and predicted slopes for the QUIET condition of experiment 1. Otherwise, as Table II.

$F0$ (Hz)	Resolvability	Actual slope	Pred. slope	Diff. from pred.	Diff. from zero
100	RES	0.224	0.25		X
100	INT	0.111	0.11		X
100	UNRES	0.002	0.071		
200	RES	0.266	0.25		X
200	INT	0.117	0.11		X
200	UNRES	0.021	0.071	X	
400	RES	0.248	0.25		X
400	INT	0.079	0.11		X
400	UNRES	0.043	0.071		

harmonic numbers were partially based on matching the excitation patterns of the stimuli.

Smoorenburg (1970) tried to prevent subjects from matching the frequencies of individual components or the excitation patterns by including conditions using test and matching stimuli with different harmonic numbers. However, the data presented were averaged across conditions where the harmonic numbers of the test and matching stimuli were the same and were different. Also, the matches were made in a sequence starting with a harmonic test stimulus, and progressively introducing larger frequency shifts. This may have led subjects to anticipate shifts in the expected direction, especially as the subjects were not naive to the theoretical background of the study.

In summary, two main conclusions can be drawn from the results. First, for complex tones with at least five harmonics, the pitch is not influenced by combination tones. Second, for complex tones with only high unresolved components (condition UNRES), there was no significant pitch shift with Δf, suggesting that pitch was based on envelope cues rather than on temporal fine structure.

We wished to investigate further the difference between our results and those of Schouten *et al.* (1962), so in experiment 2 we used stimuli similar to those of Schouten *et al.* (1962), in which the excitation patterns of the test stimuli did alter with frequency shift.

III. EXPERIMENT 2. "CLASSICAL" PITCH MATCHING TASK

A. Method

1. Stimuli

The test complex was derived from three components of a 50-, 100-, 200-, or 400-Hz $F0$. For each $F0$, three sets of nominal harmonic numbers were used: 3, 4, and 5 (resolved, RES), 8, 9, and 10 (intermediate, INT) and 13, 14, and 15 (unresolved, UNRES). The value of Δf was varied from 0% to 48% of $F0$ in steps of 8%. Larger ranges of $F0$ and frequency shift were used in this experiment because preliminary trials showed that reliable pitch matches could be obtained over these ranges. The matching complex consisted of three harmonics of an $F0$ controlled by the subject. The harmonic numbers were the same as the nominal values of the test complex. The component starting phase was cosine, and the middle component amplitude was twice that of the flanking components (corresponding to 100% AM). The timing and levels of the stimuli were the same as for experiment 1. Stimuli were generated in the same way as before, except that continuous noise (see Table I) was always presented and was generated using a white-noise source (IHR-PWNG) and four Kemo VBF 8/04 filters. The upper band of noise was generated by multiplying a low-pass noise by a sinusoid at the center frequency of the upper band. The slope on the upper edge of the noise band closest in frequency to the complex for the INT and UNRES conditions was at least 900 dB/oct. For the RES condition, for which a single low-pass noise was used, the slope was 180 dB/oct.

The noise was used in this experiment for two reasons. First, since the stimuli contained only three harmonics, it was more likely that combination tones would play a role, and we wished to explore the pitch shifts that would occur without the influence of these combination tones. Second, when only a few harmonics are present, noise seems to promote a synthetic mode of listening, helping to keep attention focused on the low pitch, rather than the pitches of individual components (Houtgast, 1976; Hall and Peters, 1981).

2. Subjects

Six normal-hearing subjects were tested, four musically trained and two not musically trained, with ages ranging from 19–24 years old. All had absolute thresholds better than 20 dB HL at the standard audiometric frequencies. One subject (author GM) also took part in experiment 1.

3. Procedure

The procedure was similar to that of experiment 1. The starting $F0$ step size of 2% was reduced to 1% after two reversals and to 0.5% after three reversals. Blocks consisted of seven consecutive trials (one for each shift) for a given $F0$ and resolvability, and the order of block presentation was randomized. For each combination of $F0$, resolvability, and shift, three pitch matches were recorded in different blocks.

B. Results

Results were similar for subjects GM and LD, and the mean results for these two subjects are plotted in Fig. 4. Results for subjects ST, SL, TA, and RA were similar, and their mean results are plotted in Fig. 5. Pitch shifts were found for all $F0$s and resolvabilities. The variability of pitch matching was much smaller than in experiment 1. The average s.d. in experiment 1, expressed as a percentage of the

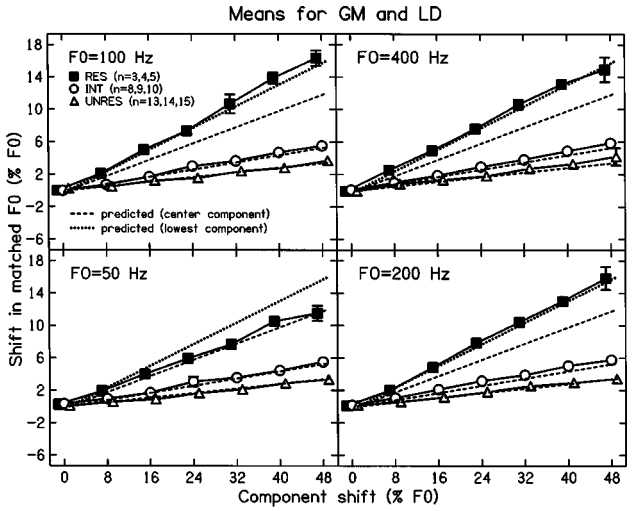


FIG. 4. Mean results of experiment 2 for subjects GM and LD. Dashed lines show predicted matches taking *n* as the center component. The heavy dotted lines show predictions taking *n* as the bottom component. Otherwise as Fig. 2.

unshifted $F0$, was 1.05 for condition NOISE, and 0.95 for condition QUIET; the average s.d. for comparable frequency shifts in experiment 2 was only 0.28.

Linear best fits were calculated for each combination of $F0$ and resolvability for the data combined across all subjects. As before, *t*-tests were performed to test the hypothesis that a pitch shift occurred. The value of *p* used to calculate the critical *t* value was 0.05/24, as 24 comparisons were being made. Table IV shows mean slopes and corresponding *t*-test results. All slopes were significantly different from zero. To assess whether there was a second effect of pitch shift, obtained slopes were compared with those predicted from the first effect, taking *n* as corresponding to the middle component. Results are shown in Table IV. The slopes for all conditions apart from $F0=100$ and 200 Hz-RES were not significantly different from those predicted from the first effect of pitch shift. Slopes *were* significantly greater than predicted for the 100- and 200-Hz $F0$ for the RES condition,

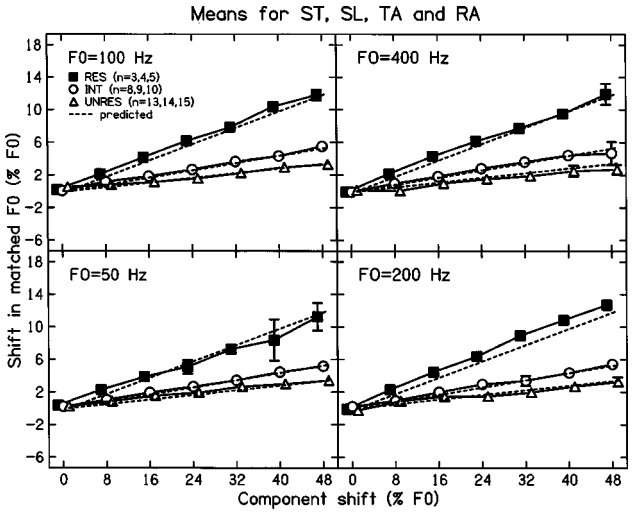


FIG. 5. Mean results of experiment 2 for subjects ST, SL, TA, and RA. Dashed lines show predicted matches taking *n* as the center component. Otherwise as Fig. 2.

but it is likely that this occurred because of the results of subjects GM and LD for these conditions. These subjects (Fig. 4) gave markedly different results than the other subjects for the RES condition with $F0$s of 100, 200, and 400 Hz. However, the slopes in the RES condition for GM and LD could be well predicted by taking *n* as corresponding to the lowest component ($n=3$). The lowest component was probably dominant for GM and LD, despite the fact that it had half the amplitude of the middle component (Moore *et al.*, 1984, 1985; Moore and Ohgushi, 1993). For the lowest $F0$ (50 Hz), harmonics above the lowest were probably dominant (Patterson and Wightman, 1976; Moore and Glasberg, 1988). This may explain why the results of GM and LD for this $F0$ were similar to those for the other subjects.

In summary, clear pitch shifts were found for all conditions including UNRES, which contrasts with the results for experiment 1, in which no shift was found in the UNRES condition. Pitch-matching variability was much lower than in experiment 1. No second effect of pitch shift was found when *n* was taken for the lowest or middle component.

IV. DISCUSSION

A. The role of temporal fine structure versus envelope cues for the UNRES condition

In experiment 1, the spectral envelope of the test and matching stimuli was held constant for a given condition. Thus, for the INT and UNRES conditions, changes in Δf or in the $F0$ of the matching stimulus had little effect on the excitation patterns of the stimuli. In experiment 1, shifts in the pitch matches were found for the INT condition but not for the UNRES condition. In experiment 2, the spectral envelopes of the test and matching stimuli changed with Δf or with the $F0$ of the matching stimulus. Such changes in spectral envelope were also a feature of most previous experiments involving pitch matches to frequency-shifted complex tones. Our results indicate that, for complex tones containing only very high partials, whose frequencies are closely spaced relative to the ERB of the auditory filter, the low pitch does not shift when the frequencies of the components are shifted, keeping the spectral envelope constant. This in turn implies that the low pitch is not based on time intervals between peaks in the temporal fine structure close to adjacent envelope maxima. Rather, the pitch seems to be based on the envelope repetition rate.

The shifts in the matches observed in experiment 2 for the UNRES condition, and perhaps in earlier experiments involving pitch matches to frequency-shifted complexes with high harmonic numbers (Schouten *et al.*, 1962; Patterson, 1973), were presumably produced by shifts in the spectral envelope of the test stimuli. Note that these shifts occur for complex tones with as many as six high partials (Patterson, 1973). In these experiments, subjects probably adjusted the $F0$ of the matching stimulus until its excitation pattern matched that of the test stimulus as closely as possible. Even in previous experiments where the test and matching stimuli fell in different spectral regions (Smoorenburg, 1970), the shift in excitation pattern of the test stimuli with frequency shift of the components may have influenced the results.

TABLE IV. Actual and predicted slopes from experiment 2. Predictions are based on the first effect of pitch shift with n taken for the middle component. Otherwise as Table II.

$F0$ (Hz)	Resolvability	Actual slope	Pred. slope	Diff. from pred.	Diff. from zero
50	RES	0.225	0.250		X
50	INT	0.105	0.111		X
50	UNRES	0.067	0.071		X
100	RES	0.280	0.250	X	X
100	INT	0.112	0.111		X
100	UNRES	0.064	0.071		X
200	RES	0.290	0.250	X	X
200	INT	0.111	0.111		X
200	UNRES	0.067	0.071		X
400	RES	0.268	0.250		X
400	INT	0.108	0.111		X
400	UNRES	0.068	0.071		X

Variability of pitch matching was much smaller in experiment 2 than in experiment 1. This also is in line with subjects using different cues in the two experiments. Excitation pattern cues may be less ambiguous than low pitch cues, especially for stimuli containing only a few high-frequency components, as was the case in experiment 2.

The question arises as to whether the pitch of complex tones is ever determined by the temporal fine structure evoked by the interference of harmonics on the basilar membrane. The results for the INT condition of experiment 1 throw some light on this issue. Shifts in pitch occurred for that condition, even though the spectral envelopes of the stimuli were held constant. For the unshifted test stimulus, the lowest harmonic within the flat region of the passband was the ninth (level$\approx$63 dB SPL), and the eighth and seventh harmonics were present in the sloping region of the spectral envelope with levels of about 55 and 45 dB SPL, respectively. When noise was present, it would have masked the seventh harmonic. Even the 55-dB eighth harmonic would have been partially or completely masked by the combined effects of the noise and the 63-dB ninth harmonic. It seems reasonable to assume that, when the noise was present, the components in the sloping region of the spectral envelope did not contribute to the perceived pitch. This is consistent with our finding that the pitch shifts could be predicted from the frequency of the lowest component within the flat region of the spectral envelope. This component was separated from its neighboring components by less than 0.9 ERB. Such a separation is probably not sufficient to allow resolution of that component (Plomp, 1964; Moore and Ohgushi, 1993). Higher components would be even less well resolved. Nevertheless, pitch shifts were observed for the frequency-shifted stimuli in the INT condition, even when noise was present.

We conclude that temporal fine structure information probably is used to determine low pitch for stimuli containing components which are just unresolved (and possibly for resolved components too). However, when the component frequencies are too high relative to $F0$ (nominal harmonic numbers all above the 13th), the temporal fine structure is not used, and pitch is determined by the envelope repetition rate. Consistent with this idea, the ability to discriminate changes in the $F0$ of harmonic complex tones is generally quite good when those tones contain harmonics with numbers in the range 5–10, but worsens when no harmonics below the tenth are present, and becomes very poor when only harmonics above the 20th are present (Ritsma and Hoekstra, 1974; Moore and Glasberg, 1988; Houtsma and Smurzynski, 1990; Moore and Peters, 1992). It appears that discrimination of $F0$ is good when time fine structure information can be extracted and poor when it cannot.

B. Why are combination tones unimportant for stimuli with many components?

Our results suggest that combination tones had no effect on the low pitch shift of our multicomponent complexes, whereas previous studies using stimuli with two or three components indicate that combination tones play a significant role. The most likely explanation for this difference is based on the idea that low pitch is derived by combining information from many partials, including resolved and unresolved partials and combination tones (Moore *et al.*, 1984, 1985; Houtsma and Smurzynski, 1990; Meddis and O'Mard, 1997; Moore, 2003). The contribution of any single partial or combination tone to the low pitch depends on how many other partials are present. When there are many partials, any single partial or combination tone would make only a small contribution to the low pitch. Essentially, the contribution of combination tones would be ''swamped'' by the contributions of other partials. When there are very few partials present, each one, including combination tones, may make a substantial contribution to the low pitch.

V. CONCLUSIONS

(1) Masking combination tones had no effect on the low pitch shift of our multicomponent complexes. Low pitch could be predicted on the basis of the frequencies of components within the complex. Together, these findings suggest that combination tones do not influence the pitch of complex tones with an overall level of 70 dB SPL and with five or more harmonics.
(2) When the spectral envelope of the test and matching stimuli was held constant, no pitch shifts were observed for the UNRES condition (experiment 1). This suggests that temporal envelope rather than fine-structure cues

mediate low pitch when partials in a complex are high relative to $F0$ (harmonic numbers all above the 13th).

(3) Pitch shifts were observed in the UNRES condition when the spectral envelope of the test stimuli varied with Δf, and the spectral envelope of the matching stimuli varied with $F0$ (experiment 2). The shifts in this case may have been produced by subjects matching the excitation patterns of the stimuli, rather than by matching low pitch. Matches based on excitation patterns may also have influenced the results of some earlier studies, e.g., that of Schouten *et al.* (1962).

ACKNOWLEDGMENTS

This work was supported by a studentship from Defeating Deafness (author G.A.M.) and by the Medical Research Council (UK). We thank Thomas Baer and Brian Glasberg for their help, and three anonymous reviewers for useful comments.

Cariani, P. A., and Delgutte, B. (**1996a**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Cariani, P. A., 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**, 1717–1734.

Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). ''A spectral network model of pitch perception,'' J. Acoust. Soc. Am. **98**, 862–879.

de Boer, E. (**1956**). ''On the residue in hearing,'' Ph.D. thesis, University of Amsterdam.

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

Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.

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

Houtgast, T. (**1976**). ''Subharmonic pitches of a pure tone at low S/N ratio,'' J. Acoust. Soc. Am. **60**, 405–409.

Houtsma, A. J. M., and Fleuren, J. F. M. (**1991**). ''Analytic and synthetic pitch of two-tone complexes,'' J. Acoust. Soc. Am. **90**, 1674–1676.

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

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.

Howell, D. C. (**1997**). *Statistical Methods for Psychology*, 4th ed. (Duxbury, Belmont, CA).

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.

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

Moore, B. C. J., and Glasberg, B. R. (**1988**). ''Effects of the relative phase of the components on the pitch discrimination of complex tones by subjects with unilateral and bilateral cochlear impairments,'' in *Basic Issues in Hearing*, edited by H. Duifhuis, H. Wit, and J. Horst (Academic, London).

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

Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). ''Frequency and intensity difference limens for harmonics within complex tones,'' J. Acoust. Soc. Am. **75**, 550–561.

Moore, B. C. J., and Ohgushi, K. (**1993**). ''Audibility of partials in inharmonic complex tones,'' J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., and Peters, R. W. (**1992**). ''Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity,'' J. Acoust. Soc. Am. **91**, 2881–2893.

Ohgushi, K. (**1978**). ''On the role of spatial and temporal cues in the perception of the pitch of complex tones,'' J. Acoust. Soc. Am. **64**, 764–771.

Palmer, A. R. (**1995**). ''Neural signal processing,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Patterson, R. D. (**1973**). ''The effects of relative phase and the number of components on residue pitch,'' J. Acoust. Soc. Am. **53**, 1565–1572.

Patterson, R. D., and Wightman, F. L. (**1976**). ''Residue pitch as a function of component spacing,'' J. Acoust. Soc. Am. **59**, 1450–1459.

Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.

Pressnitzer, D., and Patterson, R. D. (**2001**). ''Distortion products and the pitch of harmonic complex tones,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht).

Ritsma, R. J. (**1962**). ''Existence region of the tonal residue. I.,'' J. Acoust. Soc. Am. **34**, 1224–1229.

Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.

Ritsma, R. J. (**1970**). ''Periodicity detection,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).

Ritsma, R. J., and Hoekstra, A. (**1974**). ''Frequency selectivity and the tonal residue,'' in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin).

Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418–1424.

Smoorenburg, G. F. (**1970**). ''Pitch perception of two-frequency stimuli,'' J. Acoust. Soc. Am. **48**, 924–941.

Snedecor, G. W., and Cochran, W. G. (**1967**). *Statistical Methods* (Iowa University Press, Ames, IA).

Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.

Walliser, K. (**1969**). ''Zusammenhänge zwischen dem Schallreiz und der Periodentonhöhe,'' Acustica **21**, 319–328.

Yost, W. A. (**1996**). ''Pitch of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 511–518.

Modulation rate discrimination for unresolved components: Temporal cues related to fine structure and envelope

Joseph W. Hall III,[a] Emily Buss, and John H. Grose
Department of Otolaryngology Head and Neck Surgery, University of North Carolina Medical School, 610 Burnett-Womack Building, Campus Box 7070, Chapel Hill, North Carolina 27599-7070

(Received 17 January 2002; revised 22 October 2002; accepted 4 November 2002)

The present study investigated the hypothesis that the cues for modulation rate discrimination for unresolved spectral components differ as a function of the spectral region occupied by the stimuli. Specifically, it was hypothesized that when components occupy relatively low spectral regions, phase locking both to the fine structure and to the envelope are useful cues. However, as the spectral region occupied by the components increases, phase locking to the fine structure becomes less robust, whereas phase locking to the envelope remains as a potentially strong cue. Observers were asked to detect a decrease in modulation rate for carrier frequencies between 1500 and 6000 Hz. Both amplitude-modulated (AM) and quasifrequency-modulated (QFM) tones were used in order to produce stimuli having strong and weak envelope cues, respectively. Although there were marked individual differences, the results showed an interaction between modulation type and spectral region, with AM and QFM performance being relatively similar at low spectral region, but with QFM showing a steeper reduction in performance as the spectral region of the carrier frequency increased. Overall, the data are consistent with an interpretation that pitch perception for unresolved components depends upon both fine structure and envelope cues, and that the relative importance of these cues depends upon the spectral region occupied by the stimuli. [DOI: 10.1121/1.1532004]

PACS numbers: 43.66.Hg, 43.66.Lj [NFV]

I. INTRODUCTION

Harmonic complex stimuli are typically perceived as having a pitch associated with the fundamental frequency, even when the fundamental is not actually present (Seebeck, 1841). When resolved harmonics of a fundamental frequency are present, they dominate the perceived pitch (Ritsma, 1967). In such cases, pitch perception may be determined either by spectral cues, temporal cues, or by some combination of the two (Licklider, 1951; Schouten, 1970; Goldstein, 1973; Terhardt, 1974; Moore, 1980). Even though pitch perception for complex stimuli is dominated by low, resolved harmonics, a relatively weak low pitch can sometimes still be perceived even when only high, unresolved harmonics are present (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001). Because the harmonics are not spectrally resolved in such cases, it is assumed that the basis of the perceived pitch is temporal (Houtsma and Smurzynski, 1990). The present study is concerned with pitch perception for unresolved spectral components, and focuses on two temporal cue types that may be important: those related to the temporal fine structure of the waveform, and those related to the temporal envelope of the waveform.

Temporally based models of pitch perception require an assumption that repeating temporal patterns are somehow registered by the auditory system. Although the neural bases for the derivation of temporally based pitch are not presently known, autocorrelation, a mathematical function sensitive to repeating temporal patterns, often forms the basis of temporal pitch models (Licklider, 1959; Cariani and Delgutte, 1996b, 1996a; Meddis and O'Mard, 1997). One prerequisite for such a process is the availability of neural representation of timing information related to the stimulus. There is physiological evidence that both phase locking to the temporal fine structure (Kiang *et al.*, 1965; Rose *et al.*, 1967) and to the temporal envelope (Javel, 1980; Palmer, 1982; Joris and Yin, 1992) of the stimulus occur at the level of the VIIIth nerve. Both of these cues may be relevant to the pitch of unresolved spectral components. Animal neurophysiology indicates that phase locking to temporal fine structure has a low-pass characteristic. For example, in the cat, phase locking declines monotonically above about 1 kHz, becoming negligible above 4–5 kHz (Johnson, 1980). Relatively little is known about this function in the human auditory system. It would seem to be a reasonable assumption that when unresolved components occur in frequency regions where phase locking to fine structure is present, repeating temporal patterns related to fine structure may be used to extract pitch.

When harmonic stimuli are unresolved, interactions among harmonics can result in a temporal envelope that has the same period as that of the fundamental frequency. Physiological studies indicate neural phase locking to waveform envelope frequencies below about 1000 Hz (Joris and Yin, 1992), with this cutoff probably due, at least in part, to cochlear filtering (Palmer, 1982; Joris and Yin, 1992). Phase locking to envelope occurs most robustly in relatively high-frequency regions where phase locking to fine structure may be minimal or absent. Thus, repeating patterns related to temporal envelope may also contribute to pitch perception for unresolved components.

[a]Electronic mail: jwh@med.unc.edu

The notion that the pitch of unresolved components might be influenced by temporal cues arising from two sources (those related to fine structure and those related to envelope) raises some interesting issues. One is that the type of temporal cue underlying pitch perception for unresolved components may be quite different, depending upon the spectral location of the components: for unresolved components in relatively low spectral regions, pitch may be based both on temporal fine structure and temporal envelope cues; for unresolved components in relatively high spectral regions (where phase locking to fine structure is impoverished or absent) pitch may be based primarily or completely on temporal envelope cues. This raises the possibility that the effect of the phase relation among components may be quite different for unresolved components in low versus high spectral regions. For example, at high spectral regions, where phase locking to fine structure may be poor and pitch perception may depend primarily on temporal envelope, phase manipulations that reduce the availability of useful envelope cues may result in poor pitch discrimination. This may not be the case at low spectral regions where phase locking to fine structure is available. Here, even when envelope cues are diminished, repeating patterns of encoded temporal fine structure might provide adequate cues for pitch discrimination.

The present study examined this hypothesis by investigating modulation frequency discrimination for amplitude-modulated (AM) and quasifrequency-modulated (QFM) pure tones. Although modulation rate discrimination is not a direct measure of pitch perception, similar stimuli have been used in many previous experiments to gain insights into mechanisms accounting for pitch perception (e.g., Ritsma, 1962; van den Brink, 1970; Wightman, 1973; Hall and Soderquist, 1975). In the present study, the modulated tones comprised a set of spectrally narrow stimuli, allowing restricted frequency regions to be tested. The starting phases of components were manipulated to produce different depths of envelope modulation, allowing a degree of control over this stimulus variable. The frequencies used here were chosen such that spectral components were always unresolvable. In order to manipulate the usefulness of envelope cues, modulation was either AM (salient envelope cue) or QFM (reduced envelope cue). The hypothesis was, therefore, that performance for AM and QFM would be similar at low carrier frequency (where both fine structure and envelope cues were potentially useful), but that performance for QFM would be poorer than for AM at higher carrier frequencies (where envelope cues would be available but fine structure cues would be reduced or eliminated).

II. METHOD

A. Observers

Five observers (two male and three female) participated. Hearing thresholds at octave frequencies from 250 Hz to 8 kHz were better than 20 dB HL for all ears tested (ANSI, 1996). Observer age ranged between 18 and 50 years, and all observers were experienced in psychoacoustical tasks.

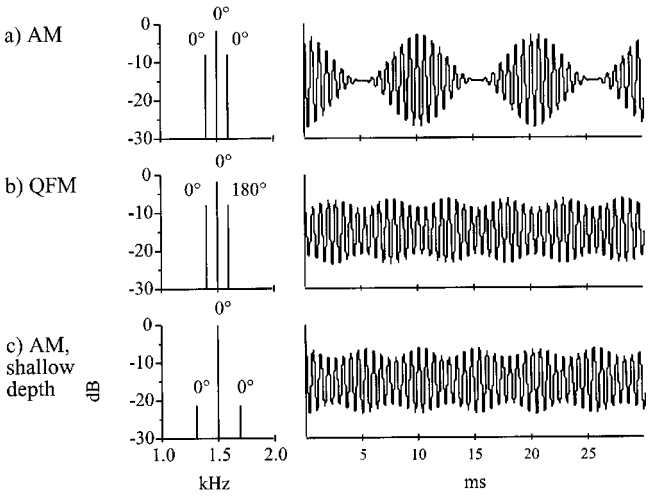


FIG. 1. Example stimuli for the 100-Hz rate conditions, with center frequency of 1.5 kHz, are shown in the frequency domain (left column) and the time domain (right column). Component starting phase is indicated above each component in the frequency domain panel. The AM (a) and QFM (b) conditions differed only in starting phase of the highest signal component. Stimuli in the shallow-AM condition (c), an additional condition introduced in the results section, differed from the SAM condition in the frequency separation between the central and flanking components (200- and 100-Hz spacing, respectively) and the relative amplitudes of the central and flanking components.

B. Stimuli

Detection of a decrease in modulation frequency was examined for unresolved three-component stimuli (AM and QFM), as a function of the spectral region of the carrier. The general spectral and temporal properties of the stimuli are summarized in Fig. 1. Stimuli were generated in the time domain as the sum of three pure tones. One tone was at the carrier frequency, and the remaining two tones were at equal frequency spacings above and below the carrier frequency, with the spacing defining the modulation rate. The flanking tones were half the amplitude of the carrier tone (-6 dB). For the AM conditions, all three tones were in sine starting phase, whereas for the QFM condition, the high-frequency flanking tone was phase shifted by 180° relative to sine starting phase. The bottom panel of Fig. 1 shows an AM stimulus that was designed to provide envelope cues comparable to those associated with the QFM stimulus. This condition will be discussed below. Stimuli were digitally generated, played at a rate of 20 kHz, antialias filtered (Kemo VBF8) at 8 kHz, and delivered monaurally through Sony MDR V6 earphones.

Detection of a decrease in modulation frequency (i.e., decrease in frequency spacing of the component tones) was determined for two standard modulation frequencies, 100 and 200 Hz. Performance was determined at fixed carrier frequencies. The rationale for maintaining a fixed carrier frequency was that cues related to frequency shifts in the excitation pattern are minimized. For the 100-Hz rate, the carrier frequencies examined were 1500, 2000, 3000, 4000, and 6000 Hz. With this stimulus set, the lowest component frequency was 1400 Hz. Because it is usually assumed that components are not resolved past the 10th harmonic (Houtsma and Smurzynski, 1990), it is very unlikely that any of

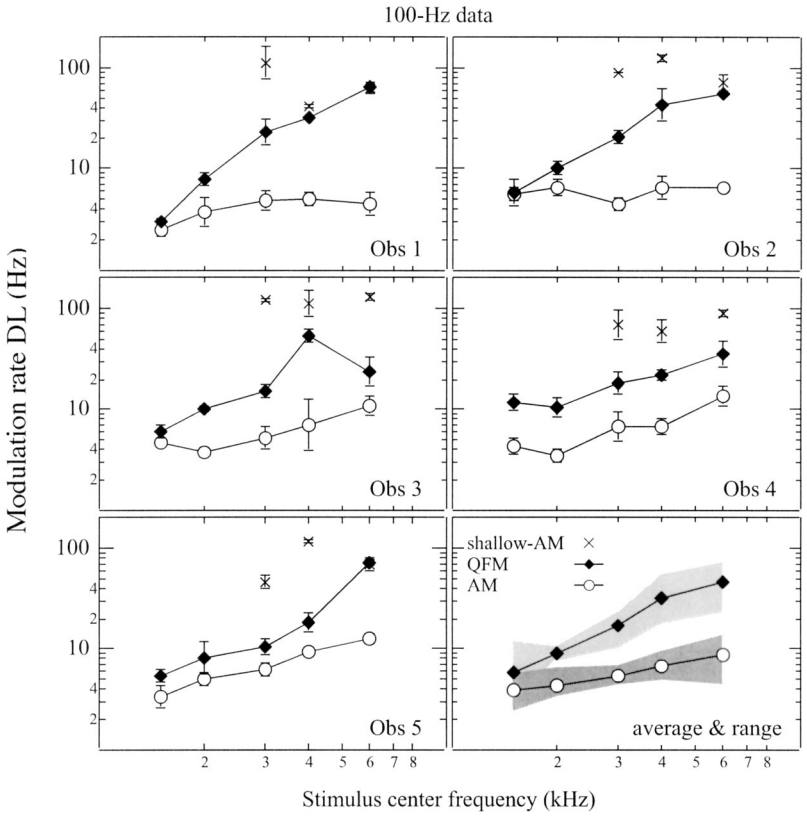


FIG. 2. Rate discrimination thresholds for the AM (open circles) and QFM (filled diamonds) conditions are plotted as a function of stimulus center frequency for the 100-Hz rate of modulation. Mean data are plotted separately for each observer, with error bars representing the standard deviation for each point. The mean across observers appears in the lower right panel, with shaded areas indicating the minimum and maximum threshold at each frequency across the five observers. Data points for the 200-Hz shallow-AM condition, described in the results section, are also included in the individual data panels (marked by the "x" symbol).

the stimuli in this set were made up of resolvable components. For the 200-Hz rate, the carrier frequencies examined were 3000, 4000, and 6000 Hz. Again, the stimuli in this set were made up of components that were unlikely to be resolvable. A continuous Gaussian noise low-pass filtered at 80% of the lowest component frequency of the standard was presented at 15-dB spectrum level to limit the availability of distortion product cues. The stimuli for modulation frequency discrimination were presented at a level of 25 dB SL with a level rove of plus and minus 5 dB around that level. The rove in level was used in order to reduce the utility of cues related to loudness or excitation pattern.

C. Procedure

The 79.4% modulation frequency difference threshold was estimated using a three-alternative, forced-choice, three-down one-up adaptive procedure (Levitt, 1971). In two of the three observation intervals, the modulation frequency was the same (100 or 200 Hz). In the remaining interval (chosen at random) the modulation frequency was lower than that of the standard. The initial difference between the modulation frequency of the standard and the modulation frequency of the target (Δf) was relatively large (e.g., 25–50 Hz), but when listeners had completed runs at a given condition, retests were started at a Δf that was a factor of 2 times the previous threshold for the condition. All stimuli were 400 ms in duration with 20-ms cosine-squared ramps and were separated by 300-ms interstimulus intervals. After three correct responses in succession, the Δf was decreased; after one incorrect response, Δf was increased. The Δf was initially changed by a multiplicative factor of 1.5. This factor was changed to 1.25 after the first reversal and 1.125 after the second reversal. If this adjustment for the 100-Hz standard ever resulted in a Δf that was more than 85 Hz, the current Δf was set to 85 Hz (this criterion was 185 Hz for the 200-Hz standard). A threshold run was stopped after six reversals in direction, and the geometric mean of the rate at the final four reversals was taken as the threshold for the run. Intervals were marked visually with 400-ms observation lights, and visual feedback was provided after each response. A block of at least five threshold estimates was obtained in each condition, with conditions completed in random order. The geometric mean of the estimates obtained was taken as representative of the set. After all thresholds for all conditions had been completed, an additional threshold was taken for each condition. If the additional threshold obtained was better than all of those obtained previously for that condition, five replacement thresholds were obtained and their geometric mean was taken as representative for the condition.

The 0-dB SL determination for the stimuli was obtained by first acquiring detection thresholds for each of the carrier frequencies. The threshold estimation procedure was a three-alternative, forced-choice, three-down one-up adaptive strategy. Signal level was adjusted in steps of 4, 2, and 1 dB for the first, second, and remaining reversals. Timing variables and the presence of the low-pass noise were the same as for the modulation discrimination conditions above. Three threshold estimates were averaged to determine 0 dB SL for each carrier frequency.

III. RESULTS AND DISCUSSION

The individual and average data across observers are summarized in Fig. 2 (100-Hz modulation) and Fig. 3

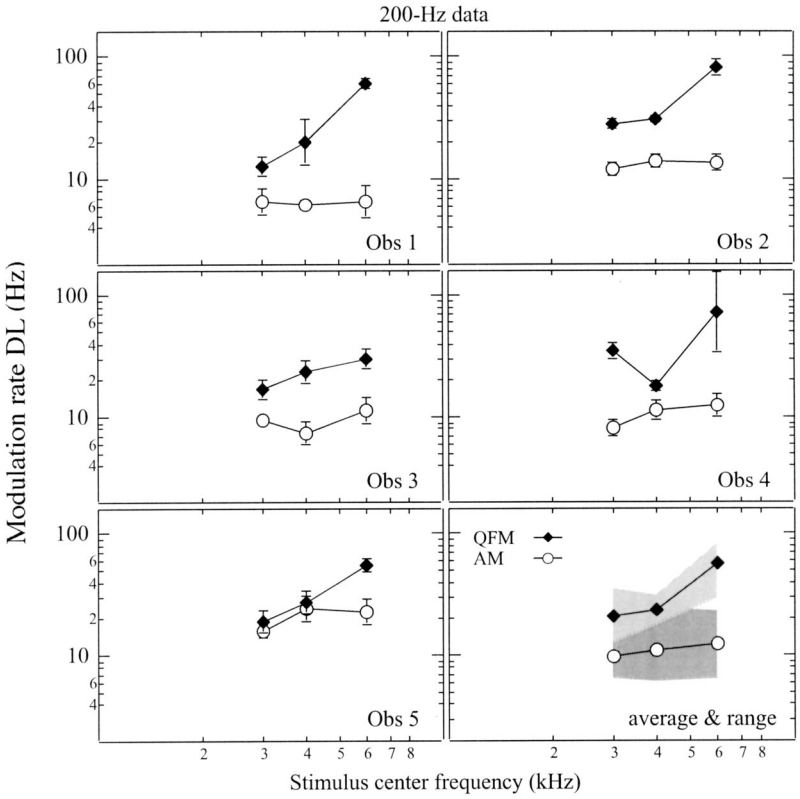


FIG. 3. Rate discrimination thresholds for the AM (open circles) and QFM (filled diamonds) conditions are plotted as a function of stimulus center frequency for the 200-Hz rate of modulation. Mean data are plotted separately for each observer, with error bars representing the standard deviation for each point. The mean across observers appears in the lower right panel, with shaded areas indicating the minimum and maximum threshold at each frequency across the five observers.

(200-Hz modulation). Some aspects of the results were consistent with the expectation that differences in performance between AM and QFM would be larger at higher carrier frequencies (where phase locking to fine structure was less likely to be appreciable) than at lower carrier frequencies (where phase locking to fine structure was more likely to be appreciable). For example, in the average data it can be seen that the AM and QFM functions diverge with increasing carrier frequency. Repeated measures analyses of variance were performed to examine the reliability of these trends. For the 100-Hz modulation rate, there was a significant effect of type of modulation (AM versus QFM) [$F(1,4) = 121.7$; $p = 0.0004$], carrier frequency [$F(4,16) = 33.1$; $p < 0.0001$], and a significant interaction between modulation type and carrier frequency [$F(4,16) = 5.1$; $p = 0.008$]. For the 200-Hz modulation rate, there was again a significant effect of type of modulation [$F(1,4) = 42.1$; $p = 0.003$], carrier frequency [$F(2,8) = 26.0$; $p = 0.0003$], and a significant interaction between modulation type and carrier frequency [$F(2,8) = 5.6$; $p = 0.03$]. Of particular interest here were the significant interactions, reflecting the divergence of the AM and QFM functions with increasing carrier frequency.

While these results reflect modulation rate discrimination rather than rate pitch discrimination *per se*, subjective reports suggest that a pitch cue was used for both modulation types at the low-frequency carriers and for AM at the high-frequency carriers. After completion of the experiment, observers 2, 3, and 5 listened to one additional threshold estimation track of the 100-Hz rate AM and QFM conditions for the 1500- and 6000-Hz carrier frequencies. After each track they were asked to describe the cue they used to identify the signal interval. The observers described the cue for both the AM and the QFM stimuli at 1500 Hz as being based upon a low pitch, with the lower rate of modulation being associated with the lower pitch. In contrast, at 6000 Hz only the cue for the AM stimulus was described in terms of a low pitch; the QFM cue was described as ''wobbly'' or ''less smooth,'' consistent with the idea that the QFM stimulus was not associated with a perceptible low pitch at high carrier frequency.

The present study employed unresolved components presented at relatively low levels in order to increase the likelihood that performance would be based upon temporal cues rather than upon cues related to the spectral domain. One way of assessing the possible contribution of spectral cues in modulation rate discrimination is by examining the associated excitation patterns (Micheyl *et al.*, 1998). Using the formulas suggested by Moore and Glasberg (1983), we examined excitation patterns associated with our 100-Hz AM stimuli centered on the lowest and highest carrier frequencies (1500 Hz and 6000 Hz). In each case, a range of sideband separations was tested: the standard 100-Hz spacing, as well as 95-, 90-, 80-, 60-, and 20-Hz-spaced signal stimuli. The level of the stimuli was computed as 25 dB above the average thresholds for our observers (7.8 and 13.8 dB SPL for 1500 and 6000 Hz, respectively). The magnitude of the spectral cue was calculated by subtracting the excitation pattern for each of the stimuli from the excitation pattern for the standard stimulus. Only points in the excitation pattern that were less than 25 dB down from the stimulus level (and were therefore likely to be audible) were considered in the calculations. The differences (in dB) between the excitation patterns are summarized in Table I. Inspection of the table shows that the maximum excitation pattern differences associated with frequency spacings near AM and QFM discrimination thresholds were generally less than 1 dB. Given that there was a 10-dB level rove associated with stimulus pre-

TABLE I. Largest difference (in dB) between the excitation pattern for the 100-Hz rate of modulation and lesser rates of modulation. Values are shown for the 1500-Hz carrier and for the 6000-Hz carrier (see the text).

	Frequency separation				
CF	95 Hz	90 Hz	80 Hz	60 Hz	20 Hz
1500 Hz	0.8	1.2	1.7	2.5	3.5
6000 Hz	0.1	0.1	0.2	0.3	0.4

sentation, it seems quite unlikely that excitation pattern cues were utilized by the observers.

Although the present study attempted to exploit the fact that the QFM stimulus possesses a relatively impoverished temporal envelope, it is important to consider how the envelope might be affected by peripheral auditory processing. For example, if some of the stimuli used here had components that were at least partially resolvable, and if the listener used an auditory filter that was centered below the carrier frequency, an increased modulation depth could be obtained for the QFM stimuli (e.g., Goldstein, 1967). Under such a circumstance, strong envelope cues might be available for both AM and QFM, perhaps accounting for the finding that performance for the 100-Hz modulation at the 1500-Hz carrier frequency was similar between AM and QFM. The 1500-Hz carrier was the lowest used for the 100-Hz modulation rate, and, thus, would be associated with the best possibility for some spectral resolution. Whereas this account cannot be ruled out, it would not seem to be compatible with a comparison of the 1500-Hz carrier, 100-Hz modulation rate with the 3000-Hz carrier, 200-Hz modulation rate. Although these conditions would be expected to be similar in terms of component resolvability, the ratio of QFM performance to AM performance was relatively higher at the 3000-Hz carrier (a ratio of approximately 2.2) than at the 1500-Hz carrier (a ratio of approximately 1.4). Whereas this difference in ratio is compatible with an explanation in terms of reduced phase locking to fine structure, an interpretation in terms of interaction between partially resolved spectral components would call for a similar ratio at the two carrier frequencies.

It should be pointed out that there were individual differences in the data, particularly for the 100-Hz modulation frequency. For example, observer 3 actually showed an improvement in performance for QFM between the 4- and 6-kHz carrier frequencies. This is difficult to account for, but it is unlikely that it reflects an improvement in phase locking between 4 and 6 kHz. Other individual differences for the 100-Hz modulation rate concern the carrier frequency at which AM and QFM functions diverged. For observers 1, 2, and 3, divergence occurred as the carrier increased above 1500 Hz. For observer 5, however, divergence of the functions occurred above 4 kHz, and for observer 4, the functions did not diverge at all. Some of the variance across observers may be related to individual differences in the ability to use envelope cues at high carrier frequencies. For example, observers 1 and 2 maintained relatively stable performance for AM with increasing carrier frequency, whereas their performance for QFM worsened with increasing carrier frequency. This would be consistent with an interpretation that these observers were able to use the envelope cues very effectively at high carrier frequencies. The divergence between AM and QFM functions with increasing carrier frequency was the most marked for these observers. For the other observers, the AM functions were not constant with increasing carrier frequency, but instead showed a modest decline in sensitivity. This is consistent with an interpretation that these observers were relatively inefficient in using temporal envelope cues at high carrier frequencies. For these observers, the divergence of the AM and QFM functions was less marked (or was absent).

As was pointed out in the Introduction, there was an expectation that the pattern of results obtained here would be influenced by both temporal fine structure and temporal envelope cues, and that the relative utility of the fine structure cues would depend upon carrier frequency. The fact that AM discrimination thresholds increased less steeply than QFM performance with increasing carrier frequency can be accounted for by assuming that robust envelope cues were still available when fine structure cues became impoverished at increasing carrier frequency. The worsening of QFM performance with increasing carrier frequency can be accounted for by assuming that envelope cues associated with the QFM stimuli were generally of insufficient quality to sustain a high level of performance as fine structure cues became impoverished with increasing carrier frequency.

As noted above, even though the QFM stimuli used here have reduced envelope salience in comparison to AM, the envelope for QFM nevertheless affords a potential performance cue. It is therefore possible that the initial steep decline in performance for QFM reflects the loss of the temporal fine structure cue, and that performance at the highest carrier frequencies reflects the use of an impoverished envelope cue. In an attempt to estimate the modulation rate DL that would be expected based on envelope cues alone, an AM stimulus with a shallow depth of modulation, equivalent to that of the QFM, was constructed. As shown in Fig. 1 the sidebands for this AM stimulus were 21.43 dB lower than the carrier and were separated from the carrier by 200 Hz, producing a 200-Hz envelope rate similar to that of the 100-Hz QFM stimulus. Because of this wider spacing, no attempt was made to obtain thresholds at frequencies below 3000 Hz, where the standard stimulus was comprised of the 14–16th harmonics of 200 Hz. Although not visually apparent in the time-domain representation in Fig. 1, the fine structure of the QFM stimulus is frequency modulated, while the fine structure for the shallow-AM condition is of constant rate through the modulation period. In order to determine the relative availability of fine structure cues in the QFM and shallow-AM stimuli, autocorrelation simulations were undertaken. Figure 4 shows summary autocorrelation functions (SACFs) produced by the AMS model (Meddis and Hewitt, 1991), parameters based upon Nuttal and Dolan (1996), Summer *et al.* (2002), and Lopez-Poveda and Meddis (2001). These functions are shown for AM, QFM, and shallow-AM stimuli centered on 1500, 3000, and 6000 Hz. No data were collected for the 1500–Hz shallow-AM conditions because the component tones would have fallen in the region of resolvability, but the summary autocorrelation

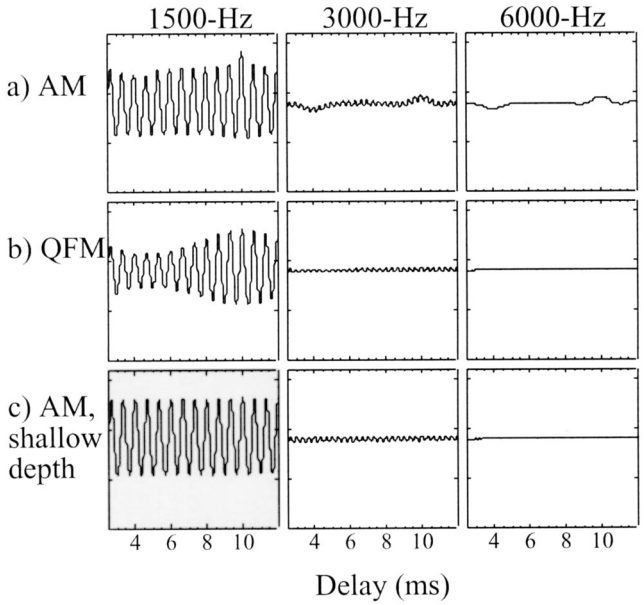


FIG. 4. Summary autocorrelation functions for three stimulus conditions (rows) and three center frequencies (columns). These were computed as the sum of autocorrelation functions across the output of a range of auditory filters following a simulation of peripheral encoding.

function is shown here (lower left panel of Fig. 4) for comparison. The peak at 10-ms delay seen in some of the panels represents 100-Hz periodicity of the encoded stimulus, after simulated peripheral encoding. The difference in the autocorrelation patterns across conditions leads to a prediction that the fine structure of the low carrier frequency QFM should provide a more robust cue for pitch discrimination than the shallow-AM. Differences in threshold would be consistent with the use of this fine structure information of the QFM stimuli, particularly in light of the fact that the envelopes should provide comparable pitch cues for both conditions.

Psychophysical results for the shallow-AM condition are shown as crosses in Fig 2. Whereas these data were highly variable across observers and across replicate threshold estimates within observer, it was generally the case that the shallow-AM condition was associated with poorer sensitivity than QFM at the 3000-Hz frequency, with thresholds then converging with QFM data at the higher frequencies. While the variability makes interpretation difficult, these results are generally consistent with the idea that residual fine structure cues are present in the QFM condition at 3000.

The AMS model was also used to compare modeled and obtained results for a subset of the stimuli used in this experiment, specifically AM, QFM, and shallow-AM stimuli centered on 1.5, 3, and 6 kHz. The variance in the SACFs (*re*: a 100-Hz standard) was computed for a range of modulation rates: 20, 40, 60, 80, 90, 95, and 97.5 Hz. Parameters for this model were again based on Nuttal and Dolan (1996), Sumner *et al.* (2002), and Lopez-Poveda and Meddis (2001). The rms difference in SACFs (the squared Euclidean distance between the two SACFs) can be used to model the discriminability of stimuli that differ in terms of perceived pitch (Meddis and O'Mard, 1997). In Fig. 5, rms difference in the SACFs is plotted as a function of the target modulation rate, with line type corresponding to modulation type. Also indicated are geometric means of the psychophysical data for the AM and QFM stimuli at each center frequency, plotted at the ordinate value corresponding to the model output for that stimulus. Some of the modeled results are broadly consistent with the results obtained, in that (1) thresholds in the AM condition are relatively unchanged by changes in stimulus center frequency (although there are slight variations in estimated cue strength at threshold); (2) thresholds in the QFM condition deteriorate radically with increases in stimulus center frequency; and (3) the shallow-AM condition is consistently associated with the poorest discrimination sensitivity. There are some aspects of these results that are not consistent with the data, however. Perhaps most salient is the fact that discrimination threshold for the QFM 3-kHz stimulus is associated with a predicted cue strength that would appear to be insufficient to support discrimination. It is possible that this discrepancy would be ameliorated by altering parameters of the model associated with phase locking (in effect, using a higher cutoff for the low-pass filter function). We also note that the predicted cue strength for QFM and shallow-AM at 6 kHz never attains a material value, regardless of modulation rate. By this indication, QFM and shallow-AM stimuli should not have produced a threshold within the range of modulation rates tested, while in fact the geometric mean for the QFM threshold fell at approximately 50 Hz. This discrepancy is less troublesome in light of the fact that observer reports suggested that pitch was not the detection cue used in these conditions (see above).

This is not the first study that has used human performance on a psychoacoustical test to make inferences about the frequency regions for which phase locking contributes to auditory performance. For example, Hartmann *et al.* (1990) argued that performance on a task involving pitch matching of a mistuned harmonic could be accounted for in terms of phase locking to temporal fine structure. Their data indicated

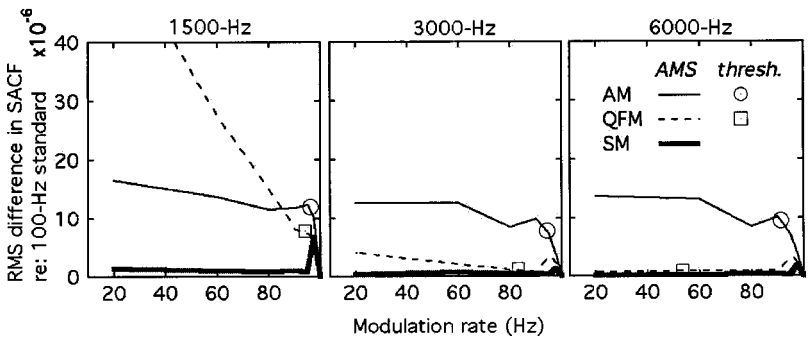


FIG. 5. The rms difference between SACF of the standard and each of a range of target stimuli is shown. The line type indicates results of the AM (solid), QFM (dotted), and shallow-AM (thick gray) modulation conditions. Mean psychophysical thresholds are indicated for the AM (open circle) and QFM (open square) conditions.

decreased sensitivity for components above 2400 Hz, a result that was interpreted in terms of reduced availability of phase locking information above this frequency. In another paradigm, Moore and Sek (1996) suggested that the ability to detect FM for low rates of FM is dominated by phase-locking information for carrier frequencies up to approximately 4000 Hz. This type of FM task has been used by others to make inferences about phase locking in cochlear-impaired ears (Lacher-Fougere and Demany, 1998). Finally, in a study of pitch perception for bandpass iterated rippled noise stimuli, Yost *et al.* (1998) interpreted their data as indicating that phase-locking information was utilized even for frequencies as high as 6–8 kHz. Although different studies have suggested different upper frequency cutoffs for the utility of synchrony information in humans, the data should not necessarily be seen as being contradictory. For example, it is likely that some tasks require more robust encoding of fine structure than others in order to maintain a high level of performance.

The present findings are consistent with an interpretation that fine structure can play a role in the pitch perception of unresolved components. This interpretation might appear to be in conflict with conclusions of the study by Kaernbach and Bering (2001) and Kaerenbach and Demany (1998). Kaerenbach and Demany (1998) investigated pitch perception for 6-kHz high-pass filtered pulse trains, and Kaernbach and Bering (2001) replicated that study using a lower filter cutoff of 2 kHz. These experiments found that pitch perception for unresolved components in pulse trains was related strongly to first-order temporal regularities (intervals between successive pulses), but not as robustly to higher-order regularities (intervals between nonsuccessive pulses). This result did not seem to depend upon the presence of energy below 4 kHz, where fine structure could plausibly have contributed to the pitch perception. Kaernbach and Bering (2001) concluded that their results supported ''the idea of an envelope analysis process that does not depend on the phase locking to the fine structure of the carrier.'' It is possible to reconcile the conclusions of Kaernbach and Bering (2001) with the present results by assuming that the effects of phase locking on the pitch for unresolved components are most likely to be revealed when envelope information is impoverished (which was the case in some of the conditions of the present study, but not in the previous studies using filtered pulse trains).

IV. SUMMARY AND CONCLUSIONS

In summary, the present study investigated sensitivity to reductions in the frequency of modulation (either AM or QFM) as a function of the modulation carrier frequency. In all cases, the spectral components comprising the stimuli were assumed to be unresolved. The results were consistent with an interpretation that the performance achieved for unresolved QFM components depended upon the spectral region occupied by those components. When the unresolved QFM components were located in a relatively low spectral region (1500 Hz), performance was usually relatively good and similar to performance for AM, presumably because fine structure phase-locking information provided useful cues for pitch. As the unresolved QFM components occupied increasingly higher spectral regions, performance declined, presumably because the fine structure phase-locking information became reduced in quality, and the observer was forced to rely upon relatively impoverished envelope cues. It is assumed that performance for AM was relatively more stable across carrier frequency because of the consistent availability of relatively strong envelope cues. Some of the individual differences in the data are probably related to individual variation in the ability to use envelope cues alone, particularly at high carrier frequencies. Overall, the data are consistent with an interpretation that pitch perception for unresolved components depends upon both fine structure and envelope cues, and that the relative importance of these cues depends upon the spectral region occupied by the stimuli.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD Grant No. 5 R01 DC00418. We thank Ray Meddis for helpful discussions regarding the AMS model. Neal Viemeister and two anonymous reviewers provided helpful comments on a previous version of this manuscript.

ANSI (**1996**). ANSI S3-1996, ''American National Standards Specification for Audiometers'' (American National Standards Institute, New York).

Cariani, P. A., and Delgutte, B. (**1996a**). ''Neural correlates of the pitch of complex tones. I. Pitch and pitch salience,'' J. Neurophysiol. **76**, 1698–1716.

Cariani, P. A., 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**, 1717–1734.

Goldstein, J. L. (**1967**). ''Auditory spectral filtering and monaural phase perception,'' J. Acoust. Soc. Am. **41**, 458–479.

Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.

Hall, J. W., and Soderquist, D. R. (**1975**). ''Encoding and pitch strength of complex tones,'' J. Acoust. Soc. Am. **58**, 1257–1261.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). ''Hearing a mistuned harmonic in an otherwise periodic complex tone,'' J. Acoust. Soc. Am. **88**, 1712–1724.

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.

Javel, E. (**1980**). ''Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of complex tones,'' J. Acoust. Soc. Am. **68**, 133–146.

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. X., and Yin, T. C. (**1992**). ''Responses to amplitude-modulated tones in the auditory nerve of the cat,'' J. Acoust. Soc. Am. **91**, 215–232.

Kaernbach, C., and Bering, C. (**2001**). ''Exploring the temporal mechanism involved in the pitch of unresolved harmonics,'' J. Acoust. Soc. Am. **110**, 1039–1048.

Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of auditory temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibres in the Cat's Auditory Nerve* (MIT Press, Cambridge, MA).

Lacher-Fougere, S., and Demany, L. (**1998**). ''Modulation detection by normal and hearing-impaired listeners,'' Audiology **37**, 109–121.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.

Licklider, J. C. R. (**1959**). ''Three auditory theories,'' in *Psychology: A Study of a Science*, edited by S. Koch (McGraw-Hill, New York).
Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3118.
Meddis, R., and Hewitt, M. J. (**1991**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Micheyl, C., Moore, B. C., and Carlyon, R. P. (**1998**). ''The role of excitation-pattern cues and temporal cues in the frequency and modulation-rate discrimination of amplitude-modulated tones,'' J. Acoust. Soc. Am. **104**, 1039–1050.
Moore, B. C., 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.
Moore, B. C. J. (**1980**). ''Neural interspike intervals and pitch,'' Audiology **19**, 363–365.
Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.
Nuttall, A. L., and Dolan, D. F. (**1996**). ''Steady-state sinusoidal velocity responses of the basilar membrane in guinea pig,'' J. Acoust. Soc. Am. **99**, 1556–1565.
Palmer, A. R. (**1982**). ''Encoding of rapid amplitude fluctuations by cochlear-nerve fibers in the guinea-pig,'' Arch. Oto-Rhino-Laryngol. **236**, 197–202.
Ritsma, R. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.
Ritsma, R. J. (**1962**). ''Existence region of the tonal residue. I.,'' J. Acoust. Soc. Am. **34**, 1224–1229.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). ''Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey,'' J. Neurophysiol. **30**, 769–793.
Schouten, J. F. (**1970**). ''The residue revisited,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Lieden, The Netherlands).
Seebeck, A. (**1841**). ''Beobachtungen über einige Bedingungen der Entstehung von Tönen,'' Ann. Phys. (Leipzig) **53**, 417–436.
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.
Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). ''A revised model of the inner-hair cell and auditory-nerve complex,'' J. Acoust. Soc. Am. **111**, 2178–2188.
Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.
van den Brink, G. (**1970**). ''Two experiments on pitch perception: Diplacusis of harmonic AM signals and pitch of inharmonic AM signals,'' J. Acoust. Soc. Am. **48**, 1355–1365.
Wightman, F. L. (**1973**). ''Pitch and stimulus fine structure,'' J. Acoust. Soc. Am. **54**, 397–406.
Yost, W. A., Patterson, R., and Sheft, S. (**1998**). ''The role of the envelope in processing iterated rippled noise,'' J. Acoust. Soc. Am. **104**, 2349–2361.

A mechanical model of vocal-fold collision with high spatial and temporal resolution[a)]

Heather E. Gunter[b)]
Division of Engineering and Applied Sciences, Harvard University, Pierce Hall, 29 Oxford Street, Cambridge, Massachusetts 02138

(Received 18 April 2002; accepted for publication 1 November 2002)

The tissue mechanics governing vocal-fold closure and collision during phonation are modeled in order to evaluate the role of elastic forces in glottal closure and in the development of stresses that may be a risk factor for pathology development. The model is a nonlinear dynamic contact problem that incorporates a three-dimensional, linear elastic, finite-element representation of a single vocal fold, a rigid midline surface, and quasistatic air pressure boundary conditions. Qualitative behavior of the model agrees with observations of glottal closure during normal voice production. The predicted relationship between subglottal pressure and peak collision force agrees with published experimental measurements. Accurate predictions of tissue dynamics during collision suggest that elastic forces play an important role during glottal closure and are an important determinant of aerodynamic variables that are associated with voice quality. Model predictions of contact force between the vocal folds are directly proportional to compressive stress ($r^2=0.79$), vertical shear stress ($r^2=0.69$), and Von Mises stress ($r^2=0.83$) in the tissue. These results guide the interpretation of experimental measurements by relating them to a quantity that is important in tissue damage. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1534100]

PACS numbers: 43.70.Aj, 43.70.Jt, 43.70.Bk [AL]

I. INTRODUCTION

Vocal-fold biomechanics play an important role in voice production. In particular, the solid mechanics underlying collision between the vocal folds are interesting due to their association with aerodynamic variables that are linked to voice quality changes (Holmberg *et al.*, 1988), and their proposed role in increasing mechanical stress levels that may cause tissue damage (Titze, 1994). A detailed theoretical model of vocal-fold tissue mechanics during collision that connects tissue-damaging variables with clinical/experimental variables and vice versa is an excellent tool with which to clarify the impact stress hypothesis. It also has the potential to predict the effect of surgical manipulation of vocal-fold structure on voice quality.

The impact stress hypothesis of vocal nodule development assumes that forces applied to the medial vocal fold during collision correlate with mechanical stresses in superficial tissue. Situations that increase collision force are thought to increase the risk of tissue damage and pathology development by increasing mechanical stresses levels (Titze, 1994). The stresses that are likely to be associated with collision and cause tissue damage include compressive stress perpendicular to the plane of contact that may cause cellular rupture and shear stress parallel to the plane of contact that may cause separation of tissue elements. Damage may also be caused by alterations in cellular function in response to the stress environment. The environment can be quantified by Von Mises stress, a scalar stress quantity that is used in engineering to represent composite stress in a material (Chandrupatla and Belegundu, 1997). The pathologies observed are 1–2 mm in size and involve the thickness of the lamina propria (Dikkers, 1994). Therefore, it is likely that mechanical stress changes occur on a similar or smaller scale. Jiang and Titze (1994) and Hess *et al.* (1998) cite the impact stress hypothesis as their motivation for identifying situations that increase intrafold collision forces and implying corresponding increases in mechanical stresses. The relationship between measured total collision forces and distributed mechanical stresses remains to be defined.

A theoretical model of vocal-fold collision that captures the dynamics of vocal-fold closure has the potential to be applied to prediction of the surgical outcomes. This would address the subset of vocal-fold surgeries where vocal-fold structure is manipulated in order to improve voice quality (Zeitels, 1998). Surgical manipulation of individual layers of vocal-fold tissue alters tissue vibration, which affects the glottal aerodynamic waveform and the acoustic voice spectrum. Voice quality is linked to changes in the aerodynamic variables maximum flow declination rate and minimum flow (Holmberg *et al.*, 1988). Vocal-fold closure and collision are the tissue correlates of these variables.

The idea to represent vocal-fold tissue mechanics mathematically is not a new one. Lumped mass models of phonation with as few as two (Ishizaka and Flanagan, 1972) or three (Story and Titze, 1995), and as many as 16 (Titze, 1973, 1974) masses predict aerodynamic and acoustic waveforms. Collision is represented by a spring that becomes active as a mass crosses the glottal midline. Story and Titze predict intraglottal collision pressures that agree qualitatively with Jiang and Titze's (1994) experimental measurements. However, the theoretical peak pressure prediction is approxi-

[a)]Portions of this work were presented in "Analysis of Factors Affecting Vocal Fold Impact Stress Using a Mechanical Model," 142nd meeting of the Acoustical Society of America, Ft. Lauderdale, FL, December 2001.
[b)]Electronic mail: gunter@fas.harvard.edu

mately fivefold less than the comparable experimental measurement (Jiang and Titze, 1994). The spatial resolution of the lumped mass models is insufficient to reflect the scale of pathology and surgery, and their empirically assigned spring, damper, and mass values have few direct implications for vocal-fold tissue physiology. Analytical continuum mechanics approaches (Titze and Strong, 1975; Berry and Titze, 1996) permit infinite spatial resolution but require simplified geometries (i.e., rectangular prisms) that make representation of pathological and surgical alterations difficult.

Alipour *et al.*'s (1996, 2000) and Jiang *et al.*'s (1998) models of phonation use finite-element techniques that overcome the barriers of low spatial resolution and restricted geometries. These techniques involve spatially and temporally discretizing a continuum mechanics problem into solid elements and time increments followed by numerical solution. Alipour *et al.*'s (2000) self-oscillating model requires increased spatial resolution, calculation of mechanical stress distributions, and extension to true three-dimensionality in order to represent vocal-fold pathologies and examine mechanical stress distributions. Jiang *et al.*'s (1998) finite-element model of vocal-fold tissue has sufficient spatial resolution to represent pathology, but limits analysis to resonant and forced vibration. It requires an explicit representation of collision forces between the vocal folds in order to be used in investigations of contact related pathology etiologies and voice quality.

A finite-element model of vocal-fold collision that is not self-oscillating is presented below. It has spatial resolution that is capable of representing the submillimeter scale of vocal-fold injury and repair, temporal resolution that captures the submillisecond time scale of vocal-fold collision, and an explicit representation of vocal-fold collision. Model predictions of collision force are validated against experimental data and the model potential is illustrated in an examination of the relationships between collision force and mechanical stress levels.

II. DEVELOPMENT

A. Assumptions

The model of vocal-fold collision presented below has three unique features. The spatial resolution of 250 μm is finer than that incorporated in other vocal-fold models. A single isotropic, linear elastic material characterizes the entire vocal-fold structure. Glottal closure during phonation is represented using a nonoscillating model consisting of appropriate initial conditions and a high-fidelity model of vocal-fold solid mechanics. Justifications for these features are outlined below.

The necessary spatial resolution is dictated by the size of geometric variations in the model, the desired resolution of model outputs, and desired computational speed. The model is intended for use in studying the effects of vocal-fold pathologies and their repair on tissue movement and for investigating the distribution of mechanical stresses in vocal-fold tissue as an indication of pathology development risk. Benign vocal-fold pathologies such as nodules and polyps have dimensions of 1–2 mm on the vocal-fold surface (Dikkers, 1994). Spanning the pathology with a minimum of four elements requires element dimensions of 250 μm and allows for some sculpting of the geometry. Surgical repair of the vocal fold can involve manipulation of only the superficial lamina propria (Zeitels, 1998), which, at approximately 500 μm thick (Hirano *et al.*, 1983), can be represented by two elements. Mechanical stresses that are important in tissue injury are on the same scale as or smaller scale than the pathologies that they are proposed to cause. Therefore, a resolution that is sufficient to represent pathologies provides appropriate information on stress distributions. Further resolution would increase geometric fidelity and output detail, but would also increase computational load significantly.

A simple material definition is used in the model. The assumption of linear elasticity is consistent with Min *et al.*'s (1995) observation of linear stress–strain behavior of human vocal ligaments for strains of less than 15%. Strains in the model do not exceed this 15% upper bound. The primary mode of deformation of the vocal fold during vocal-fold closure is compressive in the transverse (i.e., medial–lateral and inferior–superior) directions. Due to the lack of direct data on the elastic properties of vocal-fold tissue when undergoing this kind of deformation, isotropy is assumed and elastic properties based on longitudinal (i.e., anterior–posterior) tensile deformation are used for guidance. The range of elastic moduli derived by Min *et al.* (1995) using human vocal ligament samples (21.2–42.2 kPa) spans the moduli derived by Alipour-Haghighi and Titze (1991) for unstimulated canine thyroarytenoid muscle (20.7 kPa) and canine vocal-fold cover (41.1 kPa) and are the same order of magnitude as moduli derived by Kakita *et al.* (1981) using canine vocal-fold muscle and cover. The assumption of homogeneous properties is within the spread of the data in the literature.

Some studies suggest that air pressure in the glottis is low during glottal closure. The myoelastic-aerodynamic theory of voice production cites three reasons for glottal closure: elastic recoil of the deformed tissue, decreased pressure on the glottal walls due to the high velocity of the air stream through the opening, and decreased subglottal pressure (Jiang *et al.*, 2000). Experimental measurements performed on canine hemilarynges illustrate that, in an almost fully adducted larynx, air pressure in the glottis remains below 10% of subglottal pressure during glottal closure, and does not rise until full closure has occurred (Alipour and Scherer, 2000). Measurements performed on excised human hemilarynges also demonstrate reduced glottal air pressures during closure (Alipour *et al.*, 2001). It is consistent with observations of minimum air pressures and maximum vocal-fold deformation when the glottis is fully open to postulate that elastic forces dominate during glottal closure, and to focus on these forces in a model of glottal closure.

B. Implementation

The model geometry shown in Fig. 1 represents the membranous portion of a fully adducted single vocal fold. The model is defined in Cartesian coordinates. The origin is middle of the posterior glottis, inferior to the bulge of the vocal fold. The y axis defines the midline of the glottis, x dimensions indicate lateral distance from the glottal midline,

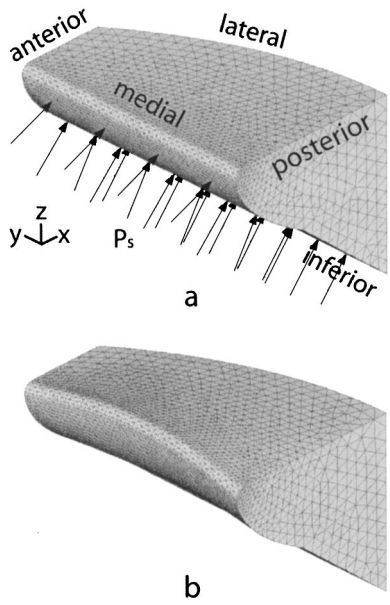


FIG. 1. Calculation of initial conditions for glottal closure using a finite-element model of the vocal fold. (a) problem formation: the three-dimensional solid represents the undeformed vocal fold. Triangular divisions indicate individual three-dimensional finite elements. Anterior, lateral, and posterior boundaries are held fixed. The arrows indicate the direction and location of application of forces representing subglottal pressure (P_s). (b) Problem solution: the three-dimensional solid represents the deformed vocal fold. This output is dependent on subglottal pressure and is the initial condition for glottal closure (see Fig. 2).

and z translations represent movement in the vertical direction. The vocal-fold dimensions are 1.4 cm long (y direction), 0.5 cm wide anteriorly, 1.0 cm wide posteriorly (x direction), and 1.0 cm high posteriorly (z direction). The anterior, posterior, and lateral surfaces are fixed to represent their attachment to the laryngeal cartilages. This structure is based on Titze and Talkin (1979) and uses their nominal parameters [i.e., shaping factor (s) equal to 0.05 radians, glottal angle (w) equal to 0.0 radians, and inferior surface angle (q) equal to 0.7 radians]. The Titze and Talkin geometry is modified by applying a fillet with a radius of 0.05 cm to the superior medial curve in order to create a smooth contour in the coronal plane. The glottal width (g) between the inferior aspect of the fillet and the superior aspect of the inferior surface is derived from an expression given by Titze and Talkin

$$g(0.05 \leqslant h \leqslant 0.5) = 0.45 - 1.9(h - 0.5) + 2(h - 0.5)^2,$$

where h is the vertical distance below the superior surface. The minimum glottal width of 0.0 cm that occurs at the inferior end of the fillet (h equal to 0.05 cm) indicates that the undeformed vocal fold is tangent to the glottal midline. A similar geometry has been used by Alipour *et al.* (2000) with good results.

A finite-element support software package (FEMAP 8.0; EDS, Inc., Plano, TX) is used to divide the geometry into 14 242 three-dimensional tetrahedral elements, which provides an elemental resolution of 250 μm on the medial surface. The model has 23 637 nodes, which reflect the ten nodes necessary to define each element and provide 70 911 degrees of freedom. Repetition of analyses using a model with 2.5-fold fewer elements affects collision force predictions by a maximum of 4%, which indicates that the mesh is fine enough to achieve convergence of results.

A common material defines all elements in the model. The material properties are nondirectional and linear, with a Young's modulus (E) of 36.1 kPa and a Poisson's ratio (ν) of 0.3. These parameters are based on experimental measurements on human vocal ligament samples by Min *et al.* (1995) and are valid for strains of less than 15%. None of the simulations presented below produces strains greater than this magnitude. The first mode of vibration is calculated using a linear perturbation eigenvalue analysis of the vocal-fold geometry in the absence of the pressure load and contact condition (ABAQUS STANDARD 5.8.1; HKS Inc.; Pawtucket, RI) to provide an evaluation of the geometric and material representations.

Interaction between the vocal fold and a rigid surface in the middle of the glottis (i.e., a yz plane at $x=0$, shown as thick black lines in Fig. 2) represents the interaction between the modeled fold and the opposing vocal fold. This surface and the nodes on the medial surface of the vocal fold form a contact pair: As a medial surface node becomes coincident with the rigid surface, sufficient force is applied to prevent the node from passing through the surface. Contact forces are in the x direction since the interaction is frictionless. Total contact force is the sum of the forces acting on all medial surface nodes. Contact area is a function of the number of nodes that are in contact with the surface. When the vocal fold it in its neutral position the contact area is 7 mm^2.

Deformation of the vocal-fold model due to subglottal pressure (P_s) defines the initial condition for vocal-fold closure. A distributed load applied perpendicularly to the element faces that form the inferior and medial surfaces as shown in Fig. 1(a) represents subglottal pressure. Equilibrium tissue deformation [Figs. 1(b), 2(a)] and stress distributions [Fig. 2(a)], calculated for each subglottal pressure ($P_s = 0.4$, 0.6, 0.8, 1.0, 1.5, and 2.0 kPa) (ABAQUS STANDARD 5.8-1), define the maximally open glottis. Geometric nonlinearities are accounted for during this calculation.

Vocal-fold closure is the progression from maximally open [Fig. 2(a)] to fully closed [Fig. 2(c)] glottis due to elastic forces within the tissue. The nonlinear, transient, dynamic solution for a given initial condition is obtained by incrementing forward through time using implicit integration algorithms (ABAQUS STANDARD 5.8-1). Time increments that result in solution convergence, calculated based on the half-step residual and changes in the contact state, range from 10–100 μs before collision occurs and 10–100 ps during collision. Typically between 50 and 100 increments are necessary to predict 3 ms of tissue movement and capture complete vocal-fold closure. Outputs of these calculations include movement of each node, mechanical stresses in each element, and interactions between nodes and the rigid midline surface.

The goals with this model are to validate predictions of closure kinematics and collision surface forces against published experimental measurements, and to examine the relationship between collision surface forces and mechanical stress levels in the tissue during collision. Model predictions that reflect these goals are discussed below.

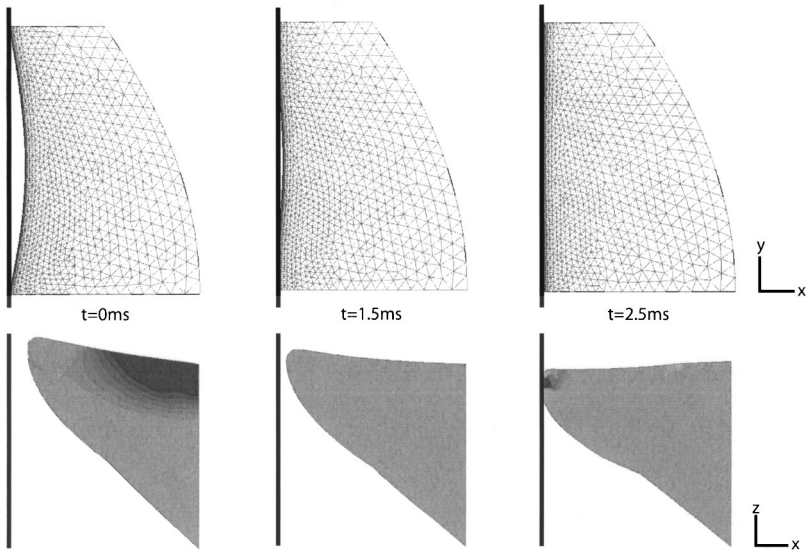


FIG. 2. Finite-element model of the membranous portion of a vocal fold during glottal closure. Each column provides two views of the model at one time increment. The vertical lines represent the plane of contact with the opposing vocal fold. Images in the top row are superior views, similar to those obtained clinically using a laryngoscope, that illustrate the changing glottal area. The bottom row of images are coronal sections through the midmembranous region of the vocal fold. The degree of dark shading in the bottom row of images represents the magnitude of compressive stress in the tissue in the direction perpendicular to the plane of contact. The images in the left-hand column represent the initial condition. The deformation is due to the application of a subglottal pressure of 2 kPa. There are compressive stresses in the superior lateral region with a maximum magnitude of 3.5 kPa. This provides the elastic energy that causes the fold to move. The middle and right-hand columns show frames in the solution. The images in the middle column represent the deformation of the vocal fold 1.5 ms after the start of glottal closure. There are no compressive stresses greater than 3 kPa. The images in the right-hand column represent the deformation of the vocal fold 2.5 ms after the start of glottal closure. There is a compressive stress concentration with a magnitude of 10 kPa at the site of contact with the opposing fold.

III. VALIDATION

The natural frequency is 127 Hz, which is consistent with experimentally measured resonant frequencies (Kaneko *et al.*, 1987) and theoretically predicted resonant frequencies for adult male voices (Alipour *et al.*, 2000). This agreement corroborates the geometric and material definitions.

Glottal closure dynamics are an important output of the model, and model behavior agrees with observations of vocal-fold vibration. The qualitative movement of the fold shown in Fig. 2 includes movement towards the midline and collision with the opposing fold. Closure begins at the anterior and posterior ends and progresses towards the midmembranous region (Fig. 2, top row), which is consistent with *in vivo* stroboscopic observations of fully adducted phonation. The initial midmembranous glottal displacement (i.e., the maximum lateral displacement of the medial vocal-fold edge) is an approximately linear function of subglottal pressure and ranges between 0.15 and 0.8 mm as illustrated in Fig. 3. These predictions are consistent with previous visual observations of normal vocal-fold vibration (Zemlin, 1998). The inferior portion of the vocal-fold edge is the first point of collision with the opposing fold. As collision proceeds, contact includes the superior vocal-fold edge (Fig. 2, bottom row). This progression is similar to phase differences that are observed clinically. Contact areas, which are calculated based on the number of elements touching the midline surface in the step 2 solutions, are also functions of P_s and time, as shown in Fig. 4. The contact areas range from 0 to 13 mm^2, which lies within the area of vocal-fold contact imaged by Jiang and Titze (1994).

Collision dynamics during closure are another important output of the model, and model behavior agrees with published experimental results. As shown in Fig. 5, the total collision force increases with time and reaches either a peak or a plateau depending on the initial P_s. The rise time of collision in a 3-mm midmembranous region for $P_s = 1.5$ and 2.0 kPa is approximately 0.5 ms, which is equal to the experimentally measured rise time for impact with a 3-mm diameter central sensor as measured in canine hemilarynges by Jiang and Titze (1994). Due to the collision force plateau, there are no clear force peaks for P_s less than 1.5 kPa. Collision force peaks are defined either as the force magnitude at

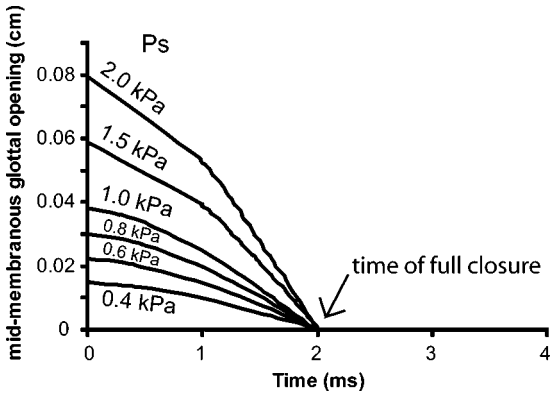


FIG. 3. Effect of subglottal pressure on the medial movement of the vocal fold—Each line traces the solution for the x position of a midmembranous point on the medial edge of the vocal-fold edge during glottal closure in a single model. The models differ by the subglottal pressure magnitude used to calculate the initial closure condition (i.e., $t = 0$). The arrow points to the time when the midmembranous glottal opening and glottal area are zero, which is defined as the time of full closure. This value is not dependent on subglottal pressure.

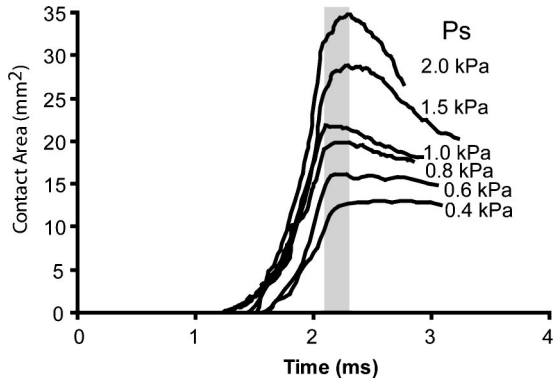


FIG. 4. Effect of subglottal pressure on the area of contact between the vocal fold and the contact surface—Each line traces the solution for contact area, which is proportional to the number of nodes that are at $x=0$ (i.e., touching the contact plane) during glottal closure in a single model. The models differ by the subglottal pressure used to calculate the initial condition of glottal closure (i.e., $t=0$). The shaded area indicates the temporal range of contact area maxima, which varies among models.

the time of full closure, which is independent of P_s as shown in Fig. 3, or as the force magnitude at the time of maximum contact area, which is dependent on P_s as shown in Fig. 4. These peaks are used to derive linear relationships between subglottal pressure and peak collision force. Table I compares the slopes and regression coefficients for linear best-fit lines derived from model predictions with data derived from *in vitro* experiments on canine hemilarynges by Jiang and Titze (1994). Both definitions of predicted peak collision force fall within the span of the experimental results.

IV. MECHANICAL STRESS DURING COLLISION

To determine whether local collision force predicts local mechanical stress, a regression analysis between contact force predictions and mechanical stress predictions is performed. For each time point during vocal-fold collision and at each 1-mm interval along the vocal-fold edge, contact force and medial superficial stresses are extracted from model results. Compressive stress perpendicular to the plane of contact, shear stress parallel to the plane of contact in the

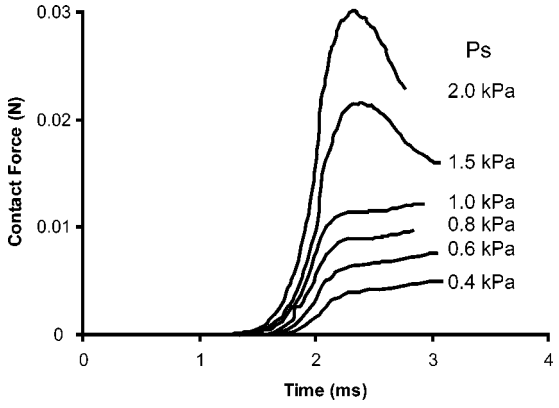


FIG. 5. Effect of subglottal pressure on the total contact force between the vocal fold and the contact surface. Each line traces the solution for contact force, which is the sum of point loads applied by the contact plane to nodes on the medial surface, during glottal closure in a single model. The models differ by the subglottal pressure magnitude used to calculate the initial condition of glottal closure (i.e., $t=0$).

TABLE I. Comparison between linear regression analysis of peak contact force and subglottal pressure for theoretical predictions using a finite-element model and Jiang and Titze's (1994) experimental measurements on canine hemilarynges.

Contact force definition	Linear slope (mN/kPa)	Regression coefficient
Model prediction—peak contact area	16.39	0.977
Model prediction—full closure	10.48	0.930
Experimental measurement (Jiang and Titze, 1994)	Range=7.70–26.60 Mean=12.46	

longitudinal direction, shear stress parallel to the plane of contact in the vertical direction, and Von Mises stress are examined. Figure 6 shows the relationship between contact force and stresses for a subglottal pressure of 1 kPa. At this subglottal pressure there is a general trend between increases in local contact force and increases in compressive stress ($r^2=0.79$), vertical shear stress ($r^2=0.69$), and Von Mises stress ($r^2=0.83$). There is no trend between local contact force and longitudinal shear stress ($r^2=0.00$). Local variations from the general trends can be appreciated by examining the data points for each 1-mm segment. The variations are most pronounced for longitudinal shear stress, where there are local inverse trends between contact force and stress despite the lack of a general trend.

V. DISCUSSION

This model examines the time and spatial course of vocal-fold closure, but not separation, during phonation and makes two contributions. The first contribution is accurate prediction of closure kinematics and collision surface forces, despite the absence of an aerodynamic representation and an assumption of homogeneous material properties. This provides insight as to the importance of elastic forces in glottal closure and supports future application to the study of the effects of vocal-fold tissue elasticity and structure on the dynamics of glottal closure, corresponding aerodynamic variables, and, by extension, voice quality.

The second contribution is an illustration of the potential of this model of glottal closure to provide a window to mechanical conditions in the vocal-fold interior that may increase injury risk. Regressions between surface contact force and relative predictions of mechanical stress in the tissue indicate that the implications of experimental impact force measurements are position dependent and identify compressive stress perpendicular to the contact plane, shear stress parallel to the contact plane in the longitudinal direction, and Von Mises stress as candidate mechanical stresses that may cause tissue damage as a result of high collision forces.

The ability of this model to predict closure kinematics and collision surface forces despite exclusion of aerodynamic forces from the model provides insight into the mechanics of voice production. Agreements between model predictions and published data suggest that elastic forces within the tissue dominate the mechanics of vocal-fold closure and collision, and are therefore a major determinant of aerodynamic variables that are associated with closure, such as minimum flow and maximum flow declination rate (MFDR).

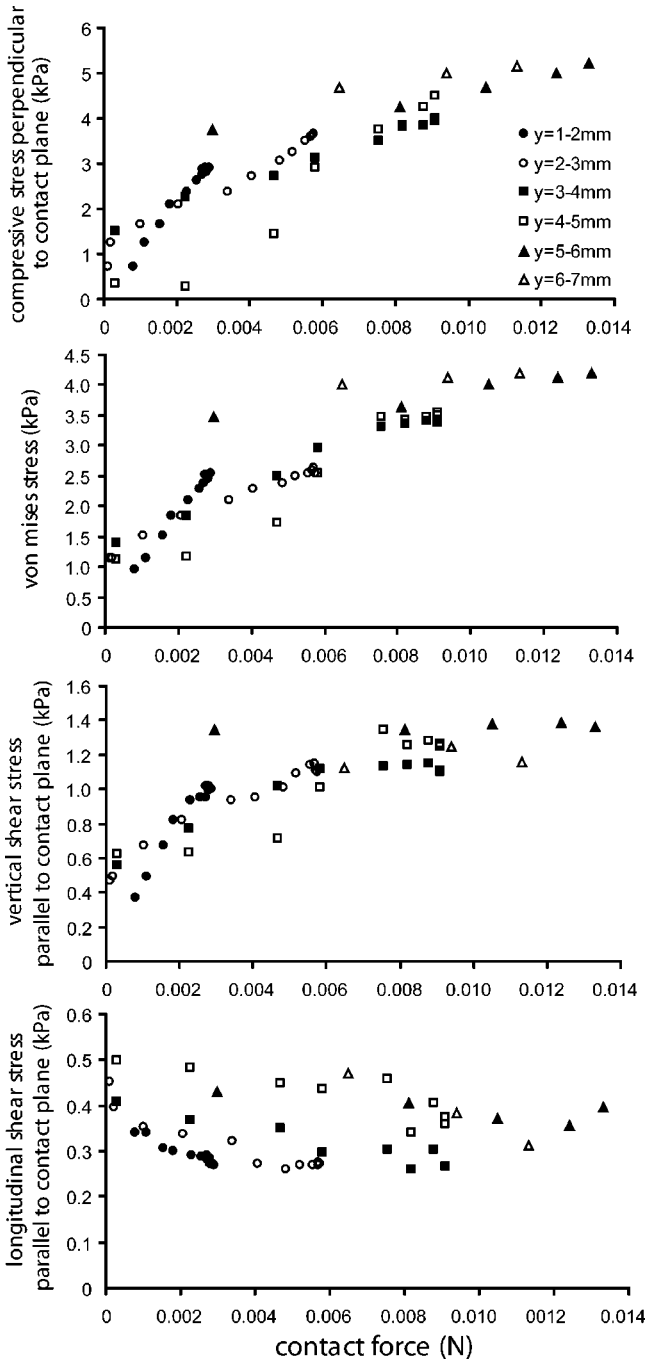


FIG. 6. Relationship between local contact force and mechanical stress on the medial vocal-fold surface during collision. The x value of each data point represents the integrated contact force over a 1-mm vertical slice of the medial surface. The y value of each data point represents the maximum mechanical stress in the surface tissue in the same 1-mm slice. Data are symmetric about the midmembranous location ($y=7$ mm). For simplicity, only data for the anterior half of the fold are shown. Correlation coefficients are given in the text.

The dependence of voice quality on MFDR and minimum flow allows extension of this hypothesis to propose that solid mechanics of vocal-fold tissue are a dominant factor in voice quality. *In vivo* associations between structural changes that affect vocal-fold closure, such as vocal nodules, increased minimum flow, decreased MFDR, and altered voice quality perceptions (Kuo, 1998) support this hypothesis.

The suggestion that tissue mechanics dominate vocal-fold closure kinematics, important aerodynamic variables, and voice quality reinforces the proposed use of a vocal-fold collision model to predict surgical outcomes. Comparison of the closure dynamics between models of vocal folds with organic pathologies and models of the same folds following surgical repair will allow prediction of the effects of surgery on voice quality. The prognostic value of this application may be further refined by incorporation of a layered geometrical structure and more complex material property definitions; these will facilitate the creation of higher fidelity models of pathology and surgical changes.

The lack of a defined contact force peak in models with low subglottal pressures and the lack of a reopening phase in all models does not agree with clinical observations of glottal opening, but was not an intended focus of this model. These results indicate that bouncing of the vocal folds following collision is not sufficient to cause glottal opening and reinforces the important role played by air pressure in vocal-fold separation. Comparison of the current results with those from a new model with a sophisticated aerodynamic representation will test the hypothesis that tissue elasticity forces dominate during glottal closure and that aerodynamic forces dominate during vocal-fold separation.

Agreement between model glottal closure kinematics and collision surface forces and experimental measurements does not necessarily imply that model predictions of mechanical stress are accurate. It is the case, in any complex model with multiple input variables, that there may not be one unique combination that produces the desired output. In the case of this model, prediction of appropriate contact surface forces and closure kinematics only implies that one appropriate set of input parameters was identified. A different combination of input parameters may have provided the same surface contact force predictions, but different absolute mechanical stress predictions. However, it is likely that the relative nature of mechanical stress predictions would remain the same. Therefore, it is reasonable to compare them against each other.

The relationships between contact forces on the vocal-fold surface and superficial tissue mechanical stress in the medial edge of the vocal fold, derived using a model of vocal-fold collision, guide the interpretation of experimental contact force measurements. Local collision force measurements are indicative of local compressive stress, vertical shear stress, and Von Mises stress. This observation suggests that pathologies associated with maneuvers that have increased local impact forces (e.g., phonation with high subglottal pressures) may be due to a compressive mode of tissue failure, a shear mode of tissue failure, or alterations in cellular behavior. Unfortunately, without knowledge of what stress magnitudes cause tissue failure, it is impossible to comment on which stress levels are most damaging to the tissue. Correlations between stress predictions and epidemiology of lesions that reflect tissue injury, such as vocal nodules, will test the impact stress hypothesis of pathology development.

Additional experimental measurements will improve definition and validation of the vocal-fold collision model and other models of vocal-fold tissue mechanics. Three-

dimensional material property data based on human vocal-fold tissue samples undergoing compression would improve the fidelity of model inputs. *In vivo* human results of surface collision forces would be a preferred source against which to validate model predictions because agreement will enhance the applicability of the model to human voice. Other experimental measurements that will be useful in validation include electroglottographic measurements of contact area and photoglottographic measurements of glottal area.

ACKNOWLEDGMENTS

Thank you to Ken Stevens, Robert Hillman, and Robert Howe for their scientific guidance and assistance with manuscript preparation, to Jaime Lee for her assistance with data analysis, to the J. Acoust. Soc. Am. reviewers and editors for their valuable feedback, and to the Whitaker Foundation for financing this work through a Biomedical Engineering Graduate Fellowship.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). ''A finite-element model of vocal-fold vibration,'' J. Acoust. Soc. Am. **108**, 3003–3012.

Alipour, F., Montequin, D., and Tayama, N. (**2001**). ''Aerodynamic profiles of a hemilarynx with a vocal tract,'' Ann. Otol. Rhinol. Laryngol. **110**, 550–555.

Alipour, F., and Scherer, R. C. (**2000**). ''Dynamic glottal pressures in an excised hemilarynx model,'' J. Voice **14**, 443–54.

Alipour, F., and Titze, I. R. (**1996**). ''Combined simulation of two-dimensional airflow and vocal fold vibration,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 17–30.

Alipour-Haghighi, F., and Titze, I. R. (**1991**). ''Elastic models of vocal fold tissues,'' J. Acoust. Soc. Am. **90**, 1326–1331.

Berry, D. A., and Titze, I. R. (**1996**). ''Normal modes in a continuum model of vocal fold tissues,'' J. Acoust. Soc. Am. **100**, 3345–3354.

Chandrupatla, T. R., and Belegundu, A. D. (**1997**). *Introduction to Finite Elements in Engineering*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ), p. 17.

Dikkers, F. G. (**1994**). *Benign Lesions of the Vocal Folds: Clinical and Histopathological Aspects* (Drukkerij Van Denderen B. V., Groningen, The Netherlands).

Hess, M. M., Verdolini, K., Bierhals, W., Mansmann, U., and Gross, M. (**1998**). ''Endolaryngeal contact pressures,'' J. Voice **12**, 50–67.

Hirano, M., Kurita, S., and Nakashima, T. (**1983**). ''Growth, development and aging of human vocal folds,'' in *Vocal Fold Physiology; Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College Hill, San Diego), pp. 23–43.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). ''Glottal air-flow and transglottal air-pressure measurements for male and female speakers in soft, normal, and loud voice,'' J. Acoust. Soc. Am. **84**, 511–529.

Ishizaka, K., and Flanagan, J. L. (**1972**). ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' Bell Syst. Tech. J. **51**, 1233–1268.

Jiang, J. J., Diaz, C. E., and Hanson, D. G. (**1998**). ''Finite element modeling of vocal fold vibration in normal phonation and hyperfunctional dysphonia: implications for the pathogenesis of vocal nodules,'' Ann. Otol. Rhinol. Laryngol. **107**, 603–610.

Jiang, J., Lin, E., and Hanson, D. G. (**2000**). ''Voice disorders and phonosurgery. I. Vocal fold physiology,'' Otolaryngol. Clin. North Am. **33**, 699–718.

Jiang, J. J., and Titze, I. R. (**1994**). ''Measurement of vocal fold intraglottal pressure and impact stress,'' J. Voice **8**, 132–144.

Kakita, Y., Hirano, M., and Ohmaru, K. (**1981**). ''Physical properties of vocal tissue: measurements on excised larynges,'' in *Vocal Fold Physiology*, edited by M. Hirano and K. Stevens (University of Tokyo Press, Tokyo), pp. 377–398.

Kaneko, T., Masuda, T., Akiki, S., Suzuki, H., Hayasaki, K., and Komatsu, K. (**1987**). ''Resonance characteristics of the human vocal fold *in vivo* and *in vitro* by an impulse excitation,'' in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaski, and K. S. Harris (Little, Brown, Boston), pp. 349–365.

Kuo, H.-K. J. (**1998**). ''Voice Source Modeling and Analysis of Speakers with Vocal-Fold Nodules,'' Ph.D. dissertation (Massachusetts Institute of Technology, Cambridge, MA), pp. 41–119.

Min, Y. B., Titze, I. R., and Alipour-Haghighi, F. (**1995**). ''Stress–strain response of the human vocal ligament,'' Ann. Otol. Rhinol. Laryngol. **104**, 563–569.

Story, B. H., and Titze, I. R. (**1995**). ''Voice simulation with a body-cover model of the vocal folds,'' J. Acoust. Soc. Am. **97**, 1249–1260.

Titze, I. R. (**1973**). ''The human vocal cords: A mathematical model. I,'' Phonetica **28**, 129–170.

Titze, I. R. (**1974**). ''The human vocal cords: A mathematical model. II,'' Phonetica **29**, 1–21.

Titze, I. R., and Strong, W. J. (**1975**). ''Normal modes in vocal cord tissues,'' J. Acoust. Soc. Am. **57**, 736–749.

Titze, I. R. (**1994**). ''Mechanical stress in phonation,'' J. Voice **8**, 99–105.

Titze, I. R., and Talkin, D. T. (**1979**). ''A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation,'' J. Acoust. Soc. Am. **66**, 60–74.

Zeitels, S. M. (**1998**). ''Phonosurgery—past, present, and future,'' Operative Tech. Otolaryngol. Head Neck Surg. **9**, 179.

Zemlin, W. R. (**1998**). *Speech and Hearing Science: Anatomy and Physiology*, 4th ed. (Allyn and Bacon, Boston), pp. 137–152.

Effects of disfluencies, predictability, and utterance position on word form variation in English conversation

Alan Bell and Daniel Jurafsky
Department of Linguistics, University of Colorado, Boulder, Colorado 80309-0925

Eric Fosler-Lussier
Bell Laboratories, Lucent Technologies

Cynthia Girand
Department of Linguistics, University of Colorado, Boulder, Colorado 80309-0925

Michelle Gregory
Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island

Daniel Gildea
University of Pennsylvania, Philadelphia, Pennsylvania

(Received 21 May 2001; revised 18 October 2002; accepted 5 November 2002)

Function words, especially frequently occurring ones such as (*the, that, and,* and *of*), vary widely in pronunciation. Understanding this variation is essential both for cognitive modeling of lexical production and for computer speech recognition and synthesis. This study investigates which factors affect the forms of function words, especially whether they have a fuller pronunciation (e.g., ðiˌ ðæt, ænd, ʌv) or a more reduced or lenited pronunciation (e.g., ðəˌ ðɨt, n, ə). It is based on over 8000 occurrences of the ten most frequent English function words in a 4-h sample from conversations from the Switchboard corpus. Ordinary linear and logistic regression models were used to examine variation in the length of the words, in the form of their vowel (basic, full, or reduced), and whether final obstruents were present or not. For all these measures, after controlling for segmental context, rate of speech, and other important factors, there are strong independent effects that made high-frequency monosyllabic function words more likely to be longer or have a fuller form (1) when neighboring disfluencies (such as filled pauses *uh* and *um*) indicate that the speaker was encountering problems in planning the utterance; (2) when the word is unexpected, i.e., less predictable in context; (3) when the word is either utterance initial or utterance final. Looking at the phenomenon in a different way, frequent function words are more likely to be shorter and to have less-full forms in fluent speech, in predictable positions or multiword collocations, and utterance internally. Also considered are other factors such as sex (women are more likely to use fuller forms, even after controlling for rate of speech, for example), and some of the differences among the ten function words in their response to the factors. [DOI: 10.1121/1.1534836]

PACS numbers: 43.70.Bk, 43.70.Fq [AL]

I. INTRODUCTION

The modern availability of large online labeled corpora of conversational speech is a boon to the researcher studying phonological production. An obvious benefit of online conversational data is their ecological validity. But a less obvious benefit is the opportunity it affords for greatly expanding the range of situational and contextual effects that can be studied. Previous studies on read speech or reiterant speech, for example, have been able to study in detail the effect of phonetic variables such as segmental context on phonological variation. A number of variables, however, have received much less attention in earlier studies. In particular, the role of larger contexts such as prosodic context, lexical context, and the environment of the production task, is much less clear, particularly in natural conversational settings, and particularly for disfluent speech. It is essential to understand the role of these contextual factors in order to inform models of speech production.

This study investigates how the forms of English words in natural conversation are systematically affected by three such contextual variables: the presence or absence of neighboring disfluencies, the predictability of the word from the neighboring lexical context, and the position of the word in utterances. More specifically, it is hypothesized that words have stronger, less lenited forms in the presence of disfluencies, when they are less predictable, and when they occur at the beginning or end of utterances.

The first of these factors concerns a ubiquitous aspect of the production process itself, namely the disfluencies that arise when a hitch occurs in the flow from concept to speech. Previous studies have suggested that the surface form of words seem to be different when the speaker is experiencing lexical production planning problems. For example, Fox Tree and Clark (1997) showed that the word *the* was more likely to be pronounced with a full vowel rather than a schwa in disfluent contexts (when followed by a pause, filled pause, or

repetition). Our goal is to extend the Fox Tree and Clark (1997) study by examining whether such planning problems affect words other than *the*. We also study the nature of the form variation itself.

The second factor in our study is contextual predictability. Frequency and predictability have played a fundamental role in models of human language processing for well over a hundred years (Schuchardt, 1885; Jespersen, 1922; Zipf, 1929). But while modern models of human language comprehension often assume that probabilistic information plays a role in the access and disambiguation of linguistic structures (Jurafsky, 1996; MacDonald, 1993; McRae *et al.*, 1998; Trueswell and Tanenhaus, 1994), the role of probability in production is much less well understood. It is known that frequent words are shorter and more often reduced or lenited (Zipf, 1929; Fidelholz, 1975; Rhodes, 1992, 1996), that a second mention of a word is shorter than the first mention (Fowler and Housum, 1987), and that words which are more contextually predictable are produced in a less intelligible manner (Lieberman, 1963). In earlier work, (Jurfsky *et al.*, 2001; Gregory *et al.*, 1999) we proposed the *Probabilistic Reduction Hypothesis* to link these phenomena: word forms are reduced when they have a higher probability. The probability of a word is conditioned on many aspects of its context, including neighboring words, syntactic and lexical structure, semantic expectations, and discourse factors. In this paper we examine the role of local lexical probability: the probability of a word given the neighboring word or words. Our goal is to understand how this kind of local probabilistic context affects surface phonological and phonetic form, and how it relates to other kinds of context. We also ask whether the influence of a word's predictability is limited to the selection of alternate wordforms during lexical access, or whether predictability also influences surface phonetic form directly.

The third contextual factor we investigate is prosodic structure. The location of a word in larger prosodic domains such as utterances, turns, intonational phrases, and phonological phrases plays an important role in reduction. Studies of language change and of pronunciation variation have long accepted three main effects—final lengthening (Klatt, 1975; Ladd and Campbell, 1991; Crystal and House, 1990, *inter alia*), initial strengthening (i.e., more extreme articulation) (Fougeron and Keating, 1997; Byrd *et al.*, 2000, *inter alia*), and final weakening (i.e., less extreme articulation). During the last several decades more and more quantitative studies have helped make our understanding of these general effects more precise; see Fougeron and Keating (1997) for a review. Many of these results, however, derive from laboratory paradigms like reiterant speech, and have not been tested on natural speech production. Furthermore, it has been difficult to tease apart prepausal lengthening from lengthening at the edge of prosodic domains. We attempt to address these questions in the domain of natural conversational speech production.

How shall we investigate the effect of these factors? Natural speech corpora offer a number of potential dependent variables to use to study variation in phonological production. Previous research suggests that lenition and reduction, or alternatively, lengthening and strengthening, are associated with context. We therefore focused on this dimension of variation, selecting three dependent factors: duration of the entire word, categories of vowel quality, and presence or absence of coda obstruents. Longer pronunciations, with citation vowels or full vowels, are more frequent in explicit (e.g., formal, lento) styles; shorter pronunciations, with reduced or elided vowels and/or elided consonants, are more frequent in elliptical (e.g., casual, allegro) styles. These three variables thus reflect a scale of lenition, weakening, or reduction. For convenience we will use the term ''reduced'' throughout this paper to refer to the more elliptical forms. Other aspects of reduction, such as elision of initial consonants or consonant weakening, were not considered.

We investigate this reduction or lenition not in every word, but only in ten of the most frequent English words, namely the function words *I*, *and*, *the*, *that*, *a*, *you*, *to*, *of*, *it*, and *in*. Why is the study limited to just these words? Briefly, there were three main reasons. A study covering all words was judged too ambitious and too complex for an initial application of the multidimensional analysis methods to be used, and one must start somewhere. The high frequencies of occurrence of these words, their especially great form variation, and their common monosyllabic form offered important advantages to the analysis. The fact they are also function words, that is strongly associated with syntactic and semantic/pragmatic structures, was not a primary consideration. Finally, and crucially, the fact that such words are not usually accented allowed us to avoid problems of controlling for the interaction of segmental form and presence of accent. If the contextual effects on reduction that we postulate exist, there should be strong evidence for them in the most frequent words; the possibly more difficult task of verifying that the effects also hold throughout the lexicon can be left for further research.

Our data is drawn from the Switchboard corpus of telephone conversations between strangers collected in the early 1990's (Godfrey *et al.*, 1992). We chose the Switchboard corpus for our research because various portions of it have been phonetically transcribed, coded for part of speech, syntactically parsed, and segmented into utterance-like units.

The next section of the paper, Sec. II, summarizes our methodology for extracting and coding forms, and analyzing form variation. Section III then describes details of the various control variables; rate of speech, phonetic context, pitch accent, etc., and summarizes their effects. Section IV focuses on our first contextual variable, the presence of disfluencies, which we take to be largely associated with problems in planning speech. Section V focuses on the second contextual variable, word predictability from neighboring words. Section VI deals with the last contextual variable, the position of a word in prosodic domains. Section VII concludes with a discussion of the results and their implications.

II. METHODOLOGY

A. The corpus

As described above, our observations of the ten function words *I*, *and*, *the*, *that*, *a*, *you*, *to*, *of*, *it*, and *in* were drawn

from the phonetically transcribed portion of the Switchboard corpus collected in the early 1990's (Godfrey *et al.*, 1992). The corpus contains 2430 conversations averaging 6 min each, totaling 240 h of speech and approximately 3 million words. The corpus was collected at Texas Instruments, mostly by soliciting paid volunteers who were connected to other volunteers via a robot telephone operator, and was then transcribed by court reporters into a word-by-word text.

Approximately 4 h of this speech was phonetically hand-transcribed at ICSI (the International Computer Science Institute) by linguistics students at UC Berkeley (Greenberg *et al.*, 1996; Greenberg, 1997) as follows. The speech files were automatically segmented into ''fragments'' at turn boundaries or at silences of 500 ms or more. The transcribers were given these strings, the word transcription, and a rough automatic phonetic transcription which was automatically aligned to the wavefile at syllable boundaries. They then corrected this rough phonetic transcription, using an augmented version of the ARPAbet. The transcribers also corrected the syllable boundary marks and the silence onsets and offsets. In general, transcribers were instructed to pay careful attention to both the waveform and spectral displays of the signal in making their decisions. In cases where no specific event could be found to mark a syllable boundary, guesses were made using tables of the duration distributions of particular segments. These boundary marks were then used to automatically compute syllable durations. Similarly, pause durations were computed for portions of the signal not attributed to a syllable. The hand-labeled and hand-segmented syllables were then automatically aligned against the word transcription, resulting in a duration for each word. Since the current study only considers monosyllabic words, in many cases these durations correspond exactly to the hand-labeled syllable boundaries. In some cases where resyllabification occurred, the automatic alignment did slightly shift the boundaries. The entire corpus contains roughly 38 000 transcribed word tokens.

Approximately two-thirds of this phonetically transcribed corpus (henceforth the ICSI corpus) was also part of the utterance-segmented portion of the Treebank III release of the Switchboard corpus (Marcus *et al.*, 1999). In this release, 1155 of the 2430 conversations were segmented by the Linguistic Data Consortium (LDC) into approximately the 205 000 utterance-like units described in Sec. VI (Meteer *et al.*, 1995).

Our database thus combines information from three sources: the original lexically transcribed Switchboard corpus, the Treebank III utterance segmentation, and the ICSI phonetically transcribed corpus. All three of these corpora, together with documentation describing them, are available from the Linguistic Data Consortium at `http://www.ldc.upenn.edu/`

From the phonetically transcribed data, we extracted 9926 occurrences of the ten function words. We immediately eliminated 801 occurrences whose surface form clearly indicated an alignment error or a transcription error, such as the word *you* pronounced [rju], or the word *you* pronounced [ði]. This left 9125 tokens of the ten function words. Of these, 404 were alternate forms such as *an*, *I'd*, *I'm*, *I'll*, and *you'd*, *you're*, etc., which because of their small numbers and incomparable forms, were excluded from our analysis. We also excluded 361 items which were coded as ''nulls,'' i.e., as having no segmental realization except possibly as a featural modification of an adjoining word. (The discussion below on coding of vowel quality comments further on the null items.) This left 8362 items as input to our analyses. The actual sample sizes of most analyses are smaller than this, because not all variables apply to all the data or could not be defined for all the data; see the discussions below.

TABLE I. Most frequent pronunciations of the ten words, grouped into basic, full, and reduced-vowel pronunciations. For each word the three most common tokens of each type of pronunciation are listed in order of frequency.

	Basic	Other full	Reduced
a	[eɪ]	[ʌ],[ɪ]	[ə],[i]
the	[ði],[i],[di]	[ðʌ],[ðɪ],[ʌ]	[ðə],[ði],[ə]
in	[ɪn],[ɪ],[ɪr],	[ɛn],[ʌn],[æn]	[ɨn],[ṇ],[ən]
of	[ʌv],[ʌ],[ʌvv]	[ɪ],[i],[ɑ]	[ə],[əv],[əf]
to	[tu],[tʉ],[ru]	[tʊ],[tɪ],[tʌ]	[tə],[ti],[ə]
and	[æn],[ænd],[ær]	[ɛn],[ɪn],[ʌn]	[ɨn],[ṇ],[ən]
that	[ðæ],[ðæt],[æ]	[ðɛ],[ðɛt],[ðɛr]	[ðɨt],[ði],[ðir]
I	[aɪ]	[ɑ],[ʌ],[æ]	[ə]
it	[ɪ],[ɪt],[ɪr]	[ʊt],[ʊ],[ʌ]	[i],[ə],[ət]
you	[yu],[u],[yʉ]	[yɪ],[ɪ],[i]	[yi],[y],[i]

B. How forms were coded

The three dependent factors of duration, vowel quality, and coda presence were coded in the following ways.

(1) *Vowel quality*: We coded each vowel as **basic**, **other full**, or **reduced**. The basic vowel is the citation or clarification pronunciation, e.g., [ði] for *the*.[1] The reduced vowels are [ə] (arpabet [ax]), [ɨ] (arpabet [ix]), [ɚ] (arpabet [axr]), and [ɵ] (midcentral reduced vowel with more [o]-like or [u]-like coloring than [ə], not in the arpabet). Any other vowel is a full vowel. This three-way distinction is split into two binary contrast variables: full/reduced (basic and other full vowel versus reduced vowel) and basic/full. See Table I for the most frequent tokens of the words in each of the vowel quality categories.

(2) *Coda obstruent*: For words which have coda obstruents (*it, that, and, of*), we coded whether the consonant is present or not. The sonorant nasal codas of *in* and *and* were not considered.

(3) *Length*: We coded the duration of the word in milliseconds.

In general we relied on the ICSI transcriptions for our coding, using software to automatically assign a category to a transcribed word. Thus, for example, if the ICSI transcription of a word was [əv], our software automatically categorized the observation as *reduced vowel* and *coda present*. We judged the interlabeler agreements of the ICSI transcribers, reported between 72.4% and 76.9%, to be quite acceptable for this task. We did, however, check the data several ways, deleting or modifying some items. As mentioned above, we first examined every pronunciation of every word, and elimi-

nated 801 incorrect pronunciations that were due to alignment errors in our automatic word-segmentation program. We then listened to the utterances in five classes of tokens that seemed likely to affect our analysis: possible misalignments in our processing, a sample of tokens transcribed as having no segment, all tokens of arpabet [ux], all tokens of arpabet [er], and a random sample of 100 of the function words. Some items from these five classes were recoded, mainly [ux] as either a nonreduced high front round vowel [ʉ], as prescribed, or reduced [ʊ]; and [er] as either full [ɝ] or reduced [ɚ]. Some items were removed, mainly those transcribed as having no segment (''nulls''), since from our sample we judged that many were equally segmental as other transcriptions. Most of the incorrect coding of these words as having ''no segments'' was due to a mismatch between incorrect word transcriptions and the phonetic transcription for the utterances. In these cases, the phonetic labelers transcribed the utterance correctly but did not correct the original word-level transcription. The mismatch between these two produced a number of alignment errors which we eliminated.

Our judgments of the tokens in the random sample in general agreed with the original transcribers. Notably, however, we judged five of the 57 full vowels in the sample to be reduced, whereas we agreed with the coding of all the 43 reduced vowels. This suggests that there may be a bias toward full vowels in the transcription.

Neither we nor the original Switchboard Transcription Project at ICSI computed interlabeler agreement statistics for syllable duration labeling. We did, however, check some segmental durations, and while in many cases we might have slightly moved segment boundaries, we found no reason to believe there were any gross systematic errors in duration labeling.

The coding for each of the three major independent variables (planning problems, predictability, and utterance position) is described in the later sections pertaining to each variable.

C. Controlling for possible confounds: Regression analysis

While the use of natural conversational corpora provides the benefits of situational validity and allows us a larger contextual window, it also presents a problem. Natural speech has myriad confounding factors that affect form variation such as phonetic factors, rate of speech, pitch accent, and sociological factors like age and sex. These factors are typically correlated. We use multiple regression, both linear and logistic, to examine the individual contributions of a variable in this situation.

A regression analysis is a statistical model that predicts a *response variable* (in this case, the word duration, or the frequency of vowel reduction) based on contributions from a number of other *explanatory factors* (Agresti, 1996). Thus, when we report that an effect was significant, it is meant to be understood that it is a significant parameter in a model that also includes the other significant variables. In other words, after accounting for the effects of the other explanatory variables, adding the explanatory variable in question produced a significantly better account of the variation in the response variable.

For duration, which is a continuous variable, we use ordinary linear regression. For vowel quality, and coda presence, which are categorical variables, we use logistic regression. Logistic regression models the effect of explanatory variables on a categorical variable in terms of the *odds* of the category, which is the ratio $[P(\text{category})]/[1\text{-}P(\text{category})]$. For a binary category like full versus reduced vowel, we estimate the odds by the ratio of the percentages of the two values: the article *a* occurs with a full vowel 17 percent of the time, and with a reduced vowel, 83 percent; the odds of a full vowel are $17/83 = 0.20$ (to 1).

It is important to understand that the goal of the regression analyses is not to create a model that will predict the forms of function words. It is primarily used as a tool to evaluate the significance and magnitude of selected factors in the presence of other correlated factors, possibly also significant.

Of course, establishing that a factor can contribute additional improvement to a model is one of the basic facts needed to construct production models. Much more, such as details of dependencies among factors and magnitudes of effects at high and low values of factors, is also needed. Some selected questions of this sort that appear to be particularly important are explored in the sections below. For example, we generally report important interactions, notably the greater effect of predictability from a preceding word for more frequent word combinations (Secs. V A 1, V A 2). Hypotheses about certain factor dependencies are tested with specific regression models, e.g., relations between disfluencies and utterance-initial position, Sec. IV B 1; and relations between word duration and vowel reduction (Secs. III A, IV A, V A 3). A few comparisons between alternative models are tested, e.g., the comparison between a two-factor model distinguishing preceding and following disfluencies and a single-factor model which does not, Sec. IV A.

The size of a factor's effect is of considerable importance, since a factor can be a significant addition, but have a relatively small effect. The level of significance of an effect is often associated with its magnitude—an effect significant at $p = 0.0001$ is likely to be greater than one that is significant at $p = 0.01$. This is not a generally appropriate measure of effect magnitude, however, so two other measures are commonly used. One is based on the estimated weight of the factor in the regression equation; the other is based on the proportion of the total variation that the factor accounts for. The weight-based measure, which is the more direct of the two, is reported for the main results. It is a ratio derived from two parameters—the estimated weight and the range of the factor. In the simplest case for a categorical factor like presence of a disfluency, the range is 1, so that the effect magnitude is simply proportional to the regression weight. In Sec. IV A, the effect of a disfluency on vowel reduction is reported as 1.68, meaning that all other factors being equal, in a disfluent context, the estimated odds that the word contains a full vowel are 1.68 times the odds of a full vowel in a fluent context. This value is calculated by taking the regression coefficient of the disfluency factor as a power of 10,

since the regression equation is based on log odds. For continuous factors, a range representing the middle 90 percent of the data is used, from the 5th to the 95th percentiles. Thus, in Table X, the magnitude of the effect on duration of the conditional probability given the previous word, 0.80, means that the estimated duration of the most predictable words (at the 95th percentile) are 0.80 times shorter than the least predictable words (at the 5th percentile).

One of the assumptions of regression analyses is that the items in the data are independent. This assumption is surely violated to some extent by our data, since many of the same items are uttered by the same speaker, or in the same conversation. A more serious violation occurs when two words are adjacent. Just how to best deal with this inherent weakness of corpus studies is not clear. Sampling one item from each conversation, or part of a conversation, was judged too costly. It would drastically reduce the power of the analyses and their generality. One reason for examining the ten most frequent function words was the expectation that in most instances such words would be separated, and occur in separate phrases. Although this is usually the case, about 20 percent of the items do occur adjacently in combinations such as *of the* and *that I*, which is not very surprising just given their high frequency of occurrence. The consequence of the nonindependence of such items is that the significance values are inflated to some extent. It is thus recommended that the reader not take the reported levels literally, but as an informed indicator of the relative significance of an effect. Where the significances are very great, this is of little concern, but becomes more of one for more marginal ones. While we have reported some effects at levels up to the conventional 0.05 level, it seems prudent to regard any result above $p = 0.01$ as marginal.

The results are of course subject to the usual limitations of such analyses, most notably that they apply strictly only to the present database and to the particular operational coding used. In many ways, the database can be considered generally representative of American English conversation. But some of the specific characteristics of the data, for example, the particular way that fragments of conversations were selected for the ICSI database, require simplifications in variable definitions and sample selections that inevitably introduce some degree of bias. Examination of many such cases has not yielded any reason to think that the distortions are large enough to invalidate the main results. Nevertheless, it is perhaps well to regard the quantitative measures of the results as pertaining to this database, and to take the results more qualitatively as a basis, together with further research, for constructing production models.

III. CONTROL FACTORS

The reduction variables are each influenced by multiple factors that must be controlled to assess the contribution of the explanatory variables—presence of disfluencies, predictability, and position in turn. While it is of course not possible to control for every factor which influences reduction, we consider here the ones that are, from prior research, most likely to play a large role.

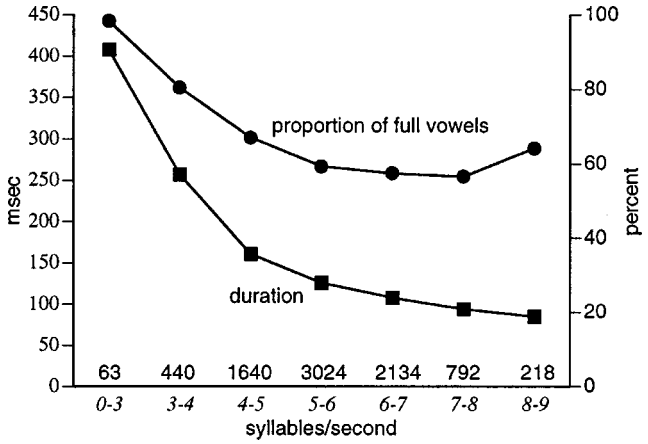


FIG. 1. Function word durations and proportions of full vowels by rate of speech. The scale for duration is on the left axis, the scale for full vowels is on the right. The number of observations for each rate category appears at the bottom of the graph.

(i) rate of speech of the speaker in syllables/second;
(ii) segmental context;
(iii) prosodic factors;
(iv) age and sex of speaker and hearer; and
(v) individual characteristics of the ten function words.

The focus of the paper leads us to regard these as control factors rather than object of study in their own right. The role of rate, phonetic context, and prosody in reduction is of course well-established. A detailed study of speaker and hearer effects in conversational speech is beyond the scope of this paper, in spite of its considerable interest. This section, therefore, reports primarily the details of the variables we selected to control these factors. Selected results about how these variables affect reduction are also presented.

A. Control factors: Rate of speech

Speech researchers have long noted the association between faster speech, informal styles, and more reduced forms. [For a recent quantitative account of rate effects in Switchboard, see Fosler-Lussier and Morgan (1999)]. We measured rate of speech at a given function word by taking the number of syllables per second in the speech fragment immediately surrounding the word, up to the nearest pause or turn boundary on each side. Fifty-one words with extremely slow or extremely fast rates were excluded from regression analyses. Unsurprisingly, rate of speech affected all measures of reduction. Words were more reduced when they were spoken more quickly. Comparing the difference between a relatively fast rate of 7.5 syllables per second and a slow rate of 2.5 syllables per second, a range which covers about 90 percent of the tokens, the estimated increase in the odds of full to reduced vowels is 2.2. That is, the odds of a full vowel at the slow rate is 2.2 times the odds at the faster rate. Figure 1 compares observed proportions (or averages, for length) with predicted values for five categories of rate along the range from 2.5 to 7.5 syllables per second. (The increased proportion of full vowels at the highest rate category is presumably not systematic.)

For all measures, there seems to be a limit effect for

faster rates; this is accounted for in the regression model by using log (rate) as the main explanatory variable as well as a (highly significant) quadratic $\log^2$ (rate) term. The overall effect of rate is weaker for coda deletion than for the other measures of reduction; in addition, the effect on deletion is largely confined to the slower rates.

Are the shortening effects for function words solely a consequence of a greater proportion of reduced (and shorter) vowels at faster rates? If they were, the apparently gradient effect of rate of speech on durational shortening might represent not a gradient effect of rate, but a categorical effect, stemming from more frequent selection of reduced vowel forms at faster speech rates. It turns out that there is a substantial additional shortening effect of rate even after accounting for vowel reduction and coda deletion. Overall, with no other variables involved, rate accounts for 17.9 percent of the variation in duration of the function words. With no other variables involved, vowel reduction and coda deletion account for 18.4 percent of the variation. After controlling for vowel reduction and coda deletion, rate still accounts for an additional 13.9 percent of the variation. A final characteristic of rate is that it did not affect all the words equally. The most strongly affected words were *a*, *the*, *to*, *and*, and *I*. Notably, regressions for *that*, *it*, and *in* did not show rate effects for any of the three vowel or coda reduction measures.

B. Control factors: Segmental context

A general fact about weakening processes is that the form of a word is influenced by the segmental context—in particular, more reduced forms tend to occur before a consonant than before a vowel (Rhodes, 1996, *inter alia*). This may result in an allophonic effect such as the widely studied loss of final /t/ and /d/ (Neu, 1980, and references therein). Alternatively, it may be an allomorphic one, as in the case of *the* with [ði] before vowels alternating with [ðə] before consonants (Keating *et al.*, 1994). The preceding segmental context is presumed to have much less influence.

Thus, for each of the function word tokens, we recorded whether the following word began with a consonant or a vowel. To account for an interaction between this following segment and the final obstruent consonant of the function word itself, we distinguished four separate contexts: V#V, V#CV, VC#V, and VC#CV. The nasals of *and* and *in* were treated as if they belonged to the nucleus, both because they can be expected to behave differently from the obstruents, and also because the interplay between vowel nasalization and nasal consonant shortening is not captured by the ICSI phonetic transcription.

In addition, the metrical strength of the following word or words can also be expected to influence reduction. Here, we attempted to capture some portion of this influence by coding each function word with a variable distinguishing whether the vowel of the following syllable is full or reduced. In general, since reduced vowels cannot be stressed or bear intonational accents, this variable may be regarded as mainly differentiating cases where the next potential prosodically strong syllable either follows directly or else one or more syllables later. A more direct effect is predicted by Bolinger's (1986) lengthening rule, which states that a full vowel is lengthened if the next vowel is also full.

TABLE II. Observed average durations and reduced vowel percentages of closed-syllable (VC) and open-syllable (V) function words before words beginning with consonants and with vowels.

	Word	Next word	Duration (ms)	Percentage of reduced vowels
consonant	VC	CV	132	33.7
follows	V	CV	102	45.6
vowel	VC	V	158	29.7
follows	V	V	128	33.0

Observed average duration and the percentage of reduced vowels are shown in Table II for the four contexts. Before a consonant in the next word, words are shorter and are more likely to be reduced. These differences were assessed by regressions after controlling for rate effects. For both vowel reduction and shortening, the onset of the following word has a very strong effect. Overall, the odds of vowel reduction are 1.63 times greater before consonants than before vowels, and item durations are 0.79 times shorter before consonants than before vowels. The consonant–vowel effect on vowel reduction is stronger for open-syllable words than for closed-syllable ones; the effect for closed syllables is still highly significant ($p = 0.0005$).

The full-reduced status of the vowel in the next word affects open-syllable items, whose vowels are more likely to be reduced if the next word has a full vowel in its first syllable (whether it begins with a consonant or not). This is a moderately significant effect ($p = 0.007$, odds ratio of 1.43); there is no significant effect of the following vowel for closed-syllable items. Duration is also affected by the category of the vowel in the next word, but in a complex way. In the VC#CV and V#V contexts, there is little effect. Open-syllable items before consonants (the V#CV context) are shorter (by a factor of 0.82) if a full vowel follows, but closed-syllable items before vowels (the VC#V context) are shorter (by a factor of .84) if a reduced vowel follows.[2]

As with rate, shortening effects are still strong after controlling for vowel reduction. Overall, for example, the onset of the following word accounts for 4.1 percent of the variance; within reduced or full vowels, it still accounts for 3.6 percent of the variance. Individual analyses by item largely confirm the overall results for reduction and shortening. Only *you* for lengthening and *that* and *in* for reduction fail to show significant effects, which of course may be partially laid at the door of the smaller sample sizes.

C. Control factors: Intonational accent

One of the most important factors influencing an English word's pronunciation is whether it receives accent or not. Presence of accent is surely highly correlated with longer duration, lack of vowel reduction, and lack of elision, and likely has systematic associations with the presence of disfluencies, a word's predictability, and its position in the intonational phrase, the explanatory variables that are considered here. The most general way of accounting for its role in wordform variation is to regard it as one of the attributes of

a word's form together with its segmental attributes of duration, vowel reduction, etc., that is, as a response or observational variable. Desirable as this might be, it entails analytic complexities and model-theoretic assumptions that seemed premature at our present stage of knowledge. The alternative is to focus on the word's segmental form, and treat the prosodic status as an explanatory variable, part of the general context in which the word occurs, and one of the factors influencing the form of the word. Since intonational accent is not transcribed in the ICSI database, we could not examine its effects or control for them directly. One of the main reasons for studying high-frequency function words was that they are unlikely to be accented. It was our hope that the possible confound of accent with variables such as disfluency and predictability would be so infrequent that it would have little influence on their analysis. Fortunately, we have been able to verify this perhaps incautious hope, making use of two small accent-coded subcorpora from Switchboard. The first was a small portion of Switchboard that has been coded for accent under the direction of Shattuck-Hufnagel and Ostendorf, an alpha-release version of which they generously made available to us. The Shattuck-Hufnagel/Ostendorf corpus used a labeling scheme called POSH (Shattuck-Hufnagel and Ostendorf, 1999), a simplification of the TOBI prosodic labeling standard (Silverman *et al.*, 1992). In addition to the Shattuck-Hufnagel/Ostendorf corpus, we coded a very small subsample of Switchboard consisting of 120 words selected from the longest tokens of each function word; it was composed of 10 tokens of each function word, except for those which may be pronouns, *I* (20 tokens), *you* (15 tokens), and *that* (15 tokens).

The overlap between the Shattuck-Hufnagel/Ostendorf corpus and the most inclusive sample used in our analyses (8311 words) was 560 words. Of this set, 53, or 9.5 percent, were accented. (A larger proportion, 23 percent, were accented in our 120-word sample, presumably because of its heavy bias toward items most likely to be accented.) A majority of the accented words was either *that* (16) or *I* (15); with *and* (6), *you* (5), and *in* (4), they accounted for all but seven of the accented words. This concentration of accent on particular function words more or less agreed with our sample, in which only four functors had more than one accented token: *I*, 12 of 20; *you*, 7 of 15; *that*, 4 of 15; and *and*, 2 of 10. It appears that function words are indeed not likely to be accented, but some function words are much less likely than others to bear accent.

In order to determine whether the accent-coded data were representative of our entire database of phonetically transcribed words, we compared relative frequencies of the function words, rates of reduction, duration, rates of preceding and following disfluencies, and preceding and following conditional and joint probabilities, using chi-square or Fisher tests for the categorical variables and t-tests for the continuous ones. Since only one of the nine comparisons was even close to significant, the subset of accented-coded data appeared to represent the overall sample reasonably well. We also examined the association of accent with disfluencies and with predictability. This confirmed our expectation that accented words would be more likely to occur in disfluent contexts and that their conditional probabilities would on average be lower than unaccented words. Only for a previous disfluency, however, was the difference significant (one-tailed Fisher test, $p = 0.001$), perhaps because of the small sample.

The main question, of course, is whether the effects of the explanatory variables remain after controlling for pitch accent. We addressed this by examining only the 385 words without accent. (The accented words were too few to make including them in an analysis useful.) The details of the comparison of this analysis with the full analyses are presented in the following sections that treat the effects of disfluencies and of predictability on duration. Overall, as will be seen, the effects that are found for the unaccented word sample are similar to those for the overall sample uncontrolled for accent. These results are necessarily preliminary and incomplete. We did not examine whether accent might be masking the role of disfluencies and predictability on vowel reduction, basic versus nonbasic vowels, and coda deletion. There were not enough data to examine effects of position or effects for individual words. The clear results for duration, however, support our strategy of examining the factors affecting form variation in function words in the absence of controls for accent. Note also that the results for the individual words which virtually never receive accent are further support. Obviously, important questions about the role of accent remain, both for function words and content words.

D. Control factors: Age and sex of speaker and hearer

Studies of socially sensitive pronunciation variation such as the alternation of *-ing* and *-in* (Wald and Shopen, 1981) have shown that the status of speaker and hearer is often a factor in such variation. It is likely that such influences extend to our reduction variables, given that all our indices of variation are doubtless linked to the choice of elliptical versus explicit styles of speech, which is in turn sensitive to the speech situation. While an earlier study of the TIMIT corpus of read speech by Byrd (1994) did not find an effect of speaker sex on the duration of centralized vowels, she did find that men use certain more reduced forms such as taps and syllabic n more frequently than women. Previous research has also shown that rate and disfluencies are sensitive to the age and sex of speakers. Byrd (1994) found that men spoke TIMIT sentences on average 6.2 percent faster than women. Shriberg (1999), in her study of disfluencies in Switchboard, found that men had slightly more disfluencies per word than women. There is thus good reason *a priori* to control for speaker and hearer status. In this section we present a simple survey of the overall differences in reduction, rate, and disfluencies associated with the age and sex of speakers and hearers in our dataset.[3] Since this survey is meant only to provide a basis for the use of these factors as controls, we do not provide detailed analyses with individual assessments of significance. Some summaries of analyses for individual items are included in Secs. IV B, V B, and VI B. The more complex analysis needed to assess their effects on production is left for future study.

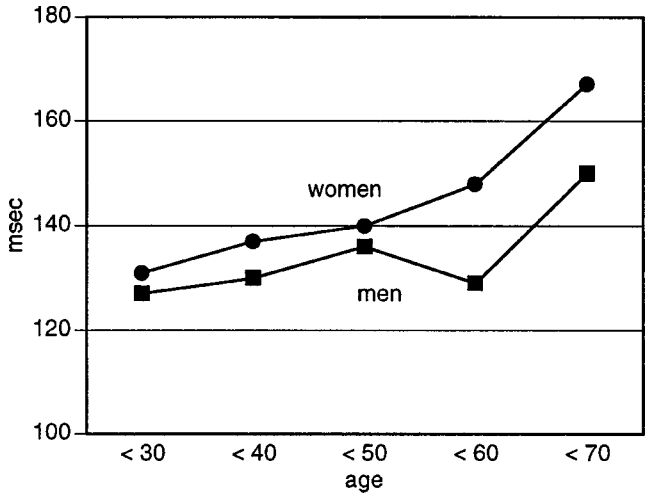


FIG. 2. Average word durations of function words of men and women speakers by age.

The ages of the 497 participants in our sample of Switchboard conversations ranged from 18 to 68; the mean age of the speakers was 37. There were 191 men speakers and 172 women, and 237 men listeners and 216 women. More items were spoken by men (58 percent) than by women (42 percent).

1. Effects of speaker and hearer on reduction variables

All the reduction measures are affected by the speaker's status.

(i) Duration: The average durations of function words are shown in Fig. 2 for men and women speakers by age category. Words spoken by women are longer (140 ms) than those spoken by men (131 ms), and words spoken by older speakers are longer (139 ms for speakers 40 and older versus 131 ms for those under 40), with the difference greater between older men and women.

(ii) Vowel reduction: The sex of speaker has the strongest effect on vowel reduction; there is little difference for older or younger speakers. Words spoken by men are reduced 41 percent of the time on average, but only 34 percent of the time for women.

(iii) Coda deletion: On the other hand, women speakers delete codas more frequently than men, 68 percent to 63 percent.

(iv) Basic vowel: Basic vowels are used more by older speakers than by younger ones; speakers under 40 use basic vowels 60 percent of the time, but this increases to 66 percent for speakers 40 and older (69 percent for speakers 60 and older).

These uncontrolled differences are significant at levels from $p<0.005$ to $p<0.0001$.

Differences associated with listener status are much smaller, and not significant, except perhaps for vowel reduction. Words are more often reduced when spoken to younger listeners under 40 than to older listeners ($p<0.05$). There were no dyad effects of speaker and listener age or of speaker and listener sex.

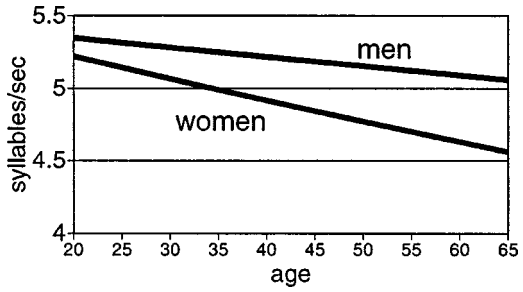


FIG. 3. Predicted average speech rates of men and women speakers by age.

Overall, reduction in function words is affected mainly by the age and sex of the speaker, and mainly in the directions that one would expect from the usual correlations of speaker status and levels of formality in speech: longer durations with less reduction of vowels and greater use of basic vowels by women and by older speakers.

2. Effect of age and sex on rate and on disfluencies

On average men spoke 6.4 percent faster than women. Men had an average rate of 5.4 syllables per second: women, an average rate of 5.0 syllables per second.[4] Younger speakers spoke more quickly, 5.5 syllables per second for speakers under 30, compared to 5.1 syllables per second for speakers 50 and older. Finally, there was an interaction of age and sex. While women on average spoke more slowly than men, older women spoke even more slowly than older men. These relationships are shown in Fig. 3, which presents the regressions of rate on age for men and women. There do not appear to be any differences in rate for different listener statuses.

The average rate of disfluency was 31.9 percent, where by disfluency we mean the presence of a disfluency either before or after a given function word. Interestingly, uncontrolled averages reveal little difference between men and women speakers of different ages, nor between men and women listeners of different ages. Since this was not consistent with Shriberg's (1999) results, we explored disfluency effects by controlling for rate and for the probability variables. The results agreed with Shriberg's finding that men have a higher rate of disfluency than women.

E. Control factors: Individual characteristics of the words

Different function words play different grammatical roles, have different distributions, have different kinds of meanings, and have different phonological forms. One should therefore expect some differences in how their reduction is affected by other factors. While it would be impractical and probably undesirable to control for item effects in analyses for overall effects of disfluencies, predictability, and utterance position, the idiosyncrasies of the ten function words unquestionably affect such results. It is thus important to compare their basic characteristics.

TABLE III. Frequencies of occurrence of the ten function words in the data.

I	and	the	that	to	you	a	of	it	in	Total
1381	1203	1123	786	769	758	745	583	562	452	8362

First, some are more frequent than others in our sample, reflecting their relative frequency in conversational speech. *I*, *and*, and *the* are the most frequent, and *of*, *it*, and *in* are the least, as can be seen in Table III.

The frequency range from most to least frequent is about 3 to 1, quite modest for lexical frequency in general, but to be expected since these are the ten most frequent words in the Switchboard corpus. One consequence of this distribution is that one cannot investigate the effect of lexical frequency on reduction with this database. This is partly because of the narrow range of frequencies, but more crucially because item frequency is confounded with other item idiosyncrasies in this small set, and there is no way to pull them apart. The other issue is the relative influence of the items on the overall results. Clearly the most frequent words will have more influence than the least frequent ones, and this needs to be kept in mind in the following discussions. It is also not advised to view this as an improper distortion of the results, since after all, the proportions of each word reflect their relative occurrence in conversational speech.

Next, the items differ considerably in their average durations and average rates of occurrence of basic, full, and reduced vowels and of coda deletion, resulting in different base levels for the overall effects on these variables.

Figure 4 shows the average durations of the ten function words. In this and following figures and tables, the words have been grouped by dominant function: articles *a, the*; prepositions/particles *in, of, to*; conjunctions *and, that*; and pronouns *I, it, you*. *And* and *that* are notably longer, in part because their vowel is intrinsically long and because they have a complex syllable structure. Similarly, the shortness of the article *a* probably reflects its single vowel.

Striking differences in the average rates of occurrence of full, unreduced vowels can be seen in Fig. 5. Six words, including *and* and *that*, have relatively high proportions of unreduced vowels, between about 65 percent and 95 percent; the others, including the article *a*, are much less likely to have an unreduced vowel (between about 25 percent and 35 percent). Thus another reason that *and* and *that* are longer is because they more often have full vowels. Likewise, the frequent reduced forms of *a* contribute to its relative shortness. Obstruent codas are present in about the same proportion for *that*, *it*, and *of* (44 percent for *that* and *it*, 54 percent for *of*); on the other hand, the very infrequent presence of the final stop of *and* (14 percent) suggests that the alternation between [n] and [nd] may stem in part from selection between distinct lexical forms of *and*.

Such item differences can affect our results in two main ways. First, the longest and shortest (or most/least reduced) items may contribute to floor or ceiling effects for some factors. As we mention below in Sec. IV B, a ceiling appears to be at least one factor responsible for *that*, *I*, and *it* not showing fewer reduced vowels in the context of following disfluencies. Their proportion of unreduced vowels is already so high it cannot become much higher. A second way is for one or more of the items with atypical forms or behaviors to be disproportionately represented over the range of a factor. One example of this is the very frequent occurrence of *and* in utterance-initial position, discussed below in Sec. VI B. Since *and* is long, it should exaggerate an initial lengthening effect; as we see later, if *and* is excluded from the analysis, the effect on duration in initial position is indeed reduced, although it remains significant.

There are of course additional differences in the behavior of the function words with respect to disfluencies, predictability variables, and utterance position, which largely reflect their functional differences. Some of the most salient such differences are discussed in the following sections. In lieu of controlling for item differences, we note below the consistency of effects over the function words, or lack of it. This provides a general indication of the robustness of the effects, and in some instances of markedly aberrant items, it suggests certain factors which may be responsible.

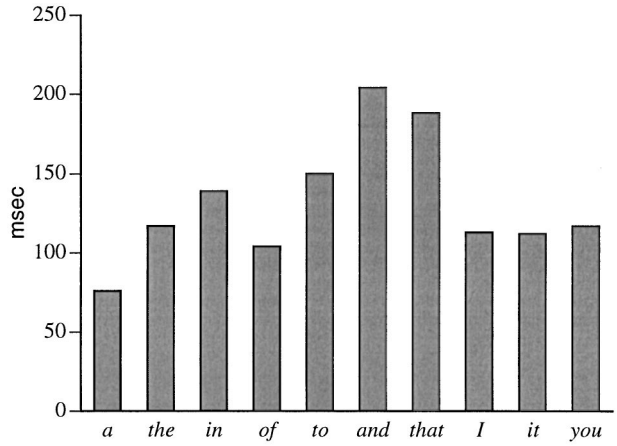


FIG. 4. Observed average durations of function words. The average duration of all words is 135 ms.

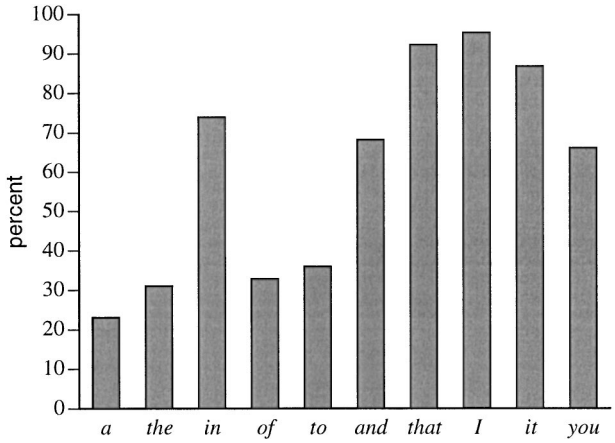


FIG. 5. Observed average frequency of occurrence of unreduced (full) vowels in function words. The average frequency for all words is 0.62.

TABLE IV. Observed durations, frequencies of basic and full vowels, and frequencies of coda presence for function words in fluent and disfluent contexts. The number of observations of the context categories appears in parentheses. Basic vowel frequencies are based on the 4886 words with full vowels. Obstruent coda presence frequencies are based on the 2947 words *and, it, of,* and *that*.

Context	Duration	Full vowel	Basic vowel	Coda presence
Fluent	109 ms (5480)	54% (5480)	64% (2936)	33% (1948)
Any disfluency	187 ms (2519)	77% (2519)	64% (1950)	39% (999)
Disfluency before	137 ms (1295)	73% (1295)	59% (940)	26% (366)
Disfluency after	222 ms (927)	80% (927)	66% (741)	42% (483)
Disfluency both	295 ms (297)	91% (297)	75% (269)	59% (150)

F. Summary of control factors

Most of the factors discussed above are controlled in our regression analyses by including appropriate variables in a base model. Our base regression models thus include the following variables.

(i) Log rate of speech and log squared of rate of speech;
(ii) Syllable type of target (open, closed);
(iii) Whether initial segment of next word begins with consonant or vowel;
(iv) Whether following vowel is reduced; and
(v) Age and sex of speaker, age of listener.

Included also were significant interaction variables, e.g., rate×speaker age. Some of these variables were dropped when they had negligible effect, e.g., listener age for duration analyses. The results presented below are based on analyses that controlled utterance position by excluding utterance-initial and utterance-final items, rather than with base models including utterance position variables, for reasons explained below. The effects of intonational accent, which could not be controlled, are assessed by comparing results from the accent-coded subsample described above with the results for the effects of disfluencies and of predictability; see note 8 in Sec. IV A and note 11 in Sec. V A 1. Similarly, rather than controlling for the differences among the function words, we summarize the results of analyses for each of the words individually in the following sections, and discuss the behaviors of selected words in more detail.

IV. PLANNING PROBLEMS AND DISFLUENCIES

The production of speech is accompanied by a variety of disfluencies, whose characteristics have been extensively documented (Shriberg, 1994, *inter alia*). In particular, it appears that certain disfluencies often have a prospective source, occurring as a reaction to speakers' trouble in formulating an upcoming idea, and expressing it with the proper syntax, words, prosody, and articulation. Fox Tree and Clark (1997) suggested that such planning problems are likely to cause neighboring words to have less reduced pronunciations. They found this to be true for *the*, and suggested that the pronunciation [ði] is used by the speaker as a signal of problems in production. Fox Tree and Clark suggested that this relationship might extend to other words. Other work has also pointed to form effects in disfluent contexts. O'Shaughnessy (1992), for example, argued that words lengthen before pauses, and Shriberg (1995) showed that forms of *I* and *the* were longer when they were repeated. It thus seems worthwhile to adopt the working hypothesis that longer and fuller forms are generally associated with planning problems, whether they function as signals of planning problems, or are part of production mechanisms to gain time to resolve planning problems, or some combination of the two.

In this section we extend such investigations to study the general relationship between disfluencies and pronunciation reduction in frequent words. Like Fox Tree and Clark (1997), we treat silent pauses, filled pauses *uh* and *um*, and repetitions as likely to be symptoms of planning problems. Each of the functors in our corpus is coded as belonging to a disfluent context if it is preceded or followed by one of these disfluencies.[5]

The following examples from our corpus illustrate the different disfluency contexts; numbers in parentheses are silence lengths in seconds.

Following disfluency	Sentence
Repetition	**I** I have strong objections to that.
Silence	...large numbers **of** (0.228) barefoot natives or something...
Filled pause (*uh*)	Somebody I talked to last week, they said they had **the** uh, they had problems doing some of the work

Preceding disfluency	Sentence
Repetition	I **I** have strong objections to that.
Silence	You know, the main things that I like about (0.214) **the** uh, job benefits...
Filled pause (*uh*)	it would encourage people, uh, **to** make more money

After eliminating uncodable items, there were 7999 function words coded for occurrence in preceding and following disfluent contexts. Of these, 2519, or 31 percent, occurred before or after a disfluency; 12 percent were followed by a disfluency, 16 percent were preceded by a disfluency, and four percent occurred between disfluencies.

A. Effects of disfluencies

Table IV compares durations, basic vowel frequency, reduced vowel frequency, and frequency of coda presence in fluent and disfluent contexts. Overall, longer and fuller forms are strongly associated with disfluencies, consistent with the hypothesis that they are symptoms of planning problems.

These observed differences, however, may not be a di-

TABLE V. Occurrence of disfluencies before and after function words. The percentages of following disfluencies are also shown.

		Disfluency before		
		Yes	No	Total
Disfluency after	Yes	297 24%	927 76%	1224 100%
	No	1295 19%	5480 81%	6775 100%
	Total	1592 23%	6407 77%	7999 100%

rect indication of the effect of the disfluency since other factors affecting the form of words might be systematically associated with disfluencies. We therefore evaluated the effect of disfluencies in regression models after controlling for the control factors listed above in Sec. III F, for the predictability variables listed in Table IX below, and for relevant interactions among these variables.[6]

Neighboring disfluencies exert a strong influence on duration and on the frequency of full versus reduced vowels, in addition to effects of the control and predictability variables. They also moderately affect the frequency of basic vowels, but have no significant effect on coda deletion. The estimated magnitudes and significances of the effects are summarized as follows.

(i) **Duration:** words in disfluent contexts are 1.34 times longer [$F(1,6200)=353.8$, $p<0.0001$].
(ii) **Vowel reduction:** the odds of a word containing a full, unreduced vowel in a disfluent context are 1.68 times greater ($\chi^2=45.9$, $p<0.0001$).
(iii) **Basic vowel:** the odds of a basic vowel form of a word occurring in a disfluent context are 1.23 times greater [$\chi^2(1)=5.8$, $p<0.02$].

Examining Table IV in more detail suggests that preceding and following disfluencies have different effects, and furthermore, that following disfluencies exert a stronger effect than preceding ones. We need to address the following questions.

(i) Is the effect of a disfluency before a word independent of the effect of one after the word?
(ii) When disfluencies occur before and after a word, are their effects cumulative? Multiplicative?
(iii) Are the effects of a disfluency after a word greater than the effects of one before a word?

Table V shows that disfluencies are more likely to occur in the presence of another disfluency. Although the increase in the likelihood of a disfluency in one position given that one occurs in the other position is not large, the association is highly significant [$\chi^2(1)=17.3$, $p<0.0001$].

The effects are multiplicative. (Recall that the response variables in the regression analysis are logs of duration or of odds, so that additivity of factors in the regression model corresponds to multiplicativity of the untransformed variables.) Regressions with the variables for preceding and for following disfluencies show no significant effect for the interaction between the two. These results suggest, at least for duration and vowel reduction, that models with separate variables for preceding and following disfluencies are preferred to ones with a single variable for disfluencies in either position. The estimated magnitudes and significances of the effects are summarized in Table VI.

Since the effects are multiplicative, the effect on a word both preceded and followed by a disfluency is given by the product of effects in Table VI. For example, the estimated duration of such a word is 1.87 times that of a word not next to a disfluency (1.22×1.51, with rounding errors). Although the effects in Table VI are qualitatively comparable to those that could be derived from the uncontrolled observations in Table IV, they are in general smaller, and in some cases, much smaller. The estimated duration effects in Table VI, for example, 1.22 for preceding and 1.51 for following disfluencies, compared to effects of 1.30 and 2.10, respectively, derived from the observed average durations in Table IV.

Turning now to following versus preceding disfluencies, the effect of a following disfluency is greater than for a preceding one for duration [$F(1,6200)=55.7$, $p<0.0001$].[7] The difference between the two, however, is not significant for vowel reduction. Nor is it significant for the basic vowel variable. In summary, then, effects on duration are clearly best modeled with separate factors for preceding and following disfluencies. For these data, simpler single-factor models are adequate to account for the effects on the presence of full vowels and of basic vowels.[8]

Disfluencies appear to affect duration more strongly than the other measures of reduction. This pattern is repeated for the other factors that are discussed in successive sections. One obvious reason for this might be that the duration of a word encompasses all lenition factors, whereas the categorical variables target more specific ones. This raises the question of the interdependence of the response variables. Here, we focus on one important aspect of this general issue: Are the effects on duration simply consequences of the shortening effects of vowel reduction, nonbasic vowels, and coda deletion? The answer is emphatically no. As one would expect, all the categorical variables, especially vowel reduc-

TABLE VI. Estimated magnitudes and significance of the effects of disfluencies before and after a target word. The magnitudes for duration are the regression estimates of how much longer words are in the disfluent context. For the full vowel variable, they are estimates of the increase in the odds of occurrence of a full vowel in a disfluent context, compared to a fluent one.

	Disfluency before		Disfluency after	
Response variable	Effect	Significance	Effect	Significance
Duration	1.22	$F(1,6200)=120.5$, $p<0.0001$	1.51	$F(1,6200)=322.5$, $p<0.0001$
Full vowel	1.59	$\chi^2(1)=27.8$, $p<0.0001$	1.68	$\chi^2(1)=18.5$, $p<0.0001$

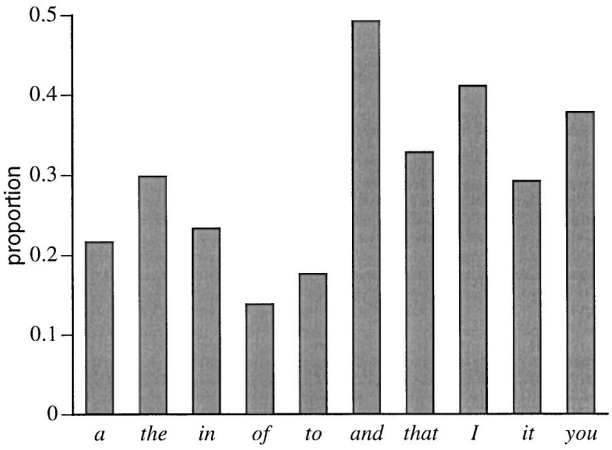


FIG. 6. Proportion of occurrences in a preceding or following disfluent context for each function word. Overall, the proportion of words occurring in disfluent contexts is 0.32. The values are based on 8045 observations.

tion, do significantly affect duration. Nevertheless, after controlling for reduced and basic vowels, the effects of preceding and following disfluencies on duration are still very strong: 1.19 times longer after a disfluency [$F(1,6198)=101.3$, $p<0.0001$] and 1.48 times longer before a disfluency [$F(1,6198)=314.1$, $p<0.0001$]. Moreover, since there is no interaction between presence of preceding or following disfluencies and vowel reduction, disfluencies lengthen full vowels and reduced vowels in the same way.

There are a number of significant interactions of the disfluency variables with rate, context variables, age of speaker, and following word predictability variables. These interactions indicate that the effects of disfluencies vary to some degree for higher or lower values of the interacting variables. The effects are relatively small and mostly limited to effects on duration.

B. Items and disfluencies

The frequencies of occurrence of the ten function words in disfluent contexts vary widely. Figure 6 shows the proportion of observations of each function word in a disfluent context, either preceded or followed by a disfluency or both.[9] A general grouping by syntactic function is evident here. The complementizers/conjunctions *and* and *that* and the pronouns have the highest rates of occurrence in disfluent contexts, while the prepositions and the articles have the lowest rates. This suggests, not surprisingly, that syntactic class plays a role in the form and behavior of function words. Consideration of this issue is limited here to the remarks about some effects of the collocation *you know* and the binomial construction *X and Y* in Secs. IV B 2, V B 1, V B 2, and VI B below; see also Jurafsky *et al.* (2002).

More crucially for the assessment of an overall effect of disfluencies on reduction, the function words more likely to occur with disfluencies—*and*, *that*, *I*, *it*, and *you*—are in general both longer (especially *and* and *that*) and more frequent overall than the words occurring less frequently with disfluencies. This has the consequence that the average disfluent duration for all the words will be longer than the average fluent duration, even if there were no difference between each word's average duration in fluent and disfluent contexts. It is thus necessary to examine disfluency effects for the individual words before accepting the results of Sec. IV A above as valid.

Table VII summarizes the effects of disfluencies for the ten function words.

Examining first the effects on function word durations, longer durations are found in the presence of disfluencies for all ten of the function words, thus confirming the general effect. The effect of a following disfluency is more general than the effect of a preceding one, in parallel with the stronger overall effect found for following disfluencies. Since *in* is the least frequent of the words, failure to find a significant effect for a preceding disfluency is possibly due to the small sample; we did not explore other possibilities. There is, however, clearly no effect of a preceding disfluency for *you*. In Sec. IV B 2 below, it will be seen that this is likely due to two facts: (1) most of the preceding disfluencies occurred before *you know*, and (2) the *you* in *you know* is reduced rather than lengthened.

On the other hand, effects on vowel quality (whether the vowel was full or reduced, and whether full vowels were the word's basic vowel or another vowel) were spottier, judging from analyses of the individual words. The results support a general effect of less vowel reduction next to disfluencies,

TABLE VII. Significances of the effects of neighboring disfluencies on individual function words. Preceding and following disfluencies have been collapsed for the vowel reduction and basic vowel variables.

Effect on	*a*	*the*	*in*	*of*	*to*	*and*	*that*	*I*	*it*	*you*
Duration by a following disfluency	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Duration by a preceding disfluency	<0.0001	<0.0001	ns	<0.005	0.01	<0.0001	0.02	<0.0001	<0.0001	ns
Reduced vowel by any disfluency	<0.0001	<0.05	0.001	0.01	<0.02	<0.0001	ns	ns	ns	ns
Basic vowel by any disfluency	ns	0.02	ns	ns	ns	<0.01	ns	ns	<0.02	ns

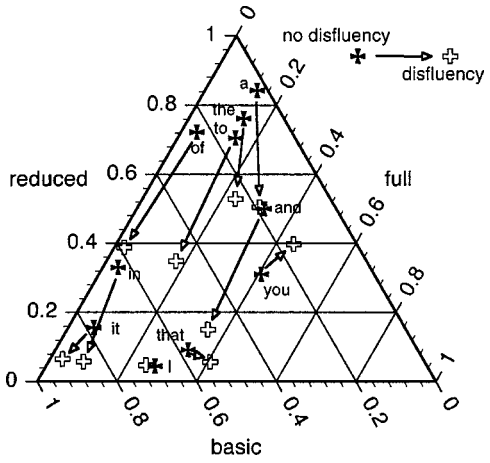


FIG. 7. Observed proportions of basic, full, and reduced vowels for the ten function words in nondisfluent contexts and in disfluent contexts. For each data point, the proportions of the three vowel categories sum to 1.0. Hence, the term full is used here in the special sense of not reduced and not basic. The proportions are based on 8045 observations, 5480 in nondisfluent contexts, 2565 in disfluent contexts.

since they reach significance for six of the function words for combined effects of preceding and following disfluencies, and, as we see below, the lack of significance for *I*, *it*, and *that* may be due to a ceiling effect. It is not clear that there is a general effect of more basic vowels next to disfluencies, though, since only three words show effects individually. The overall picture is easier to evaluate by combining the effects on reduced and basic vowels and examining them together for all the words, as presented in Fig. 7. (Note that in this figure, and in this section, we use the term ''full'' in the sense of nonbasic full, unlike the earlier use to mean any unreduced vowel.) In fluent contexts (indicated by filled crosses in the figure), the words vary greatly in the relative frequency of the vowel classes, basic, (nonbasic) full, and reduced, and this likely plays a role in how they are affected by neighboring disfluencies. In this figure, an arrow pointing down indicates that a word has fewer reduced vowels in disfluent contexts; if the arrow slants to the left, it also indicates that a word has a greater proportion of basic to nonbasic vowels in disfluent contexts. As one would expect from the overall results, most of the words exhibit one or both of these relations.

You is clearly anomalous, showing if anything an effect of disfluency in the opposite direction of the other words; some reasons for this surprising behavior are explored in Sec. IV B 2. The words along the left edge of the figure—*of*, *in*, and *it*—essentially have no nonbasic full vowels; hence, in a disfluent context, basic vowel frequency increases at the expense of reduced vowels. The lack of a significant increase in reduced vowel frequency for the three words at the bottom of the figure—*I*, *it*, and *that*—can possibly be attributed to their already very low rates of reduction, 16 percent or less. The lack of any increase in the basic vowels of *I* and *that*, on the other hand, seems to be a true item characteristic.

The four words at the top of the figure, which have the highest proportions of reduced vowels, over 70 percent, all showed significant decreases in reduced vowels, as might be expected. *The* is the only one of the four whose increase in basic vowels is significant, although the observed leftward slants for *of* and *to* in the figure suggest stronger basic vowel effects for them. The large sample size for *the* may be a factor here. *A*, *of*, and *to* are among the less frequent of the function words and their many reduced vowels leaves few items to test basic vowel effects over ($n=141$, 156, and 209, respectively). The significant basic vowel effect for *the* stands out in contrast, since its sample size is only modestly higher ($n=268$).

And, which alone of the ten words has a relatively even balance among the three vowel categories, shows the strongest effects of disfluency contexts for both reduced vowels and basic vowels. Recall that it also has the highest rate of occurrence next to disfluencies.[10]

The overall picture suggests that there is generally less vowel reduction in the neighborhood of disfluencies, with the unexplained exception of *you*, possibly diminished in strength for items which already have few reduced vowels in fluent contexts. An increased number of basic vowels in disfluent contexts is clearly not general, and is likely to be a word-specific characteristic. Since contextual selection of lexical variants is an important source for variation between basic and other full vowels, further examination of how this differs for different words is warranted. Sorting out this and other differences will clearly take much more detailed study of the individual words and their contexts.

1. Initial disfluencies and and

Not only is *and* generally more frequent, longer, and more likely to occur with disfluencies, it is much more likely than the other words to occur in utterance-initial position, making up 48 percent of the function words there (Sec. VI A). This raises the question about the role of *and* in the preference, suggested by earlier research, for disfluencies to occur in initial positions.

Shriberg (1994), for example, showed that disfluencies were more likely to occur sentence initially than sentence medially, in three corpora (Switchboard, ATIS, and American Express) ($p<0.0001$). In addition, Clark and Wasow (1998) suggested that disfluencies were more likely to occur at the beginning of large constituents like clauses than at the beginning of smaller constituents like words or phrases. This would presumably also result in a larger numbers of disfluencies in utterance-initial position. Results from our data on utterance-initial position agree with Shriberg's. After controlling for the variables mentioned in Sec. VI A below, we found that initial words in Switchboard are more likely to be disfluent than medial words ($p<0.0001$). More specifically, filled and unfilled pauses (although not repetitions) are more likely to occur after the first word than after medial words ($p<0.0001$). The greater likelihood of filled or unfilled pauses after initial words was, however, **only** true for *and*. For the other nine words, after removing the word *and*, there was no effect of increased disfluency rate on initial words. The rate of following disfluencies was the same for utterance-initial words and for noninitial words, 13 percent. Within our corpus the initial preference for disfluencies appears to be idiosyncratic to *and*. In a larger perspective it is likely to be related to the frequent use of *and* as a discourse

TABLE VIII. Observed average durations (ms) of function words in fluent and disfluent contexts. The number of observations appears in parentheses. The values are based on a sample excluding items beginning or ending a fragment, i.e., similar to the sample used in the regression analyses in this section.

	Another word		Silence		Filled pause		Repetition	
Preceded by	115	(5694)	145	(510)	147	(104)	201	(155)
Followed by	108	(5885)	187	(318)	307	(174)	186	(132)

marker. Discourse markers tend to occur initially in turns and utterances (Schiffrin, 1987). Perhaps such initial discourse markers tend to be followed by a filled or unfilled pause. A very preliminary survey over the entire 38 000-word set of Switchboard phonetic transcriptions supports this conjecture. First, turn-initial words are more likely to be followed by filled pauses or silence than noninitial words, 22 percent compared to 16 percent. Second, the vast majority of the initial disfluent words are words which frequently act as discourse markers. This suggests that the prevalence of silence and filled pauses in initial positions may be a fact more about discourse markers than about turn and utterance position.

2. The collocation you know

A number of characteristics of *you* stood out with respect to disfluencies: it was among the words most likely to occur with disfluencies, it was much more likely to occur after rather than before a disfluency than the other words, it showed no lengthening effect after a disfluency, and it showed no decrease of frequency of reduced vowels in disfluent contexts. All but the last of these can be attributed to the frequent occurrence of *you* in the collocation *you know*. This combination makes up 47 percent of the occurrences of *you* in our data. Since most of these are lexicalized fillers or editing terms, it is not surprising that the form of *you* tends to be reduced: *you* is about 25 percent shorter and about twice as likely to have a reduced vowel in *you know* than in other contexts.

You know itself very frequently occurs after a disfluency, which contributes to the apparent high rate of occurrence of *you* in disfluent contexts. Excluding *you know, you* is somewhat less likely than most of the function words to have a neighboring disfluency. The predominance of occurrence of *you* after rather than before disfluencies is partly an artifact of *you* in *you know* being almost always coded as having no following disfluency (pauses rarely separate the collocation), and partly because of the frequent occurrence of *you know* after a disfluency. In other contexts, *you* is only moderately more likely to occur after rather than before a disfluency.

The shorter and more reduced forms of *you* in *you know* obviously distorted the analyses of the effects of neighboring disfluencies. The reduced *you know*'s will count as fluent items for the following position, and hence will exaggerate the effect of a following disfluency. They will very frequently be among the disfluent items for the preceding position, and hence will dilute the effect of a previous disfluency. Indeed, when *you know* items are excluded, the effect of following disfluencies on duration is diminished, but it remains quite strong, especially considering the smaller sample [$F(1,296)=14.0$, $p=0.0002$]. And, without the *you know* items, a preceding disfluency appears to lengthen *you* [$F(1,295)=4.8$, $p<0.05$]. One the other hand, it is still the case that *you* shows no decrease in the frequency of reduced vowels in disfluency contexts when the *you know* items are excluded. It is true that the overall rate of reduced vowels is decreased to 24 percent from 66 percent (cf. Fig. 5) by the exclusion of *you know*. Since this is still well above the levels of *I*, *it*, and *that*, it is not likely that a floor effect could keep the presence of a disfluency from reducing it further, as seems plausible for the lower reduced vowel rates of *I*, *it*, and *that. You*'s vowel reduction behavior thus remains an anomaly, all the more puzzling given the evident duration effects.

C. Differential effects of disfluency types

Does the effect of disfluent items on neighboring function words extend equally to each kind of disfluency that we have considered? We address this question here mainly to be assured that the effects described above are attributable in some degree to all of the disfluencies, in keeping with their assumed status as indicators of planning problems. The limitations of our database, which focuses on individual words in a very local context, precludes any analysis of the structure of disfluencies beyond the grossest details. One of the reasons for this is that disfluencies often are not simply silent pauses, filled pauses, or repetitions, but larger events combining some or all of these, as well as editing terms, as this Switchboard example shows:

…built up in um PAUSE in the PAUSE in the PAUSE uh bureaucracy….

Some of the more detailed questions about the form structure of disfluencies are treated in O'Shaughnessy (1992), Plauché and Shriberg (1999), and Shriberg (1994, 1999).

The observed average durations of function words in fluent and in different disfluent contexts are compared in Table VIII. The significances of duration differences reported below are, however, based on regression analyses controlled for the same variables described above. The durations for filled pauses and for repetitions in Table VII cover only the simple cases not combined with a silence.

When they precede a word, all three disfluency types have a lengthening effect. The significance of the effect is least for filled pauses [$F(1,5449)=11.0$, $p<0.001$]. The significances of the other effects are $p<0.0001$. The lengthening effect of a repetition is stronger than the effect of silences and filled pauses [$F(1,6058)=34.1$, $p<0.0001$]. The effects of silences and filled pauses do not differ significantly.

All three types also have a lengthening effect when they follow a word. Again, the effect is weakest, but nevertheless

highly significant, for filled pauses [$F(1,5756)=20.9$, $p<0.0001$]. There are no significant differences between the effects of the different disfluencies, in spite of the apparently much longer durations before filled pauses.

D. Discussion

Function words which are preceded by or followed by disfluencies are longer and are more likely to have full vowels than words in fluent contexts. These effects are robust; all ten function words are longer when followed by disfluencies, and eight of ten when preceded by disfluencies. Effects on basic vowel frequency and coda presence, on the other hand, appear to depend on the lexical item or possibly, in the case of coda presence, the identity of the coda obstruent. Disfluencies after a word affect the word's form more strongly than disfluencies before a word. Preceding and following disfluencies tend to co-occur, and when they do, their effects are multiplicative. Finally, all three disfluency types have a lengthening effect.

V. WORD PREDICTABILITY FROM NEIGHBORING WORDS

In earlier work (Jurafsky *et al.*, 2001; Gregory *et al.*, 1999) we proposed the *probabilistic reduction hypothesis*: words are more reduced when they are more predictable or probable. In this section we focus on the extent to which the probability of a word given neighboring words affects reduction. There are many ways to measure the probability of a word. The simplest measure, *prior probability*, can be estimated from the relative frequency of the word in a sufficiently large corpus. The fact that the 10 words in this dataset were all very frequent, however, limited our ability to study relative frequency. The 3-to-1 range of frequency of the words is very small compared to the overall ratio of probability of about 100 000 to 1 for the highest and lowest frequency words in the entire 38 000-word phonetically transcribed portion of Switchboard. What variation there is, moreover, is inextricably confounded with the effects of form and patterns of combination of the individual items. Consequently, one cannot make useful inferences about the effects of relative frequency with the function words dataset.

We therefore limit our focus to the effect of neighboring words on predictability. Consider first the predictability of a word given the previous word. We use two measures of this. One is the *joint probability* of the two words $P(w_{i-1}w_i)$. The joint probability may be thought of as the prior probability of the two words taken together, and is estimated from the relative frequency of the two words together in a corpus. This is computed by counting the number of times the two words occur together, $C(w_{i-1}w_i)$, and dividing by N, the number of words in the corpus

$$P(w_{i-1}w_i)=\frac{C(w_{i-1}w_i)}{N}. \tag{1}$$

This is a variant of what Krug (1998) called the *string frequency* of the two words.

Used alone, joint probability is not an entirely satisfactory measure of word predictability. Pairs of words can have a high joint probability merely because the individual words are of high frequency (e.g., *of the*). But a word can occur infrequently, yet be very predictable every time it occurs. Thus most measures of predictability are based on metrics like *conditional probability* or *mutual information* which control for the frequencies of one or both of the words (Manning and Schütze, 1999). The second metric we use in this paper is such a metric: the *conditional probability of a word given the previous word*. This is also sometimes called the *transitional probability* (Saffran *et al.*, 1996b; Bush, 1999). The conditional probability of a particular target word w_i given a previous word w_{i-1} is estimated by counting the number of times the two words occur together $C(w_{i-1}w_i)$, and dividing by $C(w_{i-1})$, the occurrences of the first word

$$P(w_i|w_{i-1})=\frac{C(w_{i-1}w_i)}{C(w_{i-1})}. \tag{2}$$

In addition to considering the preceding word, the effect of the following word may be measured by the two corresponding probabilities. The *joint probability of a word with the next word* $p(w_iw_{i+1})$ is estimated from the relative frequency of the two words together

$$P(w_iw_{i+1})=\frac{C(w_iw_{i+1})}{N}. \tag{3}$$

Similarly, the *conditional probability of the target word given the next word* $p(w_i|w_{i+1})$ is the probability of the target word w_i given the next word w_{i+1}. This measures the predictability of a word given the next word the speaker is about to say, and is estimated by

$$P(w_i|w_{i+1})=\frac{C(w_iw_{i+1})}{C(w_{i+1})}. \tag{4}$$

As we see below, while conditional probabilities are the most consistent of these factors affecting reduction, joint probabilities and the relative frequencies of surrounding words also contribute additional effects. It is thus helpful to consider their relationship with the conditional probabilities. The fundamental relationship among them is given by

$$P(w_i|w_x)=\frac{P(w_iw_x)}{P(w_x)}, \tag{5}$$

where w_x denotes either the preceding or the following word. (This can be derived from the definitions above.) Since we use log probabilities as factors in the regressions to assess effects, conditional probability as a single factor with weight B ($B\times$log conditional probability) is the same as the combination ($B\times$log joint probability)$-$($B\times$log relative frequency of the neighboring word). We can thus think about the conditional probability as combining the effects of joint probability and the relative frequency of the neighboring word under the simple assumption that they have equal (but opposite) weights. If we find that either joint probability or neighboring relative frequency [but not both; by Eq. (5) any third term of the three is redundant] contributes an additional effect, this tells us that the assumption of equal weights is incorrect, and that the combined effect of the probabilities is more complex. Since any two of the three probabilities in (5) capture all of the predictability effects of a neighboring

TABLE IX. Summary of probabilistic measures and high probability examples.

Measure	Definition	Examples
Joint of target with next word	$p(w_i w_{i+1})$	**you** *know*, **I** *think*
Joint of target with previous	$p(w_{i-1} w_i)$	*and* **I**, *in* **the**
Conditional of target given previous	$p(w_i \mid w_{i-1})$	*rid* **of**, *kind* **of**
Conditional of target given next	$p(w_i \mid w_{i+1})$	**I** *do*, **you** *know*
Conditional of target given surrounding	$p(w_i \mid w_{i-1} \cdots w_{i+1})$	*matter* **of** *fact*

word, we have somewhat arbitrarily chosen to examine the conditional probabilities and the joint probabilities. Where both probabilities significantly affect reduction, we interpret the joint probability effect as either an indication that the joint probability is more heavily weighted than the neighboring word's relative frequency (in the case of less reduction with higher joint probabilities) or an indication that it is the relative frequency that is to be more heavily weighted (when there is more reduction with higher joint probabilities).

Table IX contains a summary of the probabilistic measures and some examples of high probability items from the dataset for each measure.

Other more complex conditional probabilities, often called *trigram probability* measures, played a smaller role in the analysis. Two of these were the *conditional probability of the target given the two previous words* $p(w_i \mid w_{i-2} w_{i-1})$, and the *conditional probability of the target given the two following words* $p(w_i \mid w_{i+1} w_{i+2})$. Neither of these turned out to have any effect on word forms. The other is the *conditional probability of the target given the two surrounding words* $p(w_i \mid w_{i-1} \cdots w_{i+1})$, estimated as follows:

$$P(w_i \mid w_{i-1} \cdots w_{i+1}) = \frac{C(w_{i-1} w_i w_{i+1})}{C(w_{i-1} \cdots w_{i+1})}. \tag{6}$$

We have also considered the *mutual information* (Fano, 1961) of the target word and the neighboring words in Gregory *et al.* (1999). There we showed that mutual information produces very similar results to the conditional probability of the target word given the neighboring word.

The actual computation for estimating these probabilities is somewhat more complex than the simple explanations above. Since the 38 000-word ICSI corpus is far too small to estimate word probabilities, they are estimated from the 2.4 million-word Switchboard corpus instead. We trained these probabilities via three separate stochastic grammars: a regular bigram grammar (conditioned on previous word), a reverse bigram grammar (conditioned on following word), and a centered trigram grammar. The counts were smoothed by Katz backoff with Good–Turing discounting (Jurafsky and Martin (2000), pp. 214–219, and references therein).

A. Effects of predictability

The main results that are reported in the following sections are based on regressions with the control variables listed above in Sec. III F, the preceding/following disfluency variables described in Sec. IV, and the relevant interactions among these variables. The reported results are based on a sample that excludes fragment-initial and fragment-final items, leaving a total sample of 6219 items. Separate analyses, which we do not report in detail, verified that adding additional control variables for the effects of prosodic position would not have materially changed the results. We thus chose not to use the smaller sample of some 4800 items of the ICSI corpus segmented by the LDC into sentence-like domains. See Secs. VI and VI A below for a fuller explanation of this sample and its prosodic coding. No analysis of predictability effects on basic vowel frequency or coda presence was attempted, partly because of the reduced sample sizes for those variables, and partly because they would reveal less about general effects of predictability, since their behavior varies much more from word to word.

1. Word duration

The predictability factors having the strongest effects on word duration are the conditional probabilities of the target word. As can be seen in Table X, both the conditional probability given the previous word $p(w_i \mid w_{i-1})$ and the conditional probability of the target word given the following word $p(w_i \mid w_{i+1})$ are highly significant factors. Target words which are more predictable are shorter. That is, the higher the conditional probability of the target given either of the neighboring words, the shorter the target word, as indicated by the effect magnitudes less than 1.0 in Table X. The shortening ratios used to measure effect magnitudes can be made more concrete by applying them to tokens which have typical values for other variables. This yields durations predicted by the regression models which include the other variables, as opposed to observed average durations, which are uncontrolled. Such words, if they are highly probable given the previous word (at the 95th percentile of the conditional probability), have a predicted duration of 90 ms; low conditional probability tokens (at the 5th percentile) have a predicted duration of 109 ms. The duration of words is affected similarly by their probability given the following word: highly probable tokens have a predicted duration of 86 ms; tokens with a low probability given the following word have a predicted duration of 116 ms.

There are also significant additional effects of the joint probabilities with previous and following words. When words have a higher joint probability with the following word, they are *longer*. This effect is in the opposite direction than the one we find with the conditional probabilities, and to some extent counterbalances the shortening effect of the conditional probability. In contrast, words with a higher joint probability with the previous word are shorter, affecting duration in the same way as the conditional probability. Moreover, there is a significant interaction between conditional probability given the previous word and joint probability with the previous word. This interaction captures some of the ways that the effects are uneven over the range of probabilities: joint probability has a shortening effect only for tokens whose conditional probability is above the median; and the shortening effect of conditional probability is greater for higher joint probabilities.

Thus predictability effects of the previous word and of the following word are similar in that both conditional prob-

TABLE X. Significances and magnitudes of effects of predictability variables on word duration and frequency of full vowels. The significance of each variable is obtained by adding it to a comparison regression model. The comparison model consists of the control variables for the preceding and following conditional probabilities; of the control variables plus the corresponding conditional probability for the joint probabilities; control variables plus the preceding conditional and joint probabilities for the interaction; and control variables plus all the other probability variables for the centered conditional probability. The values of F thus have degrees of freedom between $F(1,6197)$ and $F(1,6202)$. The effect magnitudes are ratios of length and ratios of odds of full vowels. They are estimated by evaluating the coefficients of the variables in the full regression equation over the range between the 5th and 95th percentiles of each variable. Effects for previous conditional and joint probabilities include the interaction, evaluated at median values of the variables.

Predictability variable	Duration			Full vowel proportion		
		Significance			Significance	
	F	p	Effect	$\chi^2(1)$	p	Effect
Conditional of target given previous	88.4	<0.0001	0.80	92.9	<0.0001	0.24
Joint of target with previous	43.7	<0.0001	0.94	55.2	<0.0001	2.44
Previous conditional×joint interaction	58.7	<0.0001		20.4	<0.0001	
Conditional of target given next	186.0	<0.0001	0.72	22.3	<0.0001	0.27
Joint of target with next	41.6	<0.0001	1.20	272.8	<0.0001	5.39
Conditional of target given surrounding	20.4	<0.0001	0.91	2.9	0.09	

abilities have shortening effects; they differ in that higher joint probability with the previous word shortens a word, but higher joint probability with following word lengthens it. In addition, no interaction was found between the previous word probabilities and following word probabilities.[11]

One further conditional probability affects word durations in addition to the variables above—the conditional probability of the word given both the previous and following words. Like the other conditional probabilities, tokens with higher conditional probabilities are shorter, but the effect is somewhat less. The predicted duration of tokens with high probabilities is 95 ms, whereas that of tokens with low probabilities is 104 ms. No significant contributions of probabilities were found involving the word before the previous word or the word following the following word (i.e., using the other trigram conditional probabilities described above). In other words, we are able to discern only strictly local probability effects, limited to the interaction of a word with the word next to it.

2. Vowel reduction

Neighboring word predictabilities affect vowel reduction in much the same way as they do word length. The conditional probability given the previous word and given the following word are both strongly associated with higher frequencies of reduction. The predicted likelihood of a full vowel in words which were highly predictable from the following word (at the 95th percentile of conditional probability) was 0.43, whereas the likelihood of a full vowel in low predictability words (at the 5th percentile) was 0.73. The predicted likelihoods for words with high and low predictability from the previous word were very similar, 0.43 and 0.72, respectively.

Again, there are also strong effects of the joint probabilities with previous and following words. For vowel reduction, however, a higher joint probability in *either* direction is associated with less reduction. Words with higher joint probabilities with either the previous or the following word are more likely to have full vowels, counterbalancing the reduction effect of the conditional probabilities. As with duration, the interaction between conditional probability given the previous word and joint probability with the previous word is highly significant, reflecting the same sort of variation in magnitude of effects that was described above for duration.

No significant additional effect was found from the preceding and the following words together. As with duration, there were no interaction effects between previous and following word predictability variables, nor were there any effects due to predictabilities involving words before the previous word or after the following word.

3. Interdependence of duration and vowel reduction

The strong effects of predictability on both shortening and on vowel reduction suggest that there may be separate sources for the two effects. Perhaps vowel reduction stems mainly from some sort of categorical choice in lexical production between full and reduced vowels, whereas shortening is mainly the result of gradient, noncategorical modifications at the level of phonetic encoding or of execution of the articulatory plans.[12] It is possible, however, that the shortening effects that we observe for function words might be solely a consequence of the vowel reduction effects, since reduced vowels are shorter than full vowels. If this were true, there might be no evidence for a gradient affect of probability on reduction. In order to test whether the effects of probability on shortening were completely due to vowel reduction, we added the full versus reduced vowel variable to the base model for duration as a control.

The probabilistic variables remain significant predictors of duration after controlling for vowel reduction. The vowel reduction variable of course accounts for a considerable amount of the duration variance (14.5 percent), so there should be less for the predictability variables to account for. Indeed, the predictability variables account for 3.8 percent of the variance in duration overall, but 2.6 percent of the variance in duration controlled for reduction. Nevertheless, except for the joint probability with the following word, all the individual predictability variables remain highly significant

TABLE XI. Significances of the effects of predictability variables on individual function words. Effects with significances above 0.01 are in boldface.

Effect on	a	the	in	of	to	and	that	I	it	you
Duration by conditional given following	**<0.05**	<0.001	<0.0001	ns	<0.0001	ns	<0.0001	ns	<0.005	<0.0001
Duration by conditional given previous	**0.05**	<0.001	ns	ns	0.0002	ns	ns	ns	**<0.02**	ns
Duration by joint with previous	ns	**0.02**	ns	<0.0001	<0.01	ns	ns	ns	**<0.05**	ns
Reduced vowel by conditional given following	ns	ns	0.0002	ns	0.0005	ns	ns	ns	ns	<0.0001
Reduced vowel by conditional given previous duration	ns	ns	**<0.05**	ns	ns	<0.0005	ns	**<0.05**	ns	ns

at levels of $p<0.0001$. Predictability not only affects whether vowels are reduced or not, but it has an additional noncategorical effect on word duration.

Further confirmation results from an examination of the words with full and with reduced vowels separately, to see whether predictability shortening affects full vowels as well as reduced vowels. Even with the smaller subsamples, the probability variables remain highly significant at levels of $p<0.0001$, with a few exceptions. The joint probability with the following word is a significant factor for reduced vowels ($p<0.005$), but not for full vowels ($p=0.15$); and conditional probability given the previous word is only marginally significant for full vowels ($p<0.01$). We also verified that the possibly categorical deletion of final obstruents in the words *and*, *it*, *of*, and *that* did not account for the predictability effects on duration within the reduced and full vowel subsamples.

B. Variability of predictability effects by word

Individual analyses of the function words show that each word's duration is affected by one or more of the predictability variables. Table XI summarizes the effects on both duration and vowel reduction for the conditional probabilities given the previous word and given the following word. It also includes the effect on duration of the joint probability with the previous word. The most general effect is that of the conditional probability given the following word, affecting the duration of six of the words. The words showing no effect or only a marginal effect of this variable, *a*, *of*, *and*, and *I*, are scattered across functional categories and include both high- and low-frequency words. Thus it does not seem possible either to attribute the pattern of effects to limitations to particular classes of words or to attribute the exceptions generally to a lack of sensitivity of the analysis due to small sample sizes. The predictability variables involving the previous word clearly affect *the*, *of*, and *to*. In addition, the interaction between the conditional probability given the previous word and the joint probability with the previous word is a significant factor for five of the words. These include *of* and *to*, indicating that the conditional probability affects the duration of the word more when it occurs in a frequent combination with a previous word. The interaction is also a factor for *in* and *I*, suggesting that, although neither the conditional nor the joint probability is significant alone, that there may be an effect of the conditional probability for frequent combinations. Finally, there are marginal effects on duration of the bilateral conditional probability given previous and following words for *a*, *and*, *that*, and *to*, with significance values ranging from <0.01 to <0.05. As for vowel reduction effects, it is evident that they are less general than those for duration. This parallels the pattern found for lengthening in disfluency contexts.

Overall, these results confirm the hypothesis that words in more predictable contexts have more reduced forms, since an effect for some predictability variable was found for each of the function words. On the other hand, the considerable variation in the strength of the effects (possibly none in some cases) underscores the importance of the interaction of each word's attributes with predictability. The hallmark of function words is that they are markers of particular pragmatic, semantic, and syntactic functions, and that they occur in particular classes of constructions. The kinds of constructions they occur in is bound to affect whether it is predominantly predictability from the left, from the right, or from both that they are subject to. Moreover, their occurrence in certain very frequent constructions may strongly influence the appearance of their overall sensitivity to predictability, since those constructions will necessarily be highly predictable contexts. While we do not explore these interesting connections here in detail, the discussions of high frequency uses of *and* and *you* that follow illustrate some of the interactions of a word's idiosyncratic behavior with predictability.

1. And *in binomial constructions*

One of the very frequent uses of *and* is as a conjunction to create binomial constructions such as *trucks and stuff*, *lockers and everything*. This immediately suggests a connection with the pattern of predictability effects on *and* discussed above, namely that *and* was one of few words to be affected by bilateral conditional probability (given both previous and following words). A very preliminary check confirms this. A fairly broad binomial category was coded by hand, which included modified and unmodified words, and adjectives and verbs as well as nouns. Excluding disfluent contexts, *and* is significantly shorter in binomials than elsewhere [$t(460)=3.65$, $p<0.0001$]. Furthermore, within binomials, *and* is significantly shorter when it is more predictable from the two surrounding words, whereas the bilateral conditional probability has no effect on the duration of *and* in its other occurrences.

2. You know *and predictability*

Recall from Sec. IV B 2 that 47 percent of the occurrences of *you* are in the collocation *you know*, and that in this context it is shorter and more likely to have a reduced vowel than in other contexts. *You* in *you know* is shorter (by about 25 ms) and much more likely to have a reduced vowel (50 percent compared to 24 percent) than *you* in other contexts.

The high frequency of the combination necessarily means that the predictability of *you* from following *know* is unusually high, 12.6 times other contexts. Its predictability from the preceding word, on the other hand, is lower, 0.42 times other contexts. This is presumably a consequence of fillers and editing terms occurring across a wide range of contexts, and hence being relatively unpredictable in any particular context. Recall from Table XI that *you* is strongly affected by predictability from the previous word, but little or not at all by the following word. The obvious question is whether this simply reflects the asymmetry of the *you know* combination, or whether it is more general. In contrast to the binomial *and* case, the results were little changed after excluding *you know: You* is shorter and more likely to have a reduced vowel when it is more predictable from the following word, but shows no effects of the predictability from the preceding word.

C. Discussion

Words that are more predictable are shorter and more likely to have reduced vowels, confirming the probabilistic reduction hypothesis introduced above. The conditional probability of the target word given the preceding word and given the following both play a role, in both duration and vowel reduction. The magnitudes of the duration effects are fairly substantial, in the order of 20 ms or more, or about 20 percent, over the range of the conditional probabilities (excluding the highest and lowest 5 percent of the items). The joint probabilities of the target words given the preceding and following words also played a role in reduction, as did the bilateral conditional probability of the target word given the two surrounding words. The local nature of the predictability variables is underscored by the lack of any effect involving words more than one word distant from the target word. The failure to find effects for all the probability variables on all the function words is possibly partly due to the smaller sample sizes, but the overall spotty pattern of effects indicates that there are real differences among the words. This sort of variation confirms the expectation that one source of the probability effects is the collocation of the function words in particular constructions. Are frequent collocations, perhaps semilexicalized, the only or primary source of the predictability observed here?

The answer seems to be no. In an earlier study (Jurafsky *et al.*, 2001), we showed that higher predictability is associated with increased reduction even in word combinations that are not lexicalized. We did this by looking at words with relatively low conditional probabilities, and showing that the effects of predictability from the preceding word hold not only for the more predictable cases, as would be expected if frequent collocations are the source of the effects, but also for the less predictable cases, which are unlikely to be lexicalized.

The fact that the effects of predictability on duration add to the effects on vowel reduction, and affect both full and reduced vowels, indicates that some of the effects of predictability on reduction are continuous and noncategorical. It is reasonable to conclude that predictability effects are not limited to lexical choice and combination at semantic and phonological form levels, so that the domains of applicability of the probabilistic reduction hypothesis include linguistic levels that allow continuous specification of phonetic form.

VI. THE POSITION OF A WORD IN PROSODIC DOMAINS

The location of a word in larger prosodic domains such as utterances, turns, intonational phrases, and phonological phrases plays an important role in reduction. Studies of language change and of pronunciation variation have long accepted three main effects—final lengthening (Klatt, 1975; Ladd and Campbell, 1991; Crystal and House, 1990, *inter alia*) initial strengthening (i.e., more extreme articulation) (Fougeron and Keating, 1997; Byrd *et al.*, 2000, *inter alia*), and final weakening (i.e., less extreme articulation) (Browman and Goldstein, 1992; Hock, 1986). During the last several decades more and more quantitative studies have helped make our understanding of these general effects more precise; see Fougeron and Keating (1997) for a review. Many of these results, however, derive from laboratory paradigms like reiterant speech, and have not been tested on natural speech production or over a wide range of lexical, prosodic, and pragmatic contexts. Furthermore, it has been difficult to tease apart prepausal lengthening from lengthening at the edge of prosodic domains.

To evaluate the effect that position in prosodic domains plays on function word reduction in conversational speech, as well as to control for positional effects in the analysis of other variables, we examine a word's position in an utterance-like domain. The domain we chose had already been transcribed for a large proportion of the Switchboard corpus by the Linguistic Data Consortium (LDC) (Meteer *et al.*, 1995), following the segmentation guidelines in Shriberg (1994). We use the term utterance for this LDC domain; Meteer *et al.* (1995) called them ''slash units.'' In general, these units are intended to model the sentence-like units which often make up spoken conversation, and hence are defined with respect to both syntactic coherence and an attempt at approximating large intonation boundaries. While this use of syntactic coherence as a heuristic for intonation boundaries is clearly inferior to a prosodic transcription of speech, the fact that grammatical boundaries and intonational boundaries are highly correlated (Croft, 1995) makes this methodological simplification less problematic.

The utterances include complete syntactic sentences.

(i) I, I have strong objections to that.
(ii) And that's not fair.
(iii) Where, where are you?
(iv) And, uh, I thought of those two things when I was, I was holding for a long time.

as well as phrases which function as complete turns

(i) And, uh, until next time.
(ii) A pop-up trailer, huh?
(iii) The news.

In most cases an utterance was contained inside a single turn. Sometimes, however, an utterance was interrupted by a

TABLE XII. Duration and vowel reduction values for function words which are in initial position in the utterance, in final position, or in medial position (noninitial, non-final).

	Initial	Medial	Final
Duration (ms)	173	125	200
Vowel reduction	82.3%	57.4%	93.2%

backchannel such as *uh-huh*, or another remark from the interlocutor. In such cases, as in the following example, A's speech was counted as one utterance; thus, the word *and* is counted as utterance initial, but the word *here* is not.

A: And, and I get mail
B: Uh-huh.
A: here at home under each of those names.

Larger turns are generally broken into utterances at syntactic boundaries which correlated with intonation boundaries.

B: And, uh, I never really, messed with anything, uh, gardening or anything like that until now,
B: but, uh, I, I keep hearing all the stories of, of different parts of town.

Readers interested in more details of the definition of utterances and the procedures followed by the LDC coders should see the coders' manual (Meteer *et al.*, 1995).

In general, utterance boundaries and turn boundaries were very highly correlated, as would be expected. For this reason, we did not examine turn-boundary position separately from utterance-boundary position. The edges of the LDC utterances should generally correspond with edges of intonational phrases (and also with edges of smaller units such as phonological phrases), whereas their interiors will sometimes contain words that are edges of intonational phrases as well as those of smaller units. Consequently, if utterance-edge strengthening effects are found, such results should be conservative.

A. Effect of utterance position

About two-thirds of the ICSI data had LDC utterance-boundary labels, so that 4777 observations were available for the analysis of utterance position.[13] Table XII shows observed values for duration and reduction in initial, medial, and final positions.

These observed differences, however, may not be valid indications of the effect of position in the prosodic domain, since other factors affecting the form of words might be systematically associated with prosodic positions. For example, Shriberg (1994) found that initial words are more likely to occur in the context of disfluencies. Since disfluencies cause words to be longer, this may exaggerate the actual effect of initial position. Initial position may have different kinds of segmental or accentual contexts than noninitial words, and may also be predictable in different ways. Pauses, which may be likely to occur after utterance-final position, would exaggerate the effect of final position. We therefore evaluated the effect of position in regression models after controlling for the factors listed in Secs. IV and V (and relevant interactions).

After controlling for all factors except predictability variables from the preceding word, initial words are longer than noninitial words [$F(1,4639)=30.1$, $p<0.0001$]. Initial words are also more likely to have a full (unreduced) vowel than noninitial words [$\chi^2(1)=192.9$, $p<0.0001$]. Conditional and joint probabilities with the preceding word were omitted from these analyses partly because they have no meaningful interpretation at the beginning of fragments, and if fragment-initial items were eliminated, the utterance-initial items would be halved, reducing the power of the analysis.

There is a more fundamental consideration, however. Low predictability is an expected characteristic of utterance-initial words, and can be expected to mask the effect of utterance position. This is the case. There remains no additional effect of initial position after adding the predictability variables as controls ($p=0.16$). (This analysis is based on the smaller subcorpus that excludes the fragment-initial items for which the predictability variables are not defined.) Low predictability, however, might well be considered to be an inherent characteristic of the position. This makes it unclear whether it is even appropriate to control for predictability. Analytically, the predictability variables mask the initial position effect, but the proper interpretation of this result awaits a deeper understanding of the interaction of predictability and prosodic domains than we possess.

Final position has long been known to play a role in lengthening (Klatt, 1975; Ladd and Campbell, 1991; Crystal and House, 1990, *inter alia*). As Table XII shows, the observed durations for final words are longer. After controlling for all factors except predictability variables from the following word, utterance-final words are longer than medial words [$F(1,3992)=25.5$, $p<0.0001$]. They are also more likely to have unreduced vowels [$\chi^2(1)=7.8$, $p=0.005$]. The utterance-final effect is not as sensitive to the masking from conditional and joint probabilities with the following word—utterance-final words are still longer [$F(1,3721)=12.7$, $p<0.0005$], and more likely to have unreduced vowels [$\chi^2(1)=7.7$, $p<0.01$] when controlled for these probabilities within fragments. Under those conditions, the estimated lengthening factor of final position is 1.23; and a word which occurred with a full vowel 60 percent of the time in medial position would have an estimated frequency of occurring with a full vowel in final position of 81 percent.

We do not report on the effect of position on the percentage of basic vowels or of coda deletion. Both these measures seem to be strongly affected by individual items. The results are difficult to interpret, but probably reflect specific high-frequency combinations of the function words with other words.

B. Variability of position effects by word

The final lengthening effect applies very generally; all ten function words are longer at the end of utterances. In contrast, only five words, *a*, *and*, *it*, *that*, and *the* have longer durations at the beginning of utterances than medially.

In addition, utterance-initial position is overwhelmingly dominated by the function words *and* and *I*—*and* makes up 48 percent, and *I* 32 percent of the function words in that position. Recall that *and* is also the longest of the function

words (Sec. III E). Is the combination of *and*'s length and frequent occurrence in initial position responsible for the utterance-initial lengthening effect above? Excluding *and*, an effect, although somewhat weaker, remains [$F(1,4078) = 6.8$, $p < 0.001$]. In addition, *and* alone shows a significant initial effect [$F(1,544) = 6.0$, $p < 0.02$]. On the other hand, there is no effect for *I* [$F(1,794) < 1$]. Thus, in contrast to our finding (Sec. IV B 1) that the association between initial position and disfluencies is limited to *and*, we conclude that the initial lengthening effect is not an artifact of the disproportionate number of longer *and*s initially, but applies more generally. The bias introduced by *and* simply exaggerates the general effect. The lack of an effect for *I*, however, indicates that there is no or little initial lengthening effect for some words, presumably due to idiosyncratic properties that we have not explored.

C. Discussion

The high-frequency function words studied here are longer and more likely to have full vowels at the beginning and end of the utterance-like domains coded by the LDC. Our results thus show that previous results on prosodic edge effects in laboratory speech (Fougeron and Keating, 1997, *inter alia*) can be extended to more natural conversational data. In addition, we found this lengthening after controlling for many contextual factors, including final pauses. This suggests that lengthening at prosodic edges plays a distinct role from prepausal lengthening. Initial strengthening is strongly associated with predictability from the previous word in ways whose understanding requires further research.

VII. CONCLUSIONS

Our results show that disfluencies, predictability, and utterance position all play strong and independent roles in whether a word is reduced, for all measures of reduction. While our regression study does not constitute a model in itself, these three results each have important implications for modeling of human lexical representation and production. First, a key result is that planning problems, as measured by disfluencies either preceding or following a function word, play a strong role in the word being longer and less reduced. This extends the results of Fox Tree and Clark (1997) on *the* to other function words. On the other hand, their suggestion that the basic form /ðɪ/ may signal a disfluency appears to be lexically specific, since we found increases in basic vowel frequencies in disfluent contexts only for *the*, *and*, and *it*. More crucially, the influence of planning problems is extended to duration, a nonphonological measure of reduction, which appears to hold generally for all the words examined.

Second, the result that function words are reduced when they are highly probable given neighboring words lends evidence to probabilistic models of human language processing (Jurafsky, 1996; Saffran *et al.*, 1996a; Seidenberg and MacDonald, 1999). While some of this reduction may be due to lexicalization of multiword phrases, some of it is due to the mental representation of some kind of probabilistic links between words, since the effects are not limited to frequent collocations. Previous work has focused on the role of probability in comprehension. Our work shows how probability can play a related role in production.

Our results on probability also extend the work of Griffin and Bock (1998), who showed that interactions between predictability and frequency argue for what they called cascade theories of word production, and against discrete two-stage models of word production. In discrete two-stage models (Jescheniak and Levelt, 1994; Levelt *et al.*, 1999), the predictability of a word in context can help cause a word to be selected. But word selection is simply binary; once a word is selected, the amount of contextual predictability does not play a role in phonological encoding. By contrast, cascade theories (Dell, 1986; Stemberger, 1985) allow the amount of evidence causing a word to be selected to be passed to lower levels in word production. Our results show that highly predictable words are shorter even after controlling for reduction or deletion at the phonological level. This suggests that the extent to which the context predicts a word cannot just play a role at lexical selection or during the compilation of syntactic and prosodic frames. Predictability (and probably also some disfluency effects) must also make its way down to the level of articulatory routines.

Third, our results show that utterance-initial and utterance-final words are longer and less likely to be reduced than utterance-medial words. Since the effect of utterance position was significant even after controlling for pauses, our results show that final lengthening in conversational speech is an attribute of the prosodic or syntactic boundary condition itself, and not of the correlated presence of pauses at boundaries. On the other hand, while final lengthening is a separate effect from any lengthening from lower predictabilities in final position, a parallel separation of position and predictability for utterance-initial position was not found. This raises the question of the proper interpretation of the interaction of predictability from neighboring words and phrasal edges, that is, whether they should be considered separate but strongly associated sources of form variation, or whether the typically low predictability of words at phrasal edges should be regarded as an intrinsic attribute of the position.

Most contextual effects in speech, like assimilation, are strongest next to their source. The factors studied here are no exception, all being local in nature, involving the immediately previous or following word or an immediately previous or following utterance boundary. This is partly because the strategy of looking for effects in the most likely circumstances dictated that such contexts be examined first. Even so, there was no additional advantage of considering predictability from the previous pair of words instead of just the previous word, and similarly for predictability from following words. (And while we did not analyze utterance-second position, the observed average duration of words in second position did not differ from those in the other medial positions.) This does not mean that there are no effects of the sort considered here that are more global in nature. One example is the shortening of repeated words in a discourse reported by Fowler and Housum (1987), although this is unlikely to be an important factor for very high-frequency words, since

they are repetitions most of the time. But the strength of the local effects, together with the suggestions that there may be at least a sharp drop in the influence of more distant factors, indicates that the local-global dimension of effects deserves closer attention, both for its contribution to the structure of production models as well as its significance for speech processing applications.

Our results also have some implications for lexical representation, suggesting that multiple lexical representations of high-frequency function words may be more numerous than models of speech production have usually assumed. For example, in addition to the more commonly noticed allomorphy of *the* and *a*, allomorphic models should also be considered for at least *to*, *of*, and *and*. Furthermore, the selection of these variants is sensitive to a wide range of factors, notably the activities of monitoring and repair. Integrating the effects of rate, style, segmental context, and prosodic context on the durations and forms of the word is also readily compatible with the models and concepts of gestural phonology (Browman and Goldstein, 1992).

Our results also have important implications for automatic speech recognition. Few of the factors that we show affect pronunciation variation are captured in current recognizers. Many of them could conceivably be added. Fosler-Lussier (1999a, 1999b) has shown first steps in this direction by showing how to build dynamic lexicons which are sensitive to speaking rate and the predictability of target words from previous words. These models could be extended to deal with predictability given following words. Similarly, planning problems could be handled with relatively simple modifications such as repetition detection and the use of a silence phone. The fact that there are key factors in reduction that are strictly local holds out the hope that good predictive models of word pronunciation may be based only on local information. We feel that these are promising directions for future investigations of ASR pronunciation models.

Much, of course, remains to be worked out in understanding the role of predictability in reduction. In addition to the exact locus of predictability in the cognitive processes involved in speech production, we still do not understand the complex interactions between conditional probabilities, joint probabilities, and item effects. Furthermore, we have simply reported first-order effects for probabilistic measures of local predictability, perhaps inviting the assumption that these effects are linear, holding in the same way from low to high probabilities. Even if this does not appear *a priori* unlikely to some, our own preliminary explorations of this question suggest that this simple model is not true. The more complex functional relationships between probability measures and reduction are yet to be determined.

If of course remains to be seen how general these effects are for all words in a conversation. In the general case, the relative frequency of each word, which we did not examine, plays a major role in the predictability of the word, and would be expected to influence word forms strongly. It may well interact with other measures, so that the effects found here might turn out not to be so strong for less frequent words. As a practical matter, disfluencies are disproportionately associated with function words, so that while we may find that longer words, less frequent words, and content words are longer and have less reduced forms in the presence of disfluencies, such occurrences may not be frequent enough to be of much practical importance for speech processing applications. Another difference that might be expected is that if the predictability effects found here are strongly associated with the connections of function words with particular constructions, then they may be weaker and less extensive for words that occur more freely.

In addition to these conclusions about lexical representation and production, we would like to end with a methodological insight. We hope to have shown that a corpus-based methodology such as ours can be paired with traditional controlled laboratory experiments to help provide insight into psychological processes like lexical production. Corpus-based methods have the advantage of ecological validity. The difficulty with corpus-based methods, of course, is that every possible confounding factor must be explicitly controlled in the statistical models. This requires time-consuming coding of data and extensive computational manipulations to make the data usable. Creating a very large hand-coded corpus is difficult, and there will always be factors that are beyond our ability to control for. But, to the extent that such control is possible, a corpus provides natural data whose frequencies and properties may be much closer to the natural task of language production than experimental materials can be. Obviously, it is important not to rely on any single method in studying human language; corpus-based study of lexical production is merely one tool in the psycholinguistic and phonetic arsenal, but one whose time, we feel, has come.

ACKNOWLEDGMENTS

This project was partially supported by the NSF, via awards IIS-9733067 and IIS-9978025. Many thanks to Joan Bybee, Steve Greenberg, Janet Pierrehumbert, Mari Ostendorf, Bill Raymond, Stefanie Shattuck-Hufnagel, Elizabeth Shriberg, and Caroline Smith, for many useful discussions on the issues raised in this article, and especially to Stefanie Shattuck-Hufnagel and an anonymous reviewer for extensive comments on an earlier draft. We are also very grateful to Stefanie Shattuck-Hufnagel and Mari Ostendorf for generously taking the time and effort to release to us a preliminary version of their prosodically coded portion of Switchboard.

[1]The choice of citation vowel is clear, even across American dialects, for all the words except *of*, which likely varies idiolectally between [ʌ] and [ɑ]. The vowel [ʌ] is arbitrarily taken to be the basic vowel of *of* here.

[2]It is noteworthy that this does not accord with Bolinger's lengthening rule, which predicts that full vowels are longer before full vowels, whether separated by consonants or not. Testing the effect of following full vowels for just items with full vowels also shows no overall effect on duration.

[3]Since the regional dialect area of Switchboard speakers was coded in our database, we checked the effect of this variable on our reduction indicators. No effect of dialect was found for duration, vowel reduction, or coda deletion. Only the frequency of basic vowels appeared to differ across dialects. We did not pursue effects of other factors, individual comparisons of dialects, or item effects.

[4]It is perhaps surprising that much the same difference between men's and women's speech rate is found for both read speech (i.e., Byrd's TIMIT result that men spoke 6.2% faster) and conversation (men's rate of 5.4 syllables/s is 8.0 versus women here). This may be in part due to the local measure (i.e., between pauses) used here. It is more like an articulation rate

measure than longer-term speaking rate measures, which would be strongly influenced by pause rate.

[5]We followed Fox Tree and Clark (1997) in choosing this definition of disfluency mainly for simplicity; deciding if a word was preceded or followed by a disfluency could be coded automatically by software, and required no subjective coding. There are problems with this simplified definition. Obviously not all instances of these disfluencies reflect planning problems. Some pauses and repetitions are fluently planned, and filled or silent pauses may initiate repair of previous speech, to mention just two alternatives. In addition, this definition means that we did not code for other disfluencies such as cutoffs and restarts, or for editing phrases such as *I mean*. Incomplete and imprecise as our disfluency set is, it is nevertheless an index of aspects of conversational structure that are strongly linked to reduction variation in word forms.

[6]The control variables used in the regressions were actually a subset of these, since not all were significant factors for each one of the response variables. The regressions exclude items at the beginning or end of fragments and thus cover samples approximately 10 percent smaller than those in Table IV. We chose not to control for utterance position in the regressions used to estimate the effects of disfluencies in this section because of the smaller sample it would entail. We did, however, verify that for the smaller sample coded for utterance position, the effect of disfluencies is much the same with or without utterance position control. See a further discussion of utterance position in Sec. VI.

[7]Tests for the difference are based on the comparison of a regression model that includes both variables with one with a single variable summing the two, forcing equal weighting of the two variables.

[8]Similar results were found for the 385 unaccented words (Sec. III C). Function words which were in the context of disfluencies were longer than those which were not. There was a highly significant effect both for preceding disfluencies $[F(1,368)=10.4,\ p<0.002]$ and for following disfluencies $[F(1,369)=26.9,\ p<0.0001]$. This indicates, at least, that the main result cannot be an artifact of the distribution of intonational accents.

[9]Figure 6 does not differentiate preceding disfluencies from following ones, because for most of the words, one is about as frequent as the other. The exceptions are *of*, which is 2.9 times more likely to occur before a disfluency than after one, and *I* and *you*, which are, respectively, 2.4 times and 4.4 times more likely to occur after a disfluency than before.

[10]This summary, which for simplicity's sake has collapsed preceding and following disfluencies, conceals some strong differences among the words in the relative strength of effects, depending on the direction. For example, reduced vowels of *a* are strongly affected by both preceding and following disfluencies, but those of *and* are much more affected by preceding disfluencies, and those of *the* much more affected by following ones.

[11]As with disfluencies, similar results were found for the 385 unaccented words (Sec. III C). Duration was affected by the five predictability variables, i.e., joint and conditional probabilities with the preceding word and with the following word, and the interaction between the preceding joint and conditional probabilities. The overall effect was highly significant $[F(5,369)=4.4,\ p<0.001]$. Preceding and following word probabilities were also individually significant, and as with the overall sample, function words that were more predictable from neighboring words were shorter.

[12]Both vowel reduction and obstruent deletion can of course have gradient sources in speech production, and the transcribers' categorical choices for the variables are unable to distinguish whether the source is from a lexical choice or gradient articulatory variation. Particularly for some of the function words, some reduction and deletion is likely to be lexical, for example [tu] versus [tə], [æn] versus [ən], [əv] versus [ə], and forms of *and* with and without final [d].

[13]The utterance-segmented subset of Switchboard is comparable to the larger ICSI sample in terms of proportions of individual function words, average rate, average duration, and proportion of reduced vowels. It contains slightly more disfluencies, and many more men speakers, however.

Agresti, A. (**1996**). *An Introduction to Categorical Data Analysis* (Wiley, New York).

Bolinger, D. (**1986**). *Intonation and its Parts: Melody in Spoken English* (Stanford University Press, Stanford).

Browman, C. P., and Goldstein, L. (**1992**). ''Articulatory phonology: An overview,'' Phonetica **49**, 155–180.

Bush, N. (**1999**). ''The predictive value of transitional probability for word-boundary palatalization in English,'' Master's thesis, University of New Mexico, Albuquerque, NM.

Byrd, D., Kaun, A., Narayanan, S., and Saltzman, E. (**2000**). ''Phrasal signatures in articulation,'' in *Papers in Laboratory Phonology V* (Cambridge University Press, Cambridge), pp. 70–87.

Byrd, D. (**1994**). ''Relations of sex and dialect to reduction,'' Speech Commun. **23**, 39–54.

Clark, H. H., and Wasow, T. (**1998**). ''Repeating words in spontaneous speech,'' Cogn. Psychol. **37**, 201–242.

Croft, W. (**1995**). ''Intonation units and grammatical structure,'' Linguistics **33**, 839–882.

Crystal, T. H., and House, A. S. (**1990**). ''Articulation rate and the duration of syllables and stress groups in connected speech,'' J. Acoust. Soc. Am. **88**, 101–112.

Dell, G. S. (**1986**). ''A spreading activation theory of retrieval in sentence production,'' Psychol. Rev. **93**, 283–321.

Fano, R. M. (**1961**). *Transmission of Information; A Statistical Theory of Communications* (MIT Press, Cambridge, MA).

Fidelholz, J. (**1975**). ''Word frequency and vowel reduction in English,'' in *CLS-75* (University of Chicago, Chicago), pp. 200–213.

Fosler-Lussier, E. (**1999a**). ''Contextual word and syllable pronunciation models,'' in Proceedings of the 1999 IEEE ASRU *Workshop*, Keystone, Colorado.

Fosler-Lussier, E. (**1999b**). ''Dynamic Pronunciation Models for Automatic Speech Recognition,'' Ph.D. thesis, University of California, Berkeley. Reprinted as ICSI Technical Report TR-99-015.

Fosler-Lussier, E., and Morgan, N. (**1999**). ''Effects of speaking rate and word frequency on conversational pronunciations,'' Speech Commun. **29**, 137–158.

Fougeron, C., and Keating, P. A. (**1997**). ''Articulatory strengthening at edges of prosodic domains,'' J. Acoust. Soc. Am. **101**, 3728–3740.

Fowler, C. A., and Housum, J. (**1987**). ''Talkers' signaling of new and old words in speech and listeners' perception and use of the distinction,'' J. Memory Lang. **26**, 489–504.

Fox Tree, J. E., and Clark, H. H. (**1997**). ''Pronouncing 'the' as 'thee' to signal problems in speaking,'' Cognition **62**, 151–167.

Godfrey, J., Holliman, E., and McDaniel, J. (**1992**). ''SWITCHBOARD: Telephone speech corpus for research and development,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing (IEEE ICASSP-92)* (IEEE, San Francisco), pp. 517–520.

Greenberg, S. (**1997**). ''Switchboard transcription system,'' unpublished manuscript labelers' manual, revision of 19 February, 1997.

Greenberg, S., Ellis, D., and Hollenback, J. (**1996**). ''Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus,'' in Proceedings of the International Conference on Spoken Language Processing (ICSLP-96), Philadelphia, PA, pp. S24–27.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (**1999**). ''The effects of collocational strength and contextual predictability in lexical production,'' in *CLS-99* (University of Chicago, Chicago), pp. 151–166.

Griffin, Z. M., and Bock, K. (**1998**). ''Constraint, word frequency, and the relationship between lexical processing levels in spoken word production,'' J. Memory Lang. **38**, 313–338.

Hock, H. H. (**1986**). *Principles of Historical Linguistics* (Mouton, The Hague).

Jescheniak, J. D., and Levelt, W. J. M. (**1994**). ''Word frequency effects in speech production: Retrieval of syntactic information and of phonological form,'' J. Exp. Psychol. Learn Mem. Cogn **20**, 824–843.

Jespersen, O. (**1922**). *Language* (Holt, New York).

Jurafsky, D. (**1996**). ''A probabilistic model of lexical and syntactic access and disambiguation,'' Cogn. Sci. **20**, 137–194.

Jurafsky, D., Bell, A., and Girand, C. (**2002**). ''The role of the lemma in form variation,'' in *Papers in Laboratory Phonology 7*, edited by N. Warner and C. Gussenhoven (Mouton de Gruyter, Berlin/New York) pp. 3–34.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (**2001**). ''Probabilistic relations between words: Evidence from reduction in lexical production,'' in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (Benjamins, Amsterdam), pp. 229–254.

Jurafsky, D., and Martin, J. H. (**2000**). *Speech and Language Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Keating, P. A., Byrd, D., Flemming, E., and Todaka, Y. (**1994**). ''Phonetic analysis of word and segment variation using the TIMIT corpus of American English,'' Speech Commun. **14**, 131–142.

Klatt, D. H. (**1975**). ''Vowel lengthening is syntactically determined in a connected discourse,'' J. Phonetics **3**, 129–140.

Krug, M. (**1998**). ''String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change,'' J. Engl. Linguistics **26**, 286–320.
Ladd, D. R., and Campbell, N. (**1991**). ''Theories of prosodic structure: Evidence from syllable duration,'' in Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en-Provence, France, pp. 290–293.
Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (**1999**). ''A theory of lexical access in speech production,'' Behav. Brain Sci. **22(1)**, 1–75.
Lieberman, P. (**1963**). ''Some effects of the semantic and grammatical context on the production and perception of speech,'' Lang Speech **6**, 172–175.
MacDonald, M. C. (**1993**). ''The interaction of lexical and syntactic ambiguity,'' J. Memory Lang. **32**, 692–715.
Manning, C. D., and Schütze, H. (**1999**). *Foundations of Statistical Natural Language Processing* (MIT Press, Cambridge, MA).
Marcus, M. P., Santorini, B., Marcinkiewicz, M. A., and Taylor, A. (**1999**). *Treebank-3*, Linguistic Data Consortium (LDC). Catalog #LDC99T42.
McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (**1998**). ''Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension,'' J. Memory Lang. **38**, 283–312.
Meteer, M. *et al.* (**1995**). *Dysfluency Annotation Stylebook for the Switchboard Corpus*, Linguistic Data Consortium. Revised June 1995 by Ann Taylor. ftp://ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFL_book.ps.gz
Neu, H. (**1980**). ''Ranking of constraints on /t,d/ deletion in American English: A statistical analysis,'' in *Locating Language in Time and Space*, edited by W. Labov (Academic, New York), pp. 37–54.
O'Shaughnessy, D. (**1992**). ''Automatic recognition of hesitations in spontaneous speech,'' in Proceedings of the IEEE International Conference on Acoustics, Speech, & Signal Processing (IEEE ICASSP-92), Vol. I, pp. 593–596.
Plauché, M., and Shriberg, E. (**1999**). ''Data-driven subclassification of disfluent repetitions based on prosodic features,'' in Proceedings of the International Congress of Phonetic Sciences (ICPhS-99), San Francisco, Vol. 2, pp. 1513–1516.
Rhodes, R. A. (**1992**). ''Flapping in American English,'' in *Proceedings of the 7th International Phonology Meeting*, edited by W. U. Dressler, M. Prinzhorn, and J. Rennison (Rosenberg and Sellier, Turin), pp. 217–232.
Rhodes, R. A. (**1996**). ''English reduced vowels and the nature of natural processes,'' in *Natural Phonology: The State of the Art*, edited by B. Hurch and R. A. Rhodes (Mouton de Gruyter, The Hague), pp. 239–259.
Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996a**). ''Statistical learning by 8-month-old infants,'' Science **274**, 1926–1928.
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996b**). ''Statistical cues in language acquisition: Word segmentation by infants,'' in COGSCI-96, pp. 376–380.
Schiffrin, D. (**1987**). *Discourse Markers* (Cambridge University Press, Cambridge).
Schuchardt, H. (**1985**). *Über die Lautgesetze: Gegen die Junggrammatiker*. Robert Oppenheim, Berlin. Excerpted with English translation in *Schuchardt, the Neogrammarians, and the Transformational Theory of Phonological Change*, edited by T. Vennemann and T. H. Wilbur (Athenaum, Frankfurt, 1972) pp. 39–72.
Seidenberg, M. S., and MacDonald, M. C. (**1999**). ''A probabilistic constraints approach to language acquisition and processing,'' Cogn. Sci. **23**, 569–588.
Shattuck-Hufnagel, S., and Ostendorf, M. (**1999**). POSH labeling guide—version 1.0. Unpublished draft.
Shriberg, E. (**1994**). ''Preliminaries to a Theory of Speech Disfluencies,'' Ph.D. thesis, University of California, Berkeley, CA. (unpublished).
Shriberg, E. (**1995**). ''Acoustic properties of disfluent repetitions,'' in Proceedings of the International Congress of Phonetic Sciences (ICPhS-95), Stockholm, Sweden, Vol. 4, pp. 384–387.
Shriberg, E. (**1999**). ''Phonetic consequences of speech disfluency,'' in Proceedings of the International Congress of Phonetic Sciences (ICPhS-99), San Francisco, Vol. I, pp. 619–622.
Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (**1992**). ''TOBI: a standard for labelling English prosody,'' in Proceedings of the International Conference on Spoken Language Processing (ICSLP-92), Vol. 2, pp. 867–870.
Stemberger, J. (**1985**). ''An interactive activation model of language production,'' in *Progress in the Psychology of Language*, edited by A. Ellis (Erlbaum, London), pp. 143–186.
Trueswell, J. C., and Tanenhaus, M. K. (**1994**). ''Toward a lexicalist framework for constraint-based syntactic ambiguity resolution,'' in *Perspectives on Sentence Processing*, edited by C. Clifton, Jr., L. Frazier, and K. Rayner (Erlbaum, Hillsdale, NJ), pp. 155–179.
Wald, B., and Shopen, T. (**1981**). ''A researcher's guide to the sociolinguistic variable (ING),'' in *Style and Variables in English*, edited by T. Shopen and J. M. Williams (Winthrop, Cambridge, MA), pp. 219–249.
Zipf, G. K. (**1929**). ''Relative frequency as a determinant of phonetic change,'' Harv. Studies Classi Philol. **15**, 1–95.

Accuracy and variability of acoustic measures of voicing onset[a)]

Alexander L. Francis,[b)] Valter Ciocca,[c)] and Jojo Man Ching Yu
Department of Speech and Hearing Sciences, University of Hong Kong

(Received 28 May 2002; revised 31 October 2002; accepted 18 November 2002)

Five commonly used methods for determining the onset of voicing of syllable-initial stop consonants were compared. The speech and glottal activity of 16 native speakers of Cantonese with normal voice quality were investigated during the production of consonant vowel (CV) syllables in Cantonese. Syllables consisted of the initial consonants /p^h/, /t^h/, /k^h/, /p/, /t/, and /k/ followed by the vowel /a/. All syllables had a high level tone, and were all real words in Cantonese. Measurements of voicing onset were made based on the onset of periodicity in the acoustic waveform, and on spectrographic measures of the onset of a voicing bar (f_0), the onset of the first formant (F1), second formant (F2), and third formant (F3). These measurements were then compared against the onset of glottal opening as determined by electroglottography. Both accuracy and variability of each measure were calculated. Results suggest that the presence of aspiration in a syllable decreased the accuracy and increased the variability of spectrogram-based measurements, but did not strongly affect measurements made from the acoustic waveform. Overall, the acoustic waveform provided the most accurate estimate of voicing onset; measurements made from the amplitude waveform were also the least variable of the five measures. These results can be explained as a consequence of differences in spectral tilt of the voicing source in breathy versus modal phonation. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1536169]

PACS numbers: 43.70.Jt [AL]

I. INTRODUCTION

The accurate determination of voicing onset from acoustic signals is important both theoretically and clinically. From a clinical perspective, the onset of voicing is often used in assessing developmental maturation of neuromotor coordination (DiSimoni, 1974; Eguchi and Hirsh, 1969; Zlatin and Koenigsknecht, 1976) and constitutes an important part of the assessment of the speech production of hearing impaired talkers (Monsen, 1976). Measurement of voicing onset is also used in therapeutic applications for stutters (Borden *et al.*, 1985). From a theoretical perspective, the voice onset time (VOT) of stop consonants often serves as a significant acoustic correlate of, and perceptual cue to, category differences between voiced and voiceless, and aspirated and unaspirated stop consonant categories (Abramson and Lisker, 1970; Klatt, 1975; Lisker, 1975, 1978; Lisker and Abramson, 1970, 1964). While the terms VOT and voicing onset are used throughout this paper in accordance with established practice, it is important to note that VOT is merely one of a large set of interrelated acoustic consequences of variation in the relative timing of glottal and oral gestures (Abramson, 1977). In principle any (or all) of these acoustic correlates of laryngeal timing could conceivably function as a perceptual cue in the right circumstances (e.g., the absence of other, more predictable or more salient cues). However, in this article we are primarily concerned with the technical requirements of accurately estimating the onset of laryngeal oscillation (''voicing'') from an acoustic signal, and do not intend to address questions of how listeners might *perceive* the onset of voicing except insofar as it may influence talkers' production patterns.

While voicing onset can be accurately determined by electroglottography (Fourcin and Abberton, 1971), there are cases in which electroglottography is not possible, for example when making field recordings at remote sites or when analyzing speech from recordings that have already been made without an accompanying record of glottal function. Thus, it is often necessary to be able to identify the onset of voicing on the basis of an acoustic analysis alone.

Currently, it appears that the most commonly used methods for measuring the onset of voicing are based on the onset of periodicity in the acoustic waveform, possibly supplemented by spectrographic analyses (Abramson, 1995) or direct measurements of airflow (e.g., Koenig, 2001). However, this issue may have been decided more on the basis of expedience than precision, and it is not clear from the existing literature whether (or why) waveform-based measures should be considered preferable to many of the other acoustic measures of voicing onset that have been proposed in the past. For example, Peterson and Lehiste (1960) identified the onset of voicing as the point at which stable striations first become visible in the frequency region of the first formant of a wide band spectrogram. In contrast, Klatt (1975) made his measurements of voicing onset at the onset of visible energy in higher formants on the grounds that voicing onset might not always be visible in the first formant region. Finally, Lisker and Abramson (1964) determined the onset of voicing according to the time of the first vertical striations visible in

[a)]Some of the material in this article was presented at the 9th meeting of the International Clinical Phonetics and Linguistic Association, May 4th, 2002, Hong Kong, China.

[b)]Present address: Audiology and Speech Sciences, Purdue University, Heavilon Hall, West Layfayette, IN 47907. Electronic mail: francisa@purdue.edu
[c)]Electronic mail: vciocca@hkusua.hku.hk

a wideband spectrogram, presumably irrespective of the frequency (or formant) at which they first appeared. In addition to these spectrographic measures, it is also possible to measure the onset of voicing as the onset of energy visible in the ''voicing bar''—the region of lowest frequency energy in a wide-band spectrogram corresponding to the fundamental frequency (f_0), typically found below the first formant (Kent and Read, 2002, p. 144). Finally, it is possible to measure voicing onset directly from the acoustic waveform itself, in terms of the onset of the first clearly periodic pattern in the acoustic signal (see, e.g., Lieberman and Blumstein, 1988, p. 216). Until now there does not seem to have been a rigorous attempt to compare the efficacy of these different techniques.

Given the wide range of possibilities for measuring the onset of voicing from an acoustic signal, why should it matter which acoustic landmark (e.g., f_0, F1, F2, F3, or onset of waveform periodicity) is used? One important issue is the relative accuracy of measurements made at each landmark, as there are obvious differences between the latency of voicing onset that each indicates. For example, voicing striations typically appear later in higher formants, though this effect may be mitigated for voiceless stops because of the reduction in amplitude of the first formant transition following the release of such stops (Liberman *et al.*, 1958). If all measures are made using the same acoustic benchmarks (e.g., F3), accurate comparison may still be possible across syllables. For example, if it is determined that measurements of voicing onset based on the onset of striations in the third formant (F3) lag behind those made from the onset of striations in the first formant (F1), it may still be possible to reliably compare measurements made from F3 across utterances or talkers. However, such comparison depends on the assumption that measurements made at all landmarks are equally variable. If measurements made at a particular landmark vary as a function of independent factors such as consonant aspiration or talker identity, then that landmark would be considered less useful. The ideal acoustic measurement of voicing onset is one that is both accurate and relatively consistent. Its latency must closely match the physiological onset of vocal-fold vibration, and must remain unaffected by factors unrelated to the physical initiation of vocal fold vibration.

The present study compared the accuracy and variability of five acoustic measures of voicing onset. Accuracy was measured in terms of the mean asynchrony between measurements made at each landmark and the time of voicing onset determined by electroglottography. Variability was measured in terms of the variance of the mean asynchrony for each landmark.

II. METHOD

A. Subjects

Sixteen native speakers of Cantonese (eight men, eight women) with normal voice quality (as judged by unanimous agreement of three final year speech therapy students) and no reported history of speaking or hearing disability participated in this study. Their age ranged from 20 to 25 years (mean $=22.06$ years).

B. Stimuli

Stimuli consisted of six monosyllabic real words in Cantonese, all produced with a high level tone. All words had a consonant–vowel (CV) syllable structure with the vowel /a/ approximately as in the English word ''father.'' The words were /pa/ (father), /p^ha/ (on all fours), /ta/ (dozen), /t^ha/ (he), /ka/ (home), and /k^ha/ (compartment). Note that Cantonese is traditionally described as maintaining a phonological contrast between voiceless unaspirated and voiceless aspirated stop series at three places of articulation. Perceptually, aspiration cues seem to be stronger indictors of this phonological contrast than are timing (VOT) cues (Tsui and Ciocca, 2000). However, it may be argued that both timing- and aspiration-related acoustic patterns reflect the same relative timing of oral and laryngeal gestures (Abramson, 1977; Davis, 1994). Thus, regardless of what listeners are listening for (aspiration noise, VOT, or some combination), here we are primarily concerned with how best to use the acoustic signal to infer the relative timing of two articulatory events—events that together result in *both* the presence or absence of aspiration noise *and* longer or shorter VOT.

C. Procedure

Stimuli were recorded in a sound-shielded room using a low noise omnidirectional microphone (Shure Beta 87) with a Bruel and Kjær model 2812 MKII preamplifier and a Kay Elemetrics model 6094 Laryngograph connected to a TASCAM DA-30 MKII DAT tape recorder. The laryngographic signal was recorded to the left channel and the acoustic signal was recorded to the right channel. During recording, the microphone was mounted on a boom and situated approximately 10 cm in front of the talkers' lips. The laryngograph electrodes were held in place with a Velcro strap and were located slightly above and to either side of the talkers' thyroid cartilage. Each word was presented individually to the talker by means of file cards (one word, written in Chinese characters, per card). For each presentation of each word, the talker was asked to read the word aloud three times in a normal voice and at a comfortable rate with a preceding vowel /a/. For example, for the word /pa/ ''father'' the talker read [apa apa apa]. Each of the six words was presented a total of five times in randomized order. Only the middle instance of the three target syllables in each utterance was analyzed, in order to minimize any effect of the initiation or anticipation of the end of the utterance. That is, all measurements were taken from intervocalic stops at the onset of a stressed syllable. For each of the six words there were five syllables from which measurements of voicing onset were made. Items that were heard to be mispronounced during the recording session were repeated at the end of the recording list. However, talkers misspoke or misread four of the stimuli without being noticed during recording, producing syllables with an incorrect place of articulation, and in two cases also an incorrect degree of aspiration. These four tokens were not included in the final analysis.

D. Data analysis

Speech samples were low-pass filtered at 22 kHz and digitally sampled at 44.1 kHz using GW Instruments' SoundScope 16 software on a Macintosh 7200/120AV via a Digidesign Audiomedia II sound card. Acoustic and laryngographic (Lx) signals were digitized simultaneously, and stored in separate time-locked files linked by name. Initial spectrographic measurements from the acoustic signal were made with GW Instruments' SoundScope 16 using a wide-band spectrogram display set to a 300-Hz analysis bandwidth and a frame advance of 0.1 ms with a time scale of 5 ms per division. Approximately 12.5 divisions were shown, or 62.5 ms at this scale. These values were selected as a compromise to achieve good temporal resolution while still retaining accurate formant resolution for both male and female talkers. In some cases other bandwidth parameters were used to confirm measurements made from the 300-Hz display. For determining onset of periodicity in the Lx and acoustic waveform displays, varying time-scales were used. An initial estimate was made from a display window encompassing at least five or six periods of the vowel, plus the entire duration of the consonant burst release (if any) and aspiration (if any). Subsequently, the temporal resolution of the window was increased until the zero-crossing preceding the first clearly periodic component of the acoustic waveform could be accurately identified. In the case of the Lx signal the onset of periodicity did not always correspond to a zero-crossing because the low-magnitude oscillations in voltage corresponding to the glottal opening and closing gestures were superimposed on larger fluctuations in current level of an unknown origin. In such cases, the onset of periodicity was determined to be the lowest point of the wave immediately preceding the first upward-going component of the first clearly periodic cycle in the Lx waveform.

Voicing onset was measured from the acoustic signal in terms of the time from the start of the sound file to one of five acoustic landmarks (f_0, F1, F2, F3, waveform), as shown in Fig. 1. Note that we did not measure VOT (voice onset time, defined as the difference between the time of the burst release and the onset of voicing). Rather, we measured the time of voicing onset in terms of the duration from the (arbitrary) start of the file because there is no accurate landmark in the Lx signal from which to identify the time of the burst release. For the f_0 measurement, the onset of voicing was determined as the time of the onset of coherent energy visible in the lowest-frequency region of the spectrogram. For the waveform measurement, the onset of voicing was identified as the time of the zero crossing preceding the upward-going portion of the first cycle of oscillation visible in the acoustic waveform. For the F1 measurement, the onset of voicing was identified as the time of onset of the first vertical striation visible in the frequency region of the first formant. For the F2 measurement, the onset of voicing was identified as the time of onset of the first vertical striation extending upward through the frequency regions of the first and second formants without interruption. For the F3 measurement, the onset of voicing was identified as the time of onset of the first vertical striation extending upward through the frequency regions of the first, second, and third formants

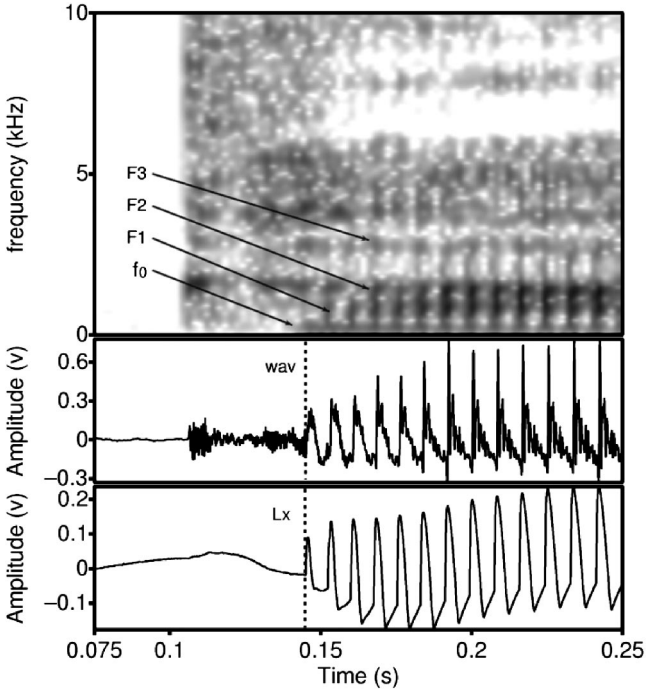


FIG. 1. Example of measurement locations in a representative syllable with an aspirated stop (male talker) in the spectrogram (top panel), acoustic waveform (middle) and Lx waveform (bottom). Measurement locations are indicated for six landmarks: fundamental frequency (f_0), first formant (F1), second formant (F2) and third formant (F3), the first upward-going zero-crossing preceding the first periodic wave component of the acoustic waveform (wav), and the lowest point preceding the first periodic wave component of the laryngographic waveform (Lx).

without interruption. These five acoustic measures of voicing onset were compared to the onset of regular vocal fold oscillation shown in the Lx waveform, defined as the lowest point immediately preceding the first cycle of regular oscillation, and calculated in terms of the measured time from the start of the sound file.

The asynchrony of each acoustic landmark was calculated as the difference between the acoustically determined time of voicing onset and that calculated from the Lx waveform, in milliseconds. Thus, an onset of voicing identified at 120 ms from the beginning of the sound file according to the F2 landmark, when compared with an onset of voicing at 110 ms from the beginning of the laryngographic waveform file according to the Lx waveform, would result in an asynchrony value of $+10$ ms. This measure of asynchrony was used as an estimate of the accuracy of measurements made at each acoustic landmark.

While accuracy is important for many purposes, in other cases it may not matter whether an acoustic measure of voicing onset is accurate (in the sense of being close in time to the actual onset of vocal fold vibration) as long as the values measured at that landmark remain relatively constant across contexts. For example, if all measures of voicing onset using a particular acoustic landmark are consistently 10 ms greater than those derived from the Lx waveform, then it should still be possible to compare measurements across different talkers, consonants, or other contextual variables since all mea-

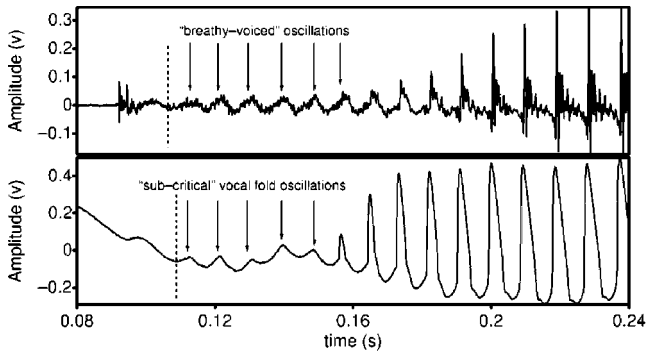


FIG. 2. Example of acoustic waveform (top) and Lx waveform (bottom) showing subcritical oscillations in vocal fold adduction visible in the Lx waveform corresponding to ''breathy-voicing'' periodicity in the acoustic signal. Measured point of onset of periodicity in the acoustic waveform is marked with a vertical cursor in the upper panel, and the vertical cursor in the lower panel indicates the measured location of the onset of oscillatory motion in the Lx signal.

sures will be biased to the same degree. However, if the asynchrony of measurements made at a particular landmark is highly variable, especially if it changes in a context dependent manner, then measurements made at that landmark cannot be considered reliable. Therefore, to determine the variability of measurements made at each landmark, we also calculated the variance of the asynchrony measured at each acoustic landmark for each syllable and each talker.

It should be noted that a large number of aspirated consonants exhibited a pattern in the Lx signal that suggested the presence of breathy voicing—subcritical oscillations of the vocal folds for around six or seven periods prior to the onset of clearly identifiable glottal opening/closing gestures, probably corresponding to the ''edge vibrations'' identified in spectrograms by Lisker and Abramson (1964, 1970). These patterns are similar to the glottal waveform of breathy vowels shown by Klatt and Klatt (1990, p. 822) (based on Stevens, 1977), and in the /aha/ productions of some of the talkers described by Löfqvist *et al.* (1995, p. 63). In the present study such patterns (illustrated in Fig. 2) were very common among female talkers, but every talker showed at least one production with such partial or breathy voicing (note that Fig. 2 shows the speech of a male talker). In all such tokens, voicing onset was estimated at the beginning of the upward-going curve of the first period of clear oscillatory motion in the Lx signal (shown with a vertical cursor in the lower panel of Fig. 2). Many utterances also showed a fluctuation in the Lx waveform coincident with the burst release in the acoustic waveform. In some cases (e.g., Fig. 1), this oscillation had a wavelength clearly much greater than that of typical voicing for the talker and was ignored (such long, solitary fluctuations may represent the influence on electrical impedance of extralaryngeal muscles involved in other aspects of speech production). In other cases, these fluctuations in the Lx waveform appeared much more like subcritical (breathy) voicing and were treated as subsequent to the onset of voicing. Finally, we note that the consonants under investigation, while syllable-initial, were not utterance-initial. As a result, it is possible that talkers continued the voicing of the preceding vowel through the stop consonant closure and into the subsequent vowel. While such articulations did occur occasionally, only productions in which there was a clear cessation of voicing preceding the burst release for the duration of at least approximately one period were included in these analyses.

III. RESULTS

A. Inter-rater reliability

In order to ensure that all measurements reported here reliably represent results that might be obtained by any experimenter, all of the waveform and Lx measures were redone by a second experimenter. In addition 20% of the tokens (one of the five repetitions of each of the consonants produced by each of the talkers) were randomly selected for remeasurement of the f_0, F1, F2, and F3 landmarks. All repeated measurements were made using the Praat 4.0 analysis software (Boersma and Weenink, 2001) with comparable spectrographic settings. Pearson's product-moment correlation analyses showed a high degree of inter-rater reliability, perhaps in part due to the strategy of using explicit, predetermined, acoustically defined measurement locations (e.g., the first visible vertical striation extending continuously through the first and second formant) facilitate inter-rater reliability. Pearson's correlation coefficients for each landmark were Lx, ($N=96$), $r=0.95$, $p<0.001$; waveform ($N=96$), $r=0.96$, $p<0.001$; f_0 ($N=96$), $r=0.97$, $p<0.001$; F1 ($N=96$), $r=0.97$, $p<0.001$; F2 ($N=96$), $r=0.96$, $p<0.001$; and F3 ($N=96$), $r=0.97$, $p<0.001$. All subsequent analyses used the measurements made by the second experimenter in those cases where measurements were repeated.

B. Accuracy

A four-way mixed factorial ANOVA (gender by aspiration by place of articulation by landmark) was calculated on the mean asynchrony values. Because of the very large differences in variance between cells in the design (for example, the mean variance for measurements of voicing onset made at the waveform landmark for [p] tokens produced by male speakers was 0.99 ms, while that for measurements made at the F3 landmark for [t^h] tokens was 428.02 ms), the Huynh–Feldt correction (Huynh and Feldt, 1970) was employed for repeated measures with more than two levels. Results showed no main effect of gender, $F(1,14)=0.059$, $p=0.81$, or of place of articulation, $F(1.48,20.72)=2.19$, $p=0.13$, but there were significant effects of landmark, $F(1.69,23.64)=57.05$, $p<0.001$, and aspiration, $F(1,14)=47.84$, $p<0.001$. There was also a significant two-way interaction between landmark and aspiration, $F(4,56)=61.54$, $p<0.00$, and a significant interaction between landmark, aspiration, and place of articulation, $F(8,112)=2.40$, $p=0.02$. None of the other interactions was significant at the $\alpha=0.05$ level. A graph of the three-way interactions (landmark by aspiration by place of articulation) is shown in Fig. 3. From this graph and posthoc analysis (Tukey HSD, $\alpha=0.05$), a number of observations can be made.

First, the different places of articulation pattern together at all levels of landmark and aspiration except that the /t^h/ tokens showed significantly greater offset ($p<0.05$) from

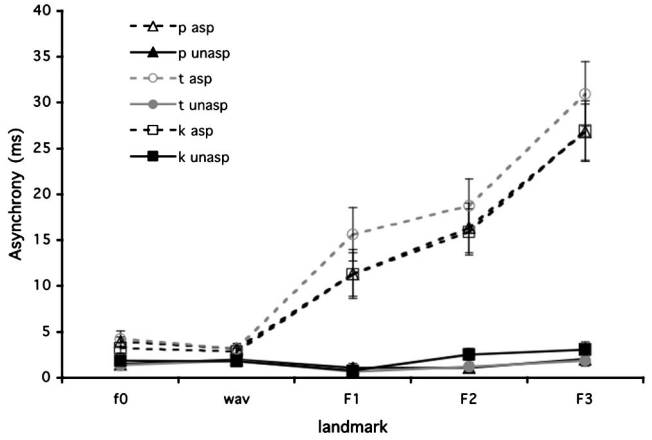


FIG. 3. Mean asynchrony between physiological (Lx) measures of voicing onset and measures of voicing onset made from the spectrographic display of the onset of energy at the fundamental frequency (f_0), first formant (F1), second formant (F2), third formant (F3), and from the first upward-going zero-crossing preceding the first periodic component of the acoustic waveform (wav). Measurements were made from syllables with aspirated stop consonants (open symbols and dotted lines) and unaspirated stop consonants (filled symbols and solid lines). Error bars indicate standard errors of the means.

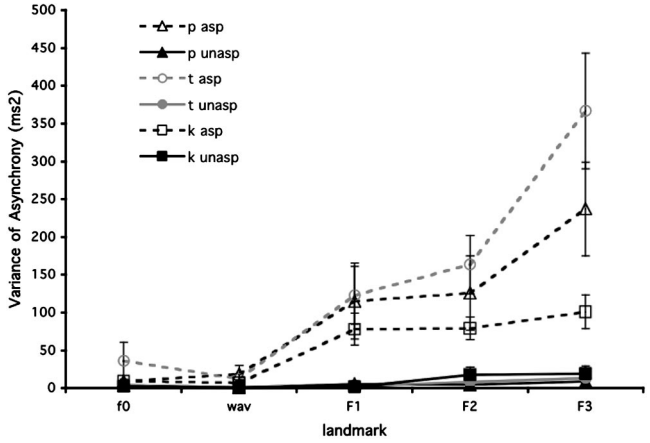


FIG. 4. Variance in mean asynchrony between physiological (Lx) measures of voicing onset and measures of voicing onset made from the spectrographic display of the onset of energy at the fundamental frequency (f_0), first formant (F1), second formant (F2), third formant (F3), and from the first upward-going zero-crossing preceding the first periodic component of the acoustic waveform (wav). Measurements were made from syllables with aspirated stop consonants (open symbols and dotted lines) and unaspirated stop consonants (filled symbols and solid lines). Error bars indicate standard errors of the means.

the /p^h/ and /k^h/ tokens at the F1 and F3 landmarks (only). These differences aside, there are no significant differences between places of articulation at the same levels of aspiration and landmark, and generalizing across place of articulation will not affect the overall validity of further conclusions. Collapsing across place of articulation, there was no significant difference between measurements of asynchrony of aspirated and unaspirated tokens made either using the f_0 or waveform landmarks ($p>0.05$). However, the asynchrony of aspirated tokens was significantly greater than that of unaspirated tokens when measured at any of the F1, F2, and F3 landmarks. These differences were due to an increased asynchrony for the aspirated tokens at the F1, F2, and F3 landmarks; in the unaspirated series there was no significant difference in voicing onset asynchrony between any of the landmarks ($p>0.05$ for all pairwise comparisons). For the aspirated series, the asynchrony of F1 measurements was significantly greater than that of measures made at f_0 or the waveform. Aspirated F1 asynchrony was also significantly smaller (more accurate) than F3, though there was no significant difference in asynchrony between the F1 and F2 measurements for aspirated consonants.

Note that the measurement criteria used for determining the location of the F2 and F3 landmarks meant that the asynchrony at these locations had to be equal to or greater than that measured from the F1 landmark because the F3 landmark could occur, by definition, no earlier than the F2, which in turn could be no earlier than the F1 landmark. However, as shown in Fig. 1, in the aspirated series the F2 and F3 landmarks could (and often did) appear at the same time, and indeed in the unaspirated series the three formant-based landmarks typically appeared at the same time. It is likely that this constraint on measurement lead to some degree of variability in measurements made from the formant-based landmarks. This is because it was often difficult to determine precisely whether the vertical striation of a particular glottal pulse extended without interruption through successive formants, even when the pulse was clearly evident within the frequency ranges of each formant. Still, the imposition of consistent display parameters may have reduced inter-rater variability due to this same uncertainty by encouraging both raters to use displays that were uncertain to a similar degree.

C. Variability

A four-way mixed factorial ANOVA (gender by aspiration by place of articulation by landmark) on the variance of the asynchrony measurements made at each landmark was also calculated. Results showed a main effect of landmark, $F(2.07,29.01)=25.56$ and aspiration, $F(1,14)=28.24$, $p<0.001$, but no effect of place of articulation, $F(2.00,28.00)=2.64$, $p=0.09$, or talker gender, $F(1,14)=0.06$, $p=0.80$. There were also significant two-way interactions between landmark and aspiration, $F(4,56)=22.42$, $p<0.001$, and landmark and place of articulation, $F(8,112)=3.79$, $p=0.001$, and a significant three-way interaction between landmark, aspiration and place of articulation, $F(8,112)=3.81$, $p=0.001$. None of the other interactions was statistically significant at the $\alpha=0.05$ level. Because talker gender did not play a role in any significant interaction, and was itself not a significant factor in the analysis, it was excluded from further analyses. A graph of the three-way interaction (landmark by aspiration by place of articulation) is shown in Fig. 4. From this graph and post-hoc (Tukey HSD, $\alpha=0.05$) analysis, a number of observations can be made.

TABLE I. Pearson's product moment correlations between mean asynchrony and variance of asynchrony. Values significant at the $\alpha = 0.05$ level are shown in bold.

	Aspirated	Unaspirated
f_0	0.44	0.26
	$p = 0.085$	$p = 0.319$
Wav	**−0.57**	**−0.79**
	$p = 0.021$	$p = 0.001$
F1	**0.82**	0.04
	$p = 0.001$	$p = 0.880$
F2	**0.76**	**0.64**
	$p = 0.001$	$p = 0.007$
F3	**0.60**	**0.75**
	$p = 0.015$	$p = 0.001$

Similar to the accuracy measurements, there was no significant difference between different places of articulation at any of the five landmarks for the unaspirated consonant series, and there was no significant effect of landmark in the unaspirated series. By contrast, in the aspirated series there were some significant differences in variance at the same landmark depending on place of articulation, though only for the F3 landmark. For example, the /k^h/ tokens showed significantly less variance at the F3 landmark than the /p^h/ or /t^h/ tokens, and the /p^h/ tokens were in turn less variable than the /t^h/ ones. However, there were no significant differences due to place of articulation in the aspirated series at any other landmark. Because the overall trend of the effect of landmark is the same for all three places of articulation, the effects of landmark were investigated by pooling across place of articulation. There was no significant difference between the variability of the asynchrony of aspirated versus unaspirated consonants when measured at either the f_0 or waveform landmarks. However, at the F1, F2, and F3 landmarks the variance of measurements made from unaspirated consonants was significantly smaller than that of measurements made from aspirated consonants. Finally, for the aspirated consonants the variance of the F1 measurements was not significantly different from the F2 measure, but both were significantly smaller than F3.

D. Relationship between asynchrony and variability

It may be noted that the distribution of variances shown in Fig. 4 appears quite similar to that of the mean scores shown in Fig. 3. Indeed, an analysis of the product moment correlations between the means and variances of each of the landmarks shows strong correlations between most of the asynchrony measurement means and their corresponding variances, as shown in Table I. These findings are consistent with the observation that variance typically increases with increasing means, at least in measurements of durations associated with speech production (e.g., Ohala, 1975; Smith *et al.*, 1983; Smith, 1992, 1994).

Because of the correlation between means and variances in this case, measurements of variance alone cannot be used to determine whether measurements at certain landmarks are *intrinsically* more variable. Although these results show that measurements made at F3 do have clearly higher variance, we cannot tell whether this is simply due to the general tendency of variability to increase with large means, or whether it is truly the case that F3 measurements are always more variable than F2 measurements. This issue is further complicated by the observation, described above, that mean measurements made at higher frequency landmarks will always be equal to or greater than those made at lower frequency landmarks. Note that the use of a measure that normalizes variances according to their corresponding means, such as the coefficient of variation (CoV, defined as the standard deviation divided by the mean), is not ideal for the present data set because many of the means involved are equal to or very close to zero. This is problematic for two reasons. First, as means approach zero the coefficient of variation approaches infinity, but there are too many such means to simply ignore these cells in the design. Similarly, because the means vary around zero (both above and below), the CoV can misrepresent equivalent differences in variance. Very small differences in means near zero result in highly divergent CoV values, while the same relative difference in means farther away from zero may result in quite similar CoV values, even for the same variance. For example, given two cells with the same standard deviation of 1.5, but means of -0.1 versus $+0.1$, the CoV for the first will be -15, and for the second it will be 15. However, for means of 0.1 and 0.3, each with a standard deviation of 1.5, the respective CoV values are 15 and 5 (one third as far apart). Thus, while we can say that measurements made at higher frequencies seem to have higher variability as well, we cannot say whether this is simply an inherent property of measurements with larger means, or whether formant-based measurements of voicing onset are necessarily more variable overall.

IV. DISCUSSION

The results of the present study suggest that the presence of aspiration after the release burst of stop consonants strongly affects the accuracy and variability of estimates of voicing onset made from acoustic measures. While the unaspirated stop series showed no significant differences in accuracy and variability of measurements of voicing onset across the five measurement landmarks, aspirated stops showed a consistently larger asynchrony and variance for measures made from the spectrogram at higher frequencies. These results imply that the most accurate acoustic measurements of voicing onset across categories can be made directly from the waveform using the onset of periodicity as an indicator of the onset of voicing or from the spectrogram using the onset of energy in the voicing bar.

Early research using acoustic measures of voicing onset typically relied on spectrographic measurements, perhaps because researchers could base their measurements on aspects of the signal known to be important in consonant perception. For example, Klatt (1975) suggested that measurements made according to the onset of energy in multiple (higher) formants may be more accurate than those made from the first formant alone because spectrographic displays (at that time) relied on display mechanisms (thermal printing) that were unable to adequately represent the subtle differences in energy over the small, low frequency ranges that indicate the onset of voicing in the voicing bar or first formant. However,

the development of digital spectrographic analyses may render moot such questions of visual precision, and the ability to employ high sampling rates (e.g., 44.1 kHz) makes it possible to identify patterns in the acoustic waveform with extremely high temporal precision as well. Certainly, the results presented here suggest that (*i*) measurements based on the waveform or voicing bar are generally more accurate and less variable than formant-based measurements, and (*ii*) F1-based measurements of voicing onset made from a digitally generated spectrogram are more accurate than F2- and F3-based measures.

The phenomenon of breathy voicing onsets observed in the productions of many aspirated consonants is also worthy of further discussion. Every talker in the present survey produced at least one token with a breathy voiced period between the burst release and the onset of modal voicing, and for some talkers, especially women, this pattern of glottal activity was the norm for aspirated stops. These results fit with the observations reported by Klatt and Klatt (1990), that female voices were typically perceived to be breathier than males. The Lx waveform displayed in Fig. 2 suggests that, as the vocal folds move from a fully abducted position to the nearly adducted configuration associated with modal voicing, they can pass through a phase of partial adduction. The Lx waveform does not directly measure the amount of vocal fold closure, but rather only indicates relative area of contact at the glottis (cf. Baken and Orlikoff, 2000, pp. 416–417). Therefore, from the Lx signal alone it cannot be determined whether this pattern of breathy voicing is accomplished by partially adducting the entire length of the vocal folds or by fully adducting a portion of the length of the vocal folds and maintaining an opening along the remaining length of the folds (cf. Hanson, 1997; Klatt and Klatt, 1990, p. 822). In either case, the resulting glottal configuration would support a voicing-like, periodic modulation of supra-glottal air pressure similar to the pattern indicated in Fig. 2, suggesting that this talker did achieve a degree of adduction sufficient to support periodic oscillation while maintaining sufficient abduction to allow a portion of the folds to produce breathy aspiration.

The prevalence of such breathy voicing following aspirated stop consonants suggests an articulatory account for certain acoustic observations made by Fischer-Jørgensen and Hutters (1981). As was found in the present experiment, Fischer-Jørgensen and Hutters (1981) found that, in open vowels, evidence for the onset of periodicity was often observed much earlier in lower-frequency regions of the spectrogram than in higher regions around the frequencies of the second and third formants. Acoustically, breathy phonation has been shown to exhibit greater spectral tilt than modal voicing, meaning that the amplitude of successive harmonics drops off more sharply as frequency increases in breathy phonation than in modal phonation, and higher harmonics are often replaced by aperiodic aspiration noise (Stevens, 1977). The fact that female talkers are more likely to be perceived as sounding breathy (Klatt and Klatt, 1990) has similarly been attributed to the greater rate of decrease in amplitude in successively higher harmonics in women's voices as compared with men's voices (cf. Holmberg *et al.*, 1988; Monson and Engebretson, 1977).

In the case of intervocalic stop consonants involving a cessation of voicing, a subsequent resumption of breathy voicing would mean that evidence for the onset of periodicity would not be found at higher frequencies until later in the utterance, when phonation has reverted to a modal pattern and the spectral tilt of the voicing source has once again become shallow enough to energize higher-frequency resonances of the vocal tract. Because of the complexity of the interrelationship between the articulatory gestures that control the transition from breathy to modal phonation, it seems plausible that this transition should be highly variable, both within and across talkers [see Löfqvist *et al.* (1995) for evidence of such variability]. In the present experiment, variability in estimates of voicing onset was indeed higher for measurements made at higher frequencies, especially following aspirated consonants. Because of the potential for a delayed appearance of voicing information at higher frequencies, and the concomitant increase in variability of measurements made at higher formants, the present results suggest that future investigations of voicing onset using acoustic measurements alone should be based on measurements made from the waveform or the voicing bar.

ACKNOWLEDGMENTS

This article is based on a dissertation submitted by the third author in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong. The authors would like to thank Arthur S. Abramson, Laura L. Koenig, and Anders Löfqvist for helpful comments on an earlier draft of this manuscript, and Raymond Wu, Donald Chan, and Kim-Ping Tsa for their technical assistance.

Abramson, A. S. (**1977**). "Laryngeal timing in consonant distinctions," Phonetica **34**, 295–303.

Abramson, A. S. (**1995**). "Laryngeal timing in Karen obstruents," in *Producing Speech: Contemporary Issues, for Katherine Safford Harris*, edited by F. Bell-Berti and L. J. Raphael (American Institute of Physics, New York), pp. 155–165.

Abramson, A. S., and Lisker, L. (**1970**). "Discriminability along the voicing continuum: Cross-language tests," in *Proceedings of the 6th International Congress of Phonetic Sciences*, pp. 569–573.

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice*, 2nd ed. (Singular, San Diego, CA), pp. 416–417.

Boersma, P., and Weenink, D. (**2001**). *Praat 4.0: A system for doing phonetics by computer* (computer software) (University of Amsterdam, Amsterdam, The Netherlands). Available online: http://www.praat.org

Borden, G. J., Baer, T., and Kenney, M. K. (**1985**). "Onset of voicing in stuttered and fluent utterances," J. Speech Hear. Res. **28**, 363–372.

Davis, K. (**1994**). "Stop voicing in Hindi," J. Phonetics **22**, 177–193.

DiSimoni, F. G. (**1974**). "Effect of vowel environment on the duration of consonants in the speech of three-, six-, and nine-year-old children," J. Acoust. Soc. Am. **55**, 360–361.

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Otolaryngol. (Stockh) **257**, 1–51.

Fischer-Jørgensen, E., and Hutters, B. (**1981**). "Aspirated stop consonants before low vowels. A problem of delimitation, its causes and consequences," Annual Report of the Institute of Phonetics, University of Copenhagen, Vol. 15.

Fourcin, A. J., and Abberton, E. (**1971**). "First applications of a new laryngograph," Medical and Biological Illustration, **21**, 172–182; reprinted in Volta Rev. **74**, 161–176.

Hanson, H. M. (**1997**). ''Glottal characteristics of female speakers: Acoustic correlates,'' J. Acoust. Soc. Am. **101**, 466–481.
Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). ''Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice,'' J. Acoust. Soc. Am. **84**, 511–529.
Huynh, H., and Feldt, L. S. (**1970**). ''Conditions under which mean square ratios in repeated measures designs have exact *F*-distributions,'' J. Am. Stat. Assoc. **65**, 1582–1589.
Kent, R. D., and Read, C. (**2002**). *The Acoustic Analysis of Speech*, 2nd ed. (Singular, San Diego, CA), p. 144.
Klatt, D. H. (**1975**). ''Voice onset time, frication, and aspiration in word-initial consonant clusters,'' J. Speech Hear. Res. **18**, 686–706.
Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.
Koenig, L. L. (**2001**). ''Distributional characteristics of VOT in children's voiceless aspirated stops and interpretation of developmental trends,'' J. Speech Lang. Hear. Res. **44**, 1058–1068.
Liberman, A. M., Delattre, P., and Cooper, F. S. (**1958**). ''Some cues for the distinction between voiced and voiceless stops in initial position,'' Lang Speech **1**, 153–167.
Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge U.P., New York), p. 216.
Lisker, L. (**1975**). ''Is it VOT or a first formant detector?'' J. Acoust. Soc. Am. **57**, 1547–1551.
Lisker, L. (**1978**). ''Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction,'' Status Report on Speech Research, SR-54 (Haskins Laboratories, New Haven, CT), pp. 127–132.
Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops,'' Word **20**, 384–422.
Lisker, L., and Abramson, A. S. (**1970**). ''Some effects of context on voice onset time in English stops,'' in *Proceedings of the 6th International Congress of Phonetic Sciences*, pp. 563–567.
Löfqvist, A., Koenig, L. L., and McGowan, R. S. (**1995**). ''Vocal tract aerodynamics in /aCa/ utterances: Measurements,'' Speech Commun. **16**, 49–66.
Monsen, R. B. (**1976**). ''Normal and reduced phonological space: The production of vowels by a deaf adolescent,'' J. Phonetics **4**, 189–198.
Monson, R. B., and Engebretson, A. M. (**1977**). ''Study of variations in the male and female glottal wave,'' J. Acoust. Soc. Am. **62**, 981–993.
Ohala, J. J. (**1975**). ''The temporal regulation of speech,'' in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. Tatham (Academic, New York), pp. 431–453.
Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllabic nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.
Smith, B. L. (**1992**). ''Relationships between duration and temporal variability in children's speech,'' J. Acoust. Soc. Am. **91**, 2165–2174.
Smith, B. L. (**1994**). ''Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adult's speech,'' J. Phonetics **22**, 155–175.
Smith, B. L., Sugarman, M. D., and Long, S. H. (**1983**). ''Experimental manipulation of speaking rate for studying temporal variability in children's speech,'' J. Acoust. Soc. Am. **74**, 744–749.
Stevens, K. N. (**1977**). ''Physics of larynx behavior and larynx modes,'' Phonetica **34**, 264–279.
Tsui, I. Y. H., and Ciocca, V. (**2000**). ''Perception of aspiration and place of articulation of Cantonese initial stops by normal and sensorineural hearing-impaired listeners,'' Int. J. Lang. Commun. Disord. **35**, 507–525.
Zlatin, M., and Koenigsknecht, R. (**1976**). ''Development of voicing contrast: A comparison of voice onset time in perception and production,'' J. Speech Hear. Res. **19**, 93–111.

Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training

Yue Wang[a)]
Department of Linguistics, University at Buffalo, State University of New York, Buffalo, New York 14260

Allard Jongman and Joan A. Sereno
Linguistics Department, University of Kansas, Lawrence, Kansas 66044

(Received 1 March 2002; accepted for publication 26 October 2002)

Training American listeners to perceive Mandarin tones has been shown to be effective, with trainees' identification improving by 21%. Improvement also generalized to new stimuli and new talkers, and was retained when tested six months after training [Y. Wang *et al.*, J. Acoust. Soc. Am. **106**, 3649–3658 (1999)]. The present study investigates whether the tone contrasts gained perceptually transferred to production. Before their perception pretest and after their post-test, the trainees were recorded producing a list of Mandarin words. Their productions were first judged by native Mandarin listeners in an identification task. Identification of trainees' post-test tone productions improved by 18% relative to their pretest productions, indicating significant tone production improvement after perceptual training. Acoustic analyses of the pre- and post-training productions further reveal the nature of the improvement, showing that post-training tone contours approximate native norms to a greater degree than pretraining tone contours. Furthermore, pitch height and pitch contour are not mastered in parallel, with the former being more resistant to improvement than the latter. These results are discussed in terms of the relationship between non-native tone perception and production as well as learning at the suprasegmental level. [DOI: 10.1121/1.1531176]

PACS numbers: 43.70.Kv [AL]

I. INTRODUCTION

Laboratory training of the discrimination of new phonetic contrasts is based on the assumption that the adult human perceptual system still has the capacity to change. Indeed, recent research has found that non-native segmental contrasts can be learned through auditory training. For example, there have been studies that trained American listeners with a three-way voice-onset time (VOT) distinction (e.g., Pisoni *et al.*, 1982; McClaskey, Pisoni, and Carrell, 1983), trained French listeners to identify the English /θ–ð/ contrast (e.g., Jamieson and Morosan, 1986, 1989), and trained Japanese listeners to identify English /r/ and /l/ (e.g., Logan, Lively, and Pisoni, 1991; Lively, Logan, and Pisoni, 1993; Lively *et al.*, 1994; Bradlow *et al.*, 1997). These studies have shown that the identification of non-native speech contrasts improved with training, and the improvement was extended to novel phonetic contexts and was retained long after training.

While these studies show the effect of training on segmental learning, Wang *et al.* (1999) extended the training procedure to the suprasegmental level by training American listeners to identify Mandarin Chinese tones. Mandarin phonemically distinguishes four tones, with tone 1 having high-level pitch, tone 2 high-rising pitch, tone 3 low-dipping pitch, and tone 4 high-falling pitch (Chao, 1948). Studies of the acoustic characteristics of Mandarin tones found that the differences in tones are manifested physically by different fundamental frequency ($F0$) values (Liu, 1924), with $F0$ height and $F0$ contour as the primary acoustic parameters characterizing Mandarin tones (Howie, 1976). For learners whose native language is nontonal, tone has presented great difficulty, since the functional association between these $F0$ characteristics and the segmental structure is unfamiliar to them (e.g., Kiriloff, 1969; Bluhme and Burr, 1971; Shen, 1989). Mandarin tone thus provides an ideal case for the study of suprasegmental training.

In Wang *et al.* (1999), eight American learners of Mandarin were trained in eight sessions during the course of 2 weeks to identify the four tones in natural words produced by native Mandarin talkers. Results show that, consistent with the previous segmental training studies, the perception of Mandarin tone improved significantly after training (21% improvement). Moreover, this improvement generalized to new stimuli (18% improvement) and new voices (25% improvement), and was retained when probed 6 months after training (21% improvement). These results are consistent with the previous findings at the segmental level, suggesting that training produces highly generalized perceptual learning that yields long-term modifications of the learners' perceptual system.

One subsequent question is whether perceptual training can affect production, so that training efforts could result in positive transfer from one modality to the other. The transfer of learning from perception to production has been reported in studies training learners to perceive non-native segmental contrasts. Rochet (1995) examined the transfer effect of perceptual training on production of French voice onset time (VOT) categories by native speakers of Mandarin Chinese.

a)Electronic mail: yuewang@buffalo.edu

Using an imitation task, productions of voiced and voiceless stops (labials, dentals, and velars) in a variety of vowel contexts were elicited both before and after perceptual training. Perceptual training involved synthetic stimuli consisting of a labial (/b/ or /p/) stop followed by a single vowel (/u/) context. An assessment of production gains following perceptual training was carried out by measuring VOT values in pretest and post-test productions. Mean VOT durations for initial stops did show improvement towards more French-like VOT values. Production accuracy was also assessed by native speaker judgments, with voiceless stops exhibiting more misidentifications in pretest productions compared to post-test productions. While perceptual training did transfer to production of voiceless stops in initial position, the improvement in production was not significant for voiced stops, and did not generalize to the production of stops in intervocalic position. Although Rochet suggests that lack of generalization may be the result of phonetic cues being actualized differently for stop consonants in initial versus intervocalic position, the training procedures may also be responsible. Using different methodological manipulations, Bradlow *et al*. (1997, 1999) also investigated the effects of perceptual training on production, by examining the production of the English /r–l/ contrast by Japanese learners. Similar to Rochet, perception and production data were gathered both at pretest and post-test, and native speakers were used to assess production gains. In the Bradlow *et al*. studies, however, perceptual training involved identification of naturally produced English /r/ and /l/ minimal word pairs using a high variability perceptual training procedure (Logan *et al.*, 1991). The results are consistent with Rochet, showing that native speakers identified the post-test productions more accurately than the pretest productions. Moreover, these production improvements were generalized to novel stimuli, and were retained 3 months after the perceptual training. However, the studies found no correlation between degree of learning in perception and production. That is, it is not the case that improvement in perception and production proceeded in parallel within individual learners (Bradlow *et al.*, 1997). Overall, these studies have shown that perceptual training has a facilitatory effect on the production domain, but the nature of the relationship between perception and production is still not clear.

Although nearly all training studies have focused on perception, Leather (1990) reports an initial attempt to examine the effect of production training on perception. This study examined a group of Dutch speakers who were trained to produce four Mandarin words (with the same syllable ''*yu*'') differing in tone. Leather found that these Dutch speakers were able to perceive the differences in tone without perceptual training. The author concluded that training in one modality tended to be sufficient to enable learners to perform in the other. However, since only one syllable was used in training as well as testing, the generalizability of the learning effect was not easily determined. We do not know if this type of training can produce long-term learning that can be extended across stimuli, voice, as well as speech modality.

In the present study, American trainees were recorded, both before and after perceptual training, producing a list of Mandarin words. Since the perceptual training of Mandarin tones has resulted in long-term perceptual improvement across stimuli and voice (Wang *et al.*, 1999), the goal of this study is to examine whether perceptual learning of this suprasegmental property can be transferred to the production domain. This study presents the results of the American learners' productions before and after training, first with an assessment by native Mandarin listeners in an identification task, and then with an acoustic analysis of pitch track comparisons. This is the first attempt to quantify the training effect with acoustic analysis, to capture the nature of the production improvements following perceptual training.

II. NATIVE LISTENER EVALUATION

A. Method

1. Participants

The participants were the 16 native speakers of American English in the perceptual study, with eight trainees who participated in the 2-week perceptual training program, and eight controls who did not receive training (Wang *et al.*, 1999). The participants were randomly selected from the students at Cornell University who had taken one or two semesters of Mandarin Chinese. None of them had ever lived in a Mandarin-speaking environment, and most of them had no experience with a tone language prior to learning Mandarin (except for four with limited Cantonese). All were paid for their participation. (For details of the characteristics of the participants, see Wang *et al.*, 1999.)

Eighty-two adult native speakers of Mandarin Chinese (18–35 years old) with no reported speech or hearing impairments participated voluntarily as judges. They were all raised and educated in Beijing, and were familiar with the *pinyin* system and the tonal diacritics which were used in this study.

2. Stimuli

The stimuli consisted of 80 real monosyllabic Mandarin words presented in isolation, 20 with each of the four tones. Half of these stimuli were used in the perceptual training (''old'' stimuli), while the other half did not appear in training (''new'' stimuli), in order to test the generalization of the production gains. The same stimuli were used for both pretest and post-test. Thus, each subject provided four sets of production stimuli: 40 old stimuli from pretest, 40 new stimuli from pretest, 40 old stimuli from post-test, and 40 new stimuli from post-test.

3. Procedure

Before the tone perception pretest, the trainees and the controls were tape recorded in a sound-insulated booth in the Cornell Phonetics Laboratory, using a cardioid microphone (Electrovoice RE20) and cassette recorder (Carver TD-1700). They were asked to read the list of 80 stimuli (blocked for old and new stimuli in the perceptual training) at a normal speaking rate, and were encouraged to repeat or correct whenever they felt necessary. The stimuli were presented on a sheet of paper in *pinyin* using the tonal diacritics familiar to the speakers. The speakers were recorded reading the same stimuli after their perception post-test 2 weeks later.

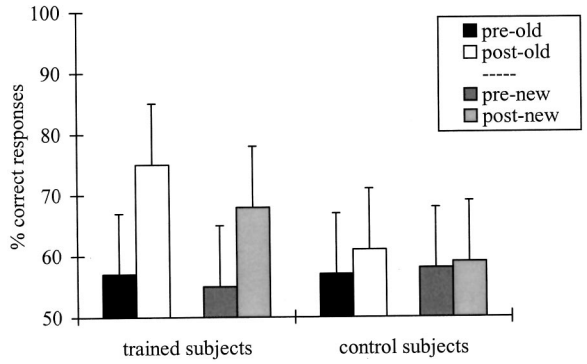


FIG. 1. Mean percent-correct identification of the four tone productions from the American trainees and the control subjects at pretest and post-test as judged by native Mandarin listeners. ''Pre-old'' and ''post-old'': pre- and post-test identification of the stimuli included in perceptual training; ''pre-new'' and ''post-new'': pre- and post-test identification of the stimuli not used in perceptual training.

The stimuli provided by the eight trainees and eight controls at both pretest and post-test were digitized at 11 kHz and low-pass filtered at 5 kHz using WAVES+/ESPS speech analysis software running on a SUN Sparc Station, after which they were edited and segmented.

These production stimuli were then transferred to a PC for play out in a perception experiment, using BLISS software (Mertus, 1989). The final output included four separate sets of data for each subject, i.e., 40 old and 40 new stimuli in perceptual training at both pretest and post-test, with a 3 s interstimulus interval. These stimuli were recorded to audio tape for evaluation by native Mandarin listeners in Beijing.

Prior to the perceptual judgment task, all potential Chinese judges were asked to identify a list of Mandarin words produced by a native Mandarin speaker. Only those who were able to identify all the tones correctly were included in the present study. Two listeners failed to reach this criterion and were excluded from participation.

The 16 American subjects' productions were evaluated by a total of 80 native Mandarin judges. Five different judges were asked to identify a single subject's productions at both pretest and post-test presented from a portable tape recorder. Answer sheets were provided containing the 40 stimuli in each of the four sets, written in *pinyin* with no tonal diacritics. Next to each stimulus, there were five categories, i.e., the four tonal diacritics, and a ''none'' category. Judges were to circle a tonal diacritic corresponding to the tone they heard, and to circle ''none'' if they decided what they heard did not correspond to any of the four tones. The order of presentation of the four stimulus sets for each of the 16 subjects was counterbalanced among judges. Thus, the identification task resulted in a total of 800 observations for each of the 16 speakers (5 judges$\times$40 stimuli$\times$4 sets).

B. Results

1. Overall improvement and generalization

Figure 1 shows the overall results of the production judgments. It displays the percentage of correctly identified productions by the American trainees and the control subjects as evaluated by the Mandarin judges. It should be noted that in the identification task, the judges could also categorize the trainees' productions as being none of the four tones. However, the results reveal that this category constitutes only a small proportion (2.7%) of the pre- and post-test judgments, indicating that for most of the cases, the trainees' productions were judged as one of the four Mandarin tones.

As shown in Fig. 1, the trainees showed an improvement in their production evaluation scores from 57% in the pretest to 75% in the post-test for the old stimuli, and from 55% to 68% for the new stimuli. This indicates the trainees' substantial improvement in production after perceptual training; that is, their improvement not only occurred on the stimuli used in perceptual training (18% increase), but was also generalized to new stimuli that were not included in the perceptual training (13% increase).

In contrast, although the control subjects started at approximately the same level as the trainees in the pretest (''old stimuli'': 57%, and ''new stimuli'': 58%), they exhibited little improvement in the post-test (''old stimuli'': 61%, and ''new stimuli'': 59%). The lack of substantial improvement for the controls occurred both for the stimuli included in the perceptual training (4% change) and also for the new stimuli (1% change).

A three-way repeated measures ANOVA was calculated with test (pretest, post-test) as within-subject factor and group (trained, control), and stimulus (old, new) as between-subject factor. The results revealed a significant main effect of test [$F(1,28)=49.3$, $p<0.0001$], and a significant test by group interaction [$F(1,28)=26.9$, $p<0.0001$]. The effects of group ([$F(1,28)=0.7$, $p>0.412$]) and stimulus ([$F(1,28)=0.19$, $p>0.663$]) did not reach significance. There was no interaction of test$\times$stimulus [$F(1,28)=2.0$, $p>0.168$], group$\times$stimulus [$F(1,28)=0.10$, $p>0.778$], or test$\times$stimulus$\times$group [$F(1,28)=0.14$, $p>0.712$]. These results suggest that the pretest and post-test performance was different for the trained and control subjects, and this difference occurred across the old and new stimuli in the perceptual training. More specifically, for the trained group, paired samples t tests showed a significant difference between pretest and post-test: [$t(15)=8.90$, $p<0.0001$]. In contrast, for the control group, no significant difference was observed for test [$t(15)=1.30$, $p>0.226$].

In sum, the above results show a significant improvement in production as the result of perception training for the trainees. Native Mandarin listeners more often perceived the intended tone after training as compared to before training. Moreover, this improvement in production was observed both for stimuli used in training and was extended to novel stimuli not included in perceptual training. However, the perceptual ratings of the native Mandarin listeners suggested no such improvement for the controls, judging controls' post-test productions as accurately as their pretest productions.

2. Individual trainees

Further analyses of these data examined productions for individual participants and individual tones. These subsequent analyses concentrated on the trainees and focused on the stimuli that were used in the perceptual training.

TABLE I. American trainees' percent-correct tone production as judged by Mandarin listeners.

Trainee	Pretest	Post-test	Increase
1	44	47	+3
2	24	55	+31
3	59	74	+15
4	69	88	+19
5	74	92	+18
6	77	90	+13
7	47	67	+20
8	61	83	+22
Mean	**57**	**75**	**+18**

Individual trainees' percent-correct tone production as judged by Mandarin listeners at pre- and post-test is presented in Table I, which shows that each trainee's production accuracy improved after perceptual training. Across all eight trainees, percent-correct tone production improved on average 18% from pretest to post-test. It is also noted that there is a large degree of variability among the eight trainees in terms of initial accuracy levels (24% to 77%), as well as amount of improvement (ranging from 3% to 31%).

3. Individual tones

The trainees' productions for each individual tone are illustrated in Fig. 2. Trainees' performance for each tone improved from the pretest to the post-test. A two-way ANOVA (test×tone) showed a main effect for test [$F(1,30) = 12.25$, $p < 0.002$], indicating significant improvement from the pre- to the post-test. There was also a main effect for tone [$F(3,28) = 7.45$, $p < 0.001$]. *Post hoc* analyses (Tukey HSD) showed that across pre- and post-test, tone 3 is significantly worse than tones 1, 2, and 4. The interaction of test and tone did not reach significance [$F(3,28) = 0.67$, $p > 0.577$], showing the improvement from pretest to post-test was consistent across all four tones.

4. Tone confusion

Table II presents a confusion matrix, for both the pre- and post-test, showing the number of production errors the trainees made for each tone as judged by native listeners of Mandarin. Tone confusions were examined, summing over the errors obtained for each tone pair. For example, the number of errors for tone pair 1 and 2 is the sum of errors of both tone 1 produced as tone 2, and tone 2 produced as tone 1. It can be seen that, at pretest, the most confusing tone pair was tones 2 and 3, followed by tones 2 and 4, tones 1 and 4, tones 1 and 2, tones 3 and 4, and tones 1 and 3. A similar rank order was retained at post-test (Spearman $\rho = 0.81$, $p < 0.05$), except for tones 1 and 4 which became the second most confusing pair.

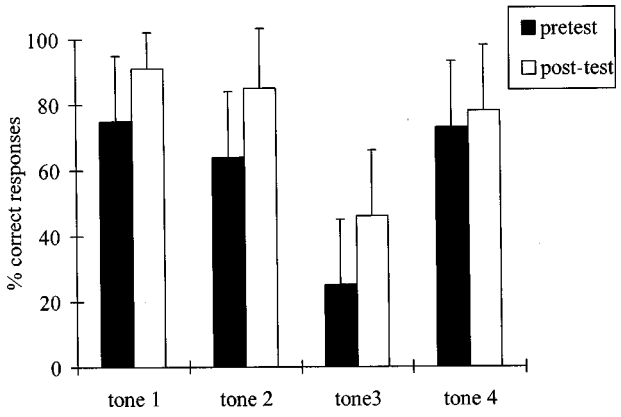


FIG. 2. American trainees' mean percent-correct productions for each tone at pretest and post-test as judged by native Mandarin listeners. The pitch contour shapes of the four tones are: tone 1, high-level pitch; tone 2, high-rising pitch; tone 3, low-dipping pitch; and tone 4, high-falling pitch.

TABLE II. Confusion matrices for the American trainees' tone productions (as judged by native Mandarin listeners) at (a) pretest and (b) post-test (10 stimuli×8 trainees×5 judges=400 responses for each tone). Correct responses are shown in bold.

	Stimulus			
	Tone 1	Tone 2	Tone 3	Tone 4
Produced as				
(a) pretest				
Tone 1	**300**	23	3	52
Tone 2	49	**255**	231	59
Tone 3	9	31	**101**	21
Tone 4	35	73	39	**262**
None	7	18	26	6
(b) post-test				
Tone 1	**365**	26	11	46
Tone 2	9	**340**	161	36
Tone 3	1	23	**184**	8
Tone 4	21	4	32	**305**
None	4	7	12	5

III. ACOUSTIC ANALYSIS

A. Method

To examine the American learners' productions before and after training, an acoustic analysis of the pitch tracks of the pre- and post-training productions was undertaken. Native speaker norms were derived and trainees' productions were compared to these native speaker norms for each of the Mandarin tones.

1. Participants and stimuli

To provide native norms for the four tones, four native speakers of Mandarin Chinese (2 males, 2 females) were asked to produce the same list of words as those used for the trainees. The setting and procedure were identical to those for the trainees.

The 40 stimuli that were included in the perceptual training were used in the acoustic analysis for both the Chinese and American speakers, yielding a total of 160 stimuli (10 syllables×4 tones×4 speakers) for the native speaker productions, and a total of 640 stimuli (10 syllables ×4 tones×8 trainees×2 tests) for the non-native productions.

2. Analysis

A total of 800 pitch contours (160 native productions, 320 non-native pretest productions, 320 non-native post-test productions) was derived using the WAVES+/ESPS software, at a sampling rate of 1 ms.

In order to directly compare the productions across speakers and stimuli, the pitch contours were normalized in two ways: $F0$ normalization, to accommodate the pitch range differences among speakers (especially between males and females), and duration normalization, to adjust for differences in speaking rate and syllable context. Both the native productions and the non-native productions were normalized in this manner.

$F0$ was normalized per speaker across the four tones. That is, the $F0$ values obtained from each speaker were converted to their logarithms, using a formula commonly adopted for such purposes (e.g., Liao, 1983; Ladd *et al.*, 1985; Rose, 1987; Shi, 1986, 1994)

$$T=[(\lg X-\lg \mathrm{L})/\lg \mathrm{H}-\lg \mathrm{L}]\times 5,$$

where H is the highest and L is the lowest $F0$ for a given speaker, and X is any given point of a pitch contour. The output (T) is a value ranging from 1 to 5, corresponding to the 5-point pitch scale for Mandarin tone proposed by Chao (1948).

Duration was normalized per tone across speakers. For each tone, the longest pitch contour was first determined (containing a certain number of $F0$ values depending on the length of the production at the sampling rate of 1 ms). Taking this number of $F0$ values, all other pitch contours for that tone were then time normalized by deriving the same number of $F0$ values, thus interpolating between observed $F0$ values. For example, as the longest pitch contour for tone 1 across speakers and tokens was 571 ms (571 $F0$ points), all tone 1 productions were ''stretched'' to have 571 $F0$ points at 1-ms intervals. The pitch contours of the other three tones were stretched in the same fashion, resulting in 569 $F0$ points for tone 2, 616 $F0$ points for tone 3, and 520 $F0$ points for tone 4. Thus, the duration of each tone was equalized across all speakers and tokens.

Using the converted $F0$ values (T values), the native norm for each tone was generated by averaging the four native Mandarin speakers' productions across all words. Likewise, for the non-native productions, two averaged productions were derived, one for the pretest and the other for the post-test by averaging across the eight trainees' productions.

For each contour, pitch values at 0% (onset), 25%, 50%, 75%, and 100% (offset), as well as the highest (peak) and the lowest (valley) points of the contour were calculated to record overall pitch height and shape. Furthermore, a number of critical attributes were analyzed on the basis of their acoustic characteristics and perceptual salience. First, pitch range (range) was calculated, since this feature has been identified as a perceptual cue distinguishing tones 1 and 2, and tones 1 and 4 (e.g., Leather, 1983; Lin and Wang, 1985; Fox and Qi, 1990). Second, both falling pitch range from onset to valley (falling range) and rising pitch range from valley to offset (rising range) were calculated. The falling pitch range, also termed ''$\Delta F0$,'' has been found to be critical in the identification of tones 2 and 3 (Shen and Lin, 1991; Moore and Jongman, 1997); whereas the rising pitch was found to cue the identification of tone 2 (e.g., Blicher *et al.*, 1990). Finally, the relative temporal position (position) from the onset to the valley of the pitch contour and from the onset to the peak of the pitch contour was calculated, since the duration from the onset to turning point (corresponding to the valley) has been found to be shorter for tone 2 than for tone 3 (Dreher and Lee, 1966; Moore and Jongman, 1997).

To eliminate occasional spurious values obtained from the pitch-tracking algorithm at the beginning and the end of the stimuli, the first and last 5 points for each pitch contour were excluded from the analysis.

B. Results

Figures 3(a)–(d) illustrate, for each tone, the pitch contours of the pretest and post-test productions averaged across trainees and stimuli, as compared to the native norm. The pitch contours are normalized for $F0$ and duration. Pitch values are represented on a 5-point pitch scale as T values.

As shown in the figure, for each tone, the post-test production resembles the native norm more closely than the pretest production both in terms of pitch height and contour shape, clearly showing an improvement in production resulting from perceptual training. An analysis of the critical points for each tone provides details about the improvement.

Table III displays the averaged T values at 0%, 25%, 50%, 75%, and 100% of the pitch contour, as well as the pitch range (range), for tone 1 at pretest and post-test across trainees, as compared to the native norm.

The native norm shows that, as a high-level tone, the T value remains constant (range: 0) with a relatively high pitch (4.2). The trainees' pretest production shows that the contour is relatively high and level, although the mean pitch values are a bit lower than the native norm and the pitch contour also decreases to a certain degree (range: 0.3). The pattern is mostly retained at post-test, except that the mean pitch values are even closer to the native norm.

A comparison of the pretest and post-test productions relative to the native norm was conducted using a two-way repeated measures ANOVA. The dependent variable is the ''deviation score'' (the absolute difference in T values between the native norm and either the pretest or the post-test production at a given point). The within-subject factor is test (pretest, post-test), and the between-subject factor is position (0%, 25%, 50%, 75%, 100%).

A significant effect of test was observed [$F(1,34)=15$, $p<0.0001$], with an overall deviation score (i.e., an averaged deviation score across all points in a contour) of 0.45 for the pretest and 0.35 for the post-test. There was no reliable effect of position [$F(4,34)=0.08$, $p>0.987$], nor was there a test by position interaction [$F(4,34)=0.3$, $p>0.904$]. These results show that across all 5 points in a contour, the difference between the post-test T value and the native norm (0.35) is significantly smaller than that between the pretest T value and the native norm (0.45), indicating greater degree of approximation to the native norm at post-test.

Overall, Table III, as well as Fig. 3(a) consistently reveal trainees' improvement at post-test for tone 1. It should also be noted that both pretest and post-test pitch contours are ''high'' and ''level,'' suggesting that the trainees' tone 1 production was relatively good at pretest, and further improved at post-test. This is consistent with the native speakers'

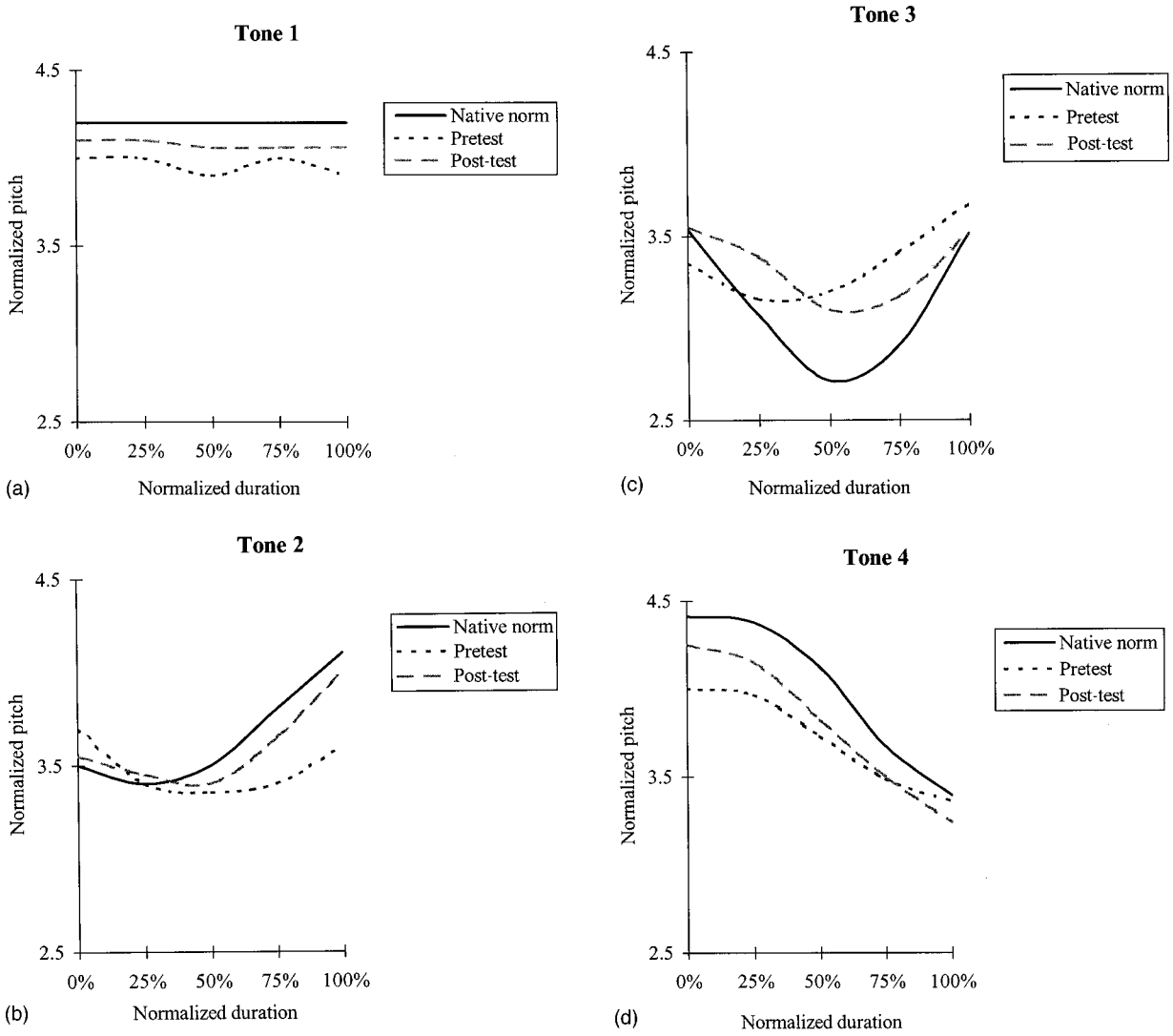


FIG. 3. (a)–(d) Pitch contours on the 5-point pitch scale (*T* values), comparing the native norm, and the pre- and post-test productions averaged across trainees, for tones 1–4, respectively. The pitch contour shapes of the four tones are: tone 1, high-level pitch; tone 2, high-rising pitch; tone 3, low-dipping pitch; and tone 4, high-falling pitch.

evaluation data, with 75% of the intended tone 1 productions judged by native speakers as correct tone 1 productions at pretest and this increased to 91% at post-test (cf. Fig. 2).

Table IV shows, for tone 2, the averaged *T* values at 0%, 25%, 50%, 75%, and 100% of the pitch contour, as well as the peak and valley *T* values and positions, and the pitch range values of the falling and rising portions of the contour, for the native norm and trainees' productions at pretest and post-test.

A two-way repeated measures ANOVA was conducted comparing the pretest and post-test productions relative to the native norm, with the deviation score as dependent variable, test as within-subject factor, and position as between-subject factor. A significant effect was observed for test [$F(1,34)=8.2$, $p<0.007$], but not for position [$F(4,34)=0.2$, $p>0.94$] or test by position interaction [$F(4,34)=1.3$, $p>0.284$]. The overall deviation values show that the difference between the pretest and native norm (deviation

TABLE III. Tone 1: Mean *T* values at 0%, 25%, 50%, 75%, and 100% (with standard deviations, s.d., in parentheses), as well as the peak (highest point), valley (lowest point), and range (pitch range), of the pitch contour of the native norm, and trainees' pre- and post-test productions.

	0% (s.d.)	25% (s.d.)	50% (s.d.)	75% (s.d.)	100% (s.d.)	Peak	Valley	Range
Native norm	4.2 (0.6)	4.2 (0.6)	4.2 (0.6)	4.2 (0.6)	4.2 (0.6)	4.2	4.2	0
Pretest	4.0 (0.6)	4.0 (0.6)	3.9 (0.6)	4.0 (0.6)	3.9 (0.6)	4.1	3.8	0.3
Post-test	4.1 (0.6)	4.1 (0.5)	4.0 (0.5)	4.0 (0.4)	4.0 (0.5)	4.1	3.9	0.2

TABLE IV. Tone 2: Mean *T* values at 0%, 25%, 50%, 75%, and 100% (with standard deviations, s.d., in parentheses) of the pitch contour, as well as mean *T* values at the lowest (valley) and highest (peak) points and their temporal position in % (expressed as a percentage of total time duration) (in brackets), and the falling pitch range (the difference between onset and valley) and the rising pitch range (the difference between valley and offset) of the pitch contour of the native norm, and trainees' pre- and post-test productions.

	0% (s.d.)	25% (s.d.)	50% (s.d.)	75% (s.d.)	100% (s.d.)	Valley [position]	Peak [position]	Falling range	Rising range
Native norm	3.5 (0.6)	3.4 (0.6)	3.5 (0.6)	3.8 (0.6)	4.1 (0.6)	3.4 [25%]	4.1 [100%]	0.1	0.7
Pretest	3.7 (0.5)	3.4 (0.5)	3.4 (0.6)	3.4 (0.6)	3.6 (0.7)	3.2 [65%]	3.7 [0%]	0.5	0.4
Post test	3.6 (0.5)	3.5 (0.4)	3.4 (0.4)	3.6 (0.4)	4.0 (0.4)	3.3 [47%]	4.0 [100%]	0.3	0.7

score: 0.50) is greater than that between the post-test and native norm (deviation score: 0.34), indicating significantly greater approximation to the native norm in the post-test than the pretest productions.

Detailed analysis of the critical points of the tone 2 native norm shows that, as a rising tone, the initial falling portion of the contour is relatively short (25% from onset) with minimal change in pitch range (0.1), while the rising portion is relatively long, reaching its peak (4.1) at the offset of the contour. Comparing the pre- and post-test test productions with the native norm, the post-test reveals greater approximation to the norm in a number of ways. First, the initial falling portion is shorter (47% versus 65%) and less steep (range: 0.3 versus 0.5) for the post-test than for the pretest. Second, similar to the native norm, in the post-test the major rising contour is long and reaches its peak at the offset (4.0). For the pretest, however, the peak (3.7) occurs at the onset, while the offset (3.6) does not rise as high as the native and post-test productions. These differences are clearly shown in Fig. 3(b).

Taken together, these results show for tone 2 a significantly higher degree of resemblance of the post-test production to the native norm both in terms of pitch contour and height, as compared to that of the pretest.

The learners' improvement is also reflected in the native speakers' evaluation described previously, as learners' correct tone 2 productions as judged by Mandarin speakers improved from 64% in the pretest to 85% in the post-test.

The average pitch values (*T*) for tone 3 at 0%, 25%, 50%, 75%, and 100% of the pitch contour, as well as the peak and valley *T* values and positions, and the pitch range values of the falling and rising portions of the contour, for the native norm and the trainees' productions are shown in Table V.

A two-way repeated measures ANOVA (test$\times$position) with deviation score as dependent variable was conducted. An analysis of the difference between the pretest and the post-test as compared to the native norm at each of these 5 points shows that the post-test production (deviation score: 0.46) is significantly more similar to the native norm than is the pretest production (deviation score: 0.60) [$F(1,34) = 7.6$, $p < 0.009$]. Neither position [$F(4,34) = 0.3$, $p > 0.85$] nor test by position interaction [$F(4,34) = 0.6$, $p > 0.654$] reached significance.

Detailed analysis of the critical points shows that, for the native norm, the turning point (valley: 2.4) is very low relative to both the peak at the onset (falling range: 1.2) and the offset (rising range: 1.1), and appears relatively late (55% from the onset) in the contour. In comparison, the turning point of the pretest production is not as low (3.2), and occurs in the initial portion (25%) of the contour. For the post-test production, although the turning point pitch value (3.1) is similar to that of the pretest, its position (65%) is much closer to the native norm, appearing relatively late in the contour as compared to that of the pretest. Moreover, the peak of the post-test production occurs in the same position as that of the native norm (0%), different from the pretest production peak position (100%).

Overall, the post-test production of tone 3 is significantly more similar to the native norm than the pretest production. Analyses show that the approximation is more in terms of pitch shape than pitch height [also cf. Fig. 3(c)], indicating that the post-test production has not fully reached the native norms. These results are supported by the native

TABLE V. Tone 3: Mean *T* values at 0%, 25%, 50%, 75%, and 100% (with standard deviations, s.d., in parentheses) of the pitch contour, as well as mean *T* values at the lowest (valley) and highest (peak) points and their temporal position in % (expressed as a percentage of total time duration) (in brackets), and the falling pitch range (the difference between onset and valley) and the rising pitch range (the difference between valley and offset) of the pitch contour of the native norm, and trainees' pre- and post-test productions.

	0% (s.d.)	25% (s.d.)	50% (s.d.)	75% (s.d.)	100% (s.d.)	Valley [position]	Peak [position]	Falling range	Rising range
Native norm	3.6 (0.6)	3.1 (0.4)	2.7 (0.4)	2.9 (0.5)	3.5 (0.3)	2.4 [55%]	3.6 [0%]	1.2	1.1
Pretest	3.4 (0.7)	3.2 (0.7)	3.2 (0.7)	3.4 (0.6)	3.7 (0.6)	3.2 [25%]	3.7 [100%]	0.2	0.5
Post-test	3.6 (0.6)	3.4 (0.5)	3.1 (0.6)	3.2 (0.5)	3.5 (0.4)	3.1 [65%]	3.6 [0%]	0.5	0.4

TABLE VI. Tone 4: Mean *T* values at 0%, 25%, 50%, 75%, and 100% (with standard deviations, s.d., in parentheses) of the pitch contour, as well as mean *T* values at the highest (peak) and lowest (valley) points, and their temporal position in % (expressed as a percentage of total time duration) (in brackets), and pitch range of the pitch contour of the native norm, and trainees' pre- and post-test productions.

	0% (s.d.)	25% (s.d.)	50% (s.d.)	75% (s.d.)	100% (s.d.)	Peak [position]	Valley [position]	Falling range
Native norm	4.4 (0.5)	4.3 (0.5)	4.1 (0.5)	3.7 (0.4)	3.4 (0.5)	4.4 [0%]	3.4 [100%]	1.0
Pretest	4.0 (0.6)	4.0 (0.6)	3.7 (0.5)	3.5 (0.5)	3.4 (0.6)	4.0 [0–25%]	3.4 [100%]	0.6
Post test	4.2 (0.6)	4.1 (0.6)	3.8 (0.6)	3.5 (0.4)	3.2 (0.5)	4.2 [0%]	3.2 [100%]	1.0

Chinese listeners' evaluation in that although significantly more post-test productions (46%) were judged to be correct than pretest productions (25%), the overall percent-correct productions is still relatively low in the post-test.

Table VI shows the pitch values for tone 4 at 0%, 25%, 50%, 75%, and 100% of the pitch contour, as well as the highest and lowest pitch values, for the native norm and the trainees' productions.

A two-way repeated measures ANOVA (test×position) with deviation score as dependent variable was conducted. An analysis of the difference between the pretest and the post-test as compared to the native norm at each of these 5 points shows that the post-test production (deviation score: 0.37) is significantly more similar to the native norm than the pretest production (deviation score: 0.47) [$F(1,34)=5.8$, $p<0.022$]. Furthermore, there is no significant effect of position [$F(1.34)=0.1$, $p>0.977$] nor a position by test interaction [$F(4,34)=0.5$, $p>0.709$].

As a high-falling tone, the native norm shows a high initial pitch (4.4) followed by a relatively steep fall (range: 1.0). Although both the pretest (4.0) and post-test (4.2) tone 4 productions start at a relatively high pitch, post-test values are higher than pretest values and consequently closer to the native norm. In terms of falling range, the post-test (range: 1.0) resembles the native norm more closely than the pretest (range: 0.6).

Overall, the post-test production of tone 4 is significantly more similar to the native norm than the pretest production. These results are also consistent with the native listener judgment data, showing that the percentage of productions that was identified as correct was greater in the post-test (78%) than in the pretest (73%). Both the acoustic analysis and the native listener evaluation data indicate that the pretest productions improved in the post-test.

The above results show that the native productions are consistent with the patterns described in Chao (1948) and Wu (1986) for the four Mandarin tones, showing a high-level pitch contour for tone 1, a mid-high rising contour for tone 2, a low-dipping contour for tone 3, and a high-falling contour for tone 4. The American learners' results show that, as a consequence of perceptual training, the post-test productions approximate the native productions more closely than do the pretest productions.

In terms of individual tones, the present results show that among the four tones, although the learners' pretest production of tone 1 was similar to the native tone 1 both in terms of pitch height and contour, tone 1 was even more native-like in the post-test. The ease of articulation for tone 1 might be attributed to the fact that, since the contour shape is constant, learners only need to grasp the single dimension of pitch height to correctly produce this tone. That the production of tone 1 is relatively easy as compared to the other three tones for American learners of Mandarin has also been reported in previous studies analyzing learners' productions both in laboratory (Leather, 1983) and classroom (Miracle, 1989) settings. It is also noted that, despite the closer approximation to a level contour, the post-test production still did not achieve full accuracy in terms of pitch height.

Tone 2 and tone 3 have consistently been found to be confusing in both first language acquisition (e.g., Li and Thompson, 1977; Clumeck, 1980) and second language acquisition (e.g., Leather, 1983; Miracle, 1989) studies. The confusion may well be due to the acoustic similarities of these two tones (Leather, 1990), in that they both involve a falling followed by a rising contour. However, differences exist in that the rising contour for tone 2 starts much earlier and ends much higher than that for tone 3. In addition, the valley for tone 2 is not as low as that for tone 3. In the present study, learners' pretest tone 2 rising contour started relatively late and did not reach high levels at the offset, showing more resemblance to a tone 3 pattern. However, in the post-test, both the frequency and temporal position of the valley, as well as the contour offset height closely approximated the native patterns.

Learners' pretest tone 3 productions reveal a rising contour starting relatively early just as a typical tone 2. Although in their post-test production the turning point position shifted later toward the native tone 3 direction, its frequency was still not as low as that of the native turning point. Together, these data show that tone 2 and tone 3 were confusable for the American learners.

The learners' pretest production of tone 4 is different from the native norm in two dimensions: it starts at a lower pitch, and its slope is less steep. In the post-test, the pitch range of the learners' productions was significantly increased, which resembled the native value. Although post-test pitch height was also increased, it was still lower than the native norm.

Taken together, the analysis of the productions of the four tones in pretest and post-test seems to suggest that the two dimensions of pitch height and pitch contour are not

TABLE VII. Tone pair confusion patterns for perception and production at pretest and post-test, in terms of percent errors for each tone pair.

	Pretest errors (%)		Post-test errors (%)	
	Perception	Production	Perception	Production
Tones 2&3	25	33	8.3	23
Tones 2&4	10	17	3.5	5.0
Tones 1&2	8.8	9.0	2.8	4.4
Tones 1&4	7.3	11	5.5	8.4
Tones 3&4	5.0	7.5	1.0	5.0
Tones 1&3	2.5	1.5	0	1.5

TABLE VIII. Confusion patterns for tone pair 2 and 3, and tone pair 3 and 4 at pretest and post-test, in terms of percent errors of the total number of stimuli, showing the difference between perception and production in terms of confusion direction.

	Pretest errors (%)		Post-test errors (%)	
Correct–incorrect	Perception	Production	Perception	Production
Tone 2 as tone 3	16	4	5.6	2
Tone 3 as tone 2	9	29	2.7	21
Tone 3 as tone 4	1.7	4.8	0	4
Tone 4 as tone 3	3.2	2.6	1	1

mastered in parallel. As compared to pitch contour, pitch height is more resistant to improvement.

IV. RELATION BETWEEN PRODUCTION AND PERCEPTION

Wang *et al.* (1999) showed that, after perceptual training of Mandarin tone, the American trainees' identification greatly improved (21% increase). The present study demonstrated that their production accuracy also increased significantly, indicating a relation between the perception and production of Mandarin tone. Consequently, American trainees' perception and production of Mandarin tone in the pretest and post-test are compared to examine the nature of this relationship.

The tone confusion data of the perception results (Wang *et al.*, 1999) show that, in the pretest, the most confusing tone pair was tones 2 and 3, followed by tones 2 and 4, tones 1 and 2, tones 1 and 4, tones 3 and 4, tones 1 and 3. This rank order was mostly retained in the post-test, except that tones 1 and 4 became the second most confusing pair. Interestingly, the present tone production confusion results reveal strikingly similar patterns in both pretest and post-test.

A comparison of the perception and production confusion patterns in the pretest and post-test is shown in Table VII, in terms of the percent errors for each tone pair. As shown in the table, in the pretest the percent errors for perception and production are highly correlated [$r(5) = 0.98$, $p < 0.0001$]. The rank order in terms of tone pair is also highly correlated for perception and production [$\rho(5) = 0.94$, $p < 0.005$]. Similarly, perception and production are significantly correlated in the post-test, both in terms of errors [$r(5) = 0.9$, $p < 0.01$] and in terms of tone pair rank order [$\rho(5) = 0.9$, $p < 0.015$].

These results show that trainees' tone perception and production are highly related. However, despite this general consistency, differences do exist between perception and production.

It is noted that although tone pair 2 and 3 is the most confusing pair for both perception and production, the direction of confusion is different for these two modalities. That is, tone 2 was incorrectly perceived as tone 3 more frequently than tone 3 was incorrectly perceived as tone 2. In contrast, tone 3 was incorrectly produced as tone 2 more frequently than the reverse. Similar patterns are also found for tones 3 and 4, in that tone 3 was more often incorrectly produced as tone 4, but less often perceived as tone 4. These patterns are illustrated in Table VIII.

Comparing the pretest and post-test data, it is also noted that tone pair 2 and 3 errors decreased to a large degree in perception. However, a similar decrease is not as evident in production comparing pretest to post-test. Similarly, tones 3 and 4 did not improve greatly in the production post-test.

These patterns are also reflected in the overall results for individual tones, in that the perception of tone 3 was relatively good to start with and significantly improved after training (see Fig. 2 in Wang *et al.*, 1999), whereas its production was poor in the pretest and remained so in the post-test (see Fig. 2). Taken together, these data show that while tone 3 was relatively easy to identify, it was difficult to produce, and was resistant to improvement.

V. DISCUSSION AND CONCLUSIONS

The present study shows that, after perceptual Mandarin tone training, the American learners' productions of Mandarin tone improved without any production training. The native Chinese listeners' evaluation of the trainees' pretest and post-test productions indicates that, after training, there was an improvement for each of the four tones, and the improvement in production was even extended to novel stimuli which were not used in the perception training. Native Chinese listeners' evaluation of the controls' pretest and post-test productions did not show a similar improvement. Acoustic analysis consistently revealed that the trainees' post-test productions were significantly more similar to the native norm in terms of both pitch height and contour than were pretest productions. These results indicate that the effect of training in perception transferred to the production domain.

The present study is consistent with previous training studies in the segmental domain showing the transfer of perceptual learning to production, such as the production of French VOT categories by native Chinese speakers (Rochet, 1995), and the production of the English /r–l/ contrast by Japanese learners (Bradlow *et al.*, 1997, 1999; Akahane-Yamada, 1999). Together, these studies coupled with the present results show that the facilitatory effect of perception training on production not only occurs for segmental learning, but also extends to suprasegmental learning.

The facilitatory effect of perception training for production learning supports the view in segmental acquisition research that the two speech modalities are related, with perception ''leading'' production (Flege, 1997). Indeed, the current phonetic learning theories are all perception oriented, stating that perceptual experience can guide sensory-motor learning (Kuhl, 2000a, b), and the accuracy with which L2

segments are perceived limits how accurately they can be produced (Flege, 1999). Studies of non-native segmental acquisition have found significant, albeit modest, correlations between learners' perception and production of L2 vowels (e.g., Flege, Bohn, and Jang, 1997) and consonants (e.g., Flege, 1993), showing that production accuracy is constrained by perception accuracy. The results in the present study provide supporting evidence of the nature of this relationship in suprasegmental learning. While the high correlation of the tone pair confusion patterns of the pretest perception and production shows the relationship of these two domains, the high correlation of the post-test perception and production tone pair confusion patterns clearly demonstrates how perceptual learning guided production. For example, the tone pairs that had been greatly improved perceptually, e.g., tones 2 and 3, tones 2 and 4, tones 1 and 2, also showed great improvement in production. In contrast, tone pair 1 and 4, which was most resistant to improvement in perception, had minimally improved in production as well.

Despite the general claim of a positive correlation between perception and production, the learning of these two modalities may not always be in parallel, as not all aspects of perceptual learning can be incorporated in production (Flege, 1999). Flege (1999) further pointed out that not all instances of non-native phonetic production have a perceptual origin. Some segments that are not used in learners' L1 phonetic system may present difficulty for production learning. The present results may also provide some support for this segment-based claim. Although, after training, tone 3 became relatively easy to perceive, it remained difficult to produce. Table VIII further revealed that, contrary to the patterns for perception, more tone 3 stimuli were incorrectly produced as tone 2 or tone 4. On this account, the difficulty in the production of tone 3 might not be due to a failure in perception learning, but rather to the novelty of the sound itself. It might be that the low dipping nature of the tone 3 pitch contour is so unfamiliar to the American learners that it makes articulation difficult. It is not known whether additional perceptual (or production) training may improve this particular tone.

The relationship of perception and production learning suggests an integrated system underlying these two mechanisms. Thus, learning a speech contrast involves mastery of both perception and production. The present finding that the effect of perceptual training not only extended to new speech contexts but was also transferred to the production domain further indicates that perceptual training results in highly generalized learning, suggesting that new tonal categories might have been established as a consequence of training. Taken together, the present results concerning the training of suprasegmental contrasts are consistent with the notion of a malleability of the adult learner's speech learning system across both perception and production.

ACKNOWLEDGMENTS

This research was conducted as part of a doctoral dissertation at Cornell University by the first author under the direction of the second and third authors. Portions of this study were reported at the 137th meeting (1999, Berlin), and the 142nd meeting of the Acoustical Society of America (2001, Fort Lauderdale). We would like to thank Michelle Spence, Brian Kim, Alicia Mackay, Eric Evans, Dorris Lok, and Bertrand Kotewall for assistance and technical support. Part of this research was supported by a grant from the University of Kansas General Research Fund to the third author.

Akahane-Yamada, R. (**1999**). ''Toward further understanding of second language speech learning: An approach utilizing speech technology,'' J. Acoust. Soc. Am. **105**, 1032.

Blicher, D. L., Diehl, R. L., and Cohen, L. B. (**1990**). ''Effects of syllable duration on the perception of the Mandarin tone2/tone3 distinction: Evidence of auditory enhancement,'' J. Phonetics **18**, 37–49.

Bluhme, H., and Burr, R. (**1971**). ''An audio-visual display of pitch for teaching Chinese tone,'' Stud. Linguist. **22**, 51–57.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (**1997**). ''Training Japanese listeners to identify English /r/ and /l/. IV. Some effects of perceptual learning on speech production,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (**1999**). ''Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production,'' Percept. Psychophys. **61**, 977–985.

Chao, Y. R. (**1948**). *Mandarin Primer* (Harvard University Press, Cambridge).

Clumeck, H. (**1980**). ''The acquisition of tone,''in *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), Vol. I.

Dreher, J. J., and Lee, P. C. (**1966**). *Instrumental Investigation of Single and Paired Mandarin Tonemes* (Advanced Research Laboratory, Douglas Aircraft Company, Huntington Beach, CA).

Flege, J. E. (**1993**). ''Production and perception of a novel, second-language phonetic contrast,'' J. Acoust. Soc. Am. **93**, 1589–1608.

Flege, J. E. (**1997**). ''The role of phonetic category formation in second-language speech learning,'' in *New Sounds 97. Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech*, edited by J. Leather and A. James, pp. 79–88.

Flege, J. E. (**1999**). ''The relation between L2 production and perception,'' in *Proceedings of the XIVth International Congress of Phonetics Sciences*, edited by J. Ohala, Y. Hasegawa, M. Ohala, D. Granveille, and A. Bailey (Department of Linguistics, University of California at Berkeley, Berkeley, CA), pp. 1273–1276.

Flege, J. E., Bohn, O-S., and Jang, S. (**1997**). ''The production and perception of English vowels by native speakers of German, Korean, Mandarin, and Spanish,'' J. Phonetics **25**, 422–470.

Fox, R., and Qi, Y. Y. (**1990**). ''Context effects in the perception of lexical tone,'' J. Chin. Ling. **18**, 261–283.

Howie, J. M. (**1976**). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge, MA).

Jamieson, D. G., and Morosan, D. E. (**1986**). ''Training non-native speech contrasts in adults: Acquisition of the English /θ/–/ð/ contrast by francophones,'' Percept. Psychophys. **40**, 205–215.

Jamieson, D. G., and Morosan, D. E. (**1989**). ''Training new, non-native speech contrasts: A comparison of the prototype and perceptual fading techniques,'' Can. J. Psychol. **43**, 88–96.

Kiriloff, C. (**1969**). ''On the auditory discrimination of tones in Mandarin,'' Phonetica **20**, 63–67.

Kuhl, P. K. (**2000a**). ''Language, Mind, and Brain: Experience alters perception,'' in *The New Cognitive Neurosciences*, 2nd ed., edited by M. S. Gazzaniga (The MIT Press, Cambridge, MA), pp. 99–115.

Kuhl, P. K. (**2000b**). ''A new view of language acquisition,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11850–11857.

Ladd, D. R., Silverman, K., Tolkmitt, F., Bergmann, G., and Scherer, K. (**1985**). ''Evidence for the independent function of intonation contour type, voice quality, and *F*0 range in signaling speaker affect,'' J. Acoust. Soc. Am. **78**, 435–444.

Leather, J. (**1983**). ''Speaker normalization in perception of lexical tone,'' J. Phonetics **11**, 373–382.

Leather, J. (**1990**). ''Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers,'' in *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language*

Speech, edited by J. Leather and A. James (University of Amsterdam, Amsterdam).
Li, C. N., and Thompson, S. (**1977**). ''The acquisition of tone in Mandarin-speaking children,'' J. Child Lang **4**, 185–199.
Liao, R. (**1983**). ''Suzhouhua danzidiao shuangzidiao de shiyan yanjiu,'' Yuyan Yanjiu **5**, 24–50.
Lin, T., and Wang, W. Y.-S. (**1985**). ''Shengdiao ganzhi wenti,'' Zhongguo Yuyan Xuebao **2**, 59–69.
Liu, F. (**1924**), *Szu Sheng Shih Yen Lu* (Ch'un Yi, Shanghai).
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., 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.
McClaskey, C. L., Pisoni, D. B., and Carrell, T. D. (**1983**). ''Transfer of training of a new linguistic contrast in voicing,'' Percept. Psychophys. **34**, 323–330.
Mertus, J. (**1989**). BLISS *Manual* (Brown University, Providence).
Miracle, W. C. (**1989**). ''Tone production of American students of Chinese: A preliminary acoustic study,'' J. Chin. Lang. Teach. Assoc. **24**, 49–65.
Moore, C. B., and Jongman, A. (**1997**). ''Speaker normalization in the perception of Mandarin Chinese tones,'' J. Acoust. Soc. Am. **102**, 1864–1877.
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. Hum. Percept. Perform. **8**, 297–314.
Rochet, B. L. (**1995**). ''Perception and production of second-language speech sounds by adults,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 379–410.
Rose, P. (**1987**). ''Considerations in the normalization of the fundamental frequency of linguistic tone,'' Speech Commun. **6**, 343–351.
Shen, X. S. (**1989**). ''Toward a register approach in teaching Mandarin tones,'' J. Chin. Lang. Teach. Assoc. **24**, 27–47.
Shen, X. S., and Lin, M. C. (**1991**). ''A perceptual study of Mandarin tones 2 and 3,'' Lang Speech **34**, 145–156.
Shi, F. (**1986**). ''Tianjin fangyan shuangzizu shengdiao fenxi,'' Yuyan Yanjiu **10**.
Shi, F. (**1994**). ''Beijinghua de shengdiao geju,'' in *Yuyin Conggao*, edited by F. Shi and R. Liao (Beijing Foreign Language Institute Press, Beijing), pp. 10–19.
Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). ''Training American listeners to perceive Mandarin tones,'' J. Acoust. Soc. Am. **106**, 3649–3658.
Wu, Z. J. (**1986**). *The Spectrographic Album of Mono-syllables of Standard Chinese* (Social Science, Beijing).

A narrow band pattern-matching model of vowel perception

James M. Hillenbrand[a)]
Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, Michigan 49008

Robert A. Houde
RIT Research Corporation, 125 Tech Park Drive, Rochester, New York 14623

(Received 5 February 2002; accepted for publication 2 August 2002)

The purpose of this paper is to propose and evaluate a new model of vowel perception which assumes that vowel identity is recognized by a template-matching process involving the comparison of narrow band input spectra with a set of smoothed spectral-shape templates that are learned through ordinary exposure to speech. In the present simulation of this process, the input spectra are computed over a sufficiently long window to resolve individual harmonics of voiced speech. Prior to template creation and pattern matching, the narrow band spectra are amplitude equalized by a spectrum-level normalization process, and the information-bearing spectral peaks are enhanced by a ''flooring'' procedure that zeroes out spectral values below a threshold function consisting of a center-weighted running average of spectral amplitudes. Templates for each vowel category are created simply by averaging the narrow band spectra of like vowels spoken by a panel of talkers. In the present implementation, separate templates are used for men, women, and children. The pattern matching is implemented with a simple city-block distance measure given by the sum of the channel-by-channel differences between the narrow band input spectrum (level-equalized and floored) and each vowel template. Spectral movement is taken into account by computing the distance measure at several points throughout the course of the vowel. The input spectrum is assigned to the vowel template that results in the smallest difference accumulated over the sequence of spectral slices. The model was evaluated using a large database consisting of 12 vowels in /hVd/ context spoken by 45 men, 48 women, and 46 children. The narrow band model classified vowels in this database with a degree of accuracy (91.4%) approaching that of human listeners. [DOI: 10.1121/1.1513647]

PACS numbers: 43.71.An, 43.72.Ar, 43.66.Ba [KRK]

I. INTRODUCTION

A longstanding goal of speech perception research is to explain the perceptual mechanisms that are involved in the recognition of vowel identity. As with most other phenomena in phonetic perception, this problem has resisted straightforward solution, due in large measure to the variation in acoustic patterns that is observed when like vowels are spoken by different talkers, in different phonetic environments, at different speaking rates, at different fundamental frequencies, or with varying levels of contrastive stress. A wide range of models have been proposed to address one or more of these variability problems (for a review, see Rosner and Pickering, 1994). This diversity in theoretical approaches stands in contrast to a rather limited set of choices in the underlying acoustic representations that drive these recognition models. The overwhelming majority of vowel identification models have assumed that the recognition process is driven by an underlying representation consisting of either the formant frequency pattern of the vowel (with or without fundamental frequency as a normalizing factor) or the gross shape of the smoothed spectral envelope. Competition between these two quite different approaches to vowel identification has occupied a fair amount of attention in the literature. Excellent reviews of this literature can be found in Bladon and Lindblom (1981), Bladon (1982), Klatt (1982a, b), Zahorian and Jagharghi (1986), and Ito *et al.* (2001). To summarize the issues briefly, the main idea underlying formant representations is the notion that the recognition of vowel identity (and many other aspects of phonetic quality) is controlled not by the detailed shape of the spectrum but rather by the distribution of formant frequencies, chiefly the three lowest formants (F_1–F_3). The virtues of formant representations include the following: (1) formant representations are quite compact relative to whole spectrum models, in keeping with commonly held notions about the low dimensionality of perceptual space for vowels (e.g., Pols *et al.*, 1969); (2) formant representations allow for a number of fairly straightforward solutions to talker normalization problems that arise from variation in vocal tract length (e.g., Disner, 1980; Miller, 1989; Nearey, 1992; Nearey *et al.*, 1979; Syrdal and Gopal, 1986; Hillenbrand and Gayvert, 1993); (3) reasonable correspondences have been found between formant representations and perceptual dimensions inferred from measures of perceived similarities among vowels (e.g., Miller, 1956; Pols *et al.*, 1969; cf. Bladon and Lindblom, 1981); (4) speech or speechlike signals that are synthesized from formant representations are typically highly intelligible, even in cases involving gross departures in detailed spectral shape between the original and reconstructed signals (e.g., Remez *et al.*, 1981); (5) several pattern recognition studies have shown that naturally spoken vowels can be recognized with reason-

a)Electronic mail: james.hillenbrand@wmich.edu

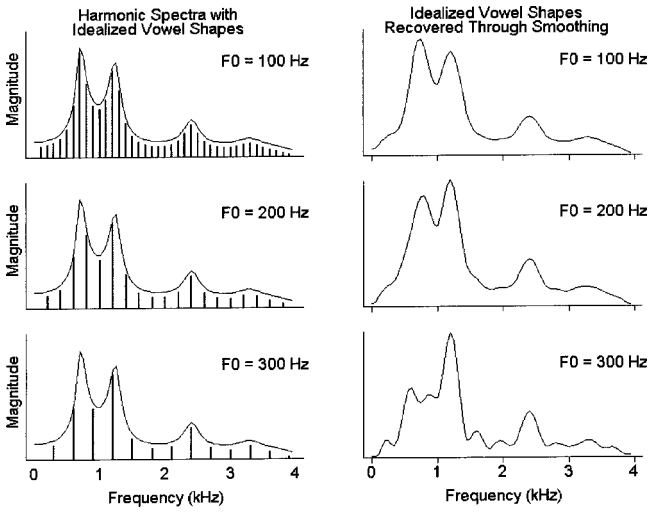


FIG. 1. Left: Harmonic spectra and estimates of the idealized smooth envelope shapes (smooth curves—see text) for the vowel /ɑ/ spoken at three fundamental frequencies, but with the same vocal tract frequency response curve and glottal source shape. Right: Attempting to recover the idealized envelope shape through cepstral smoothing. Note that the smoothing operation recovers the vocal tract filter shape well at 100 Hz, reasonably well at 200 Hz, but very poorly at 300 Hz. The distortion in the smoothed envelope shape at $F_0 = 300$ Hz is due to aliasing. For a detailed discussion of this phenomenon, see de Cheveigné and Kawahara (1999).

able accuracy based on formant measurements, particularly when fundamental frequency, formant movements, and duration are taken into account (e.g., Nearey *et al.*, 1979; Syrdal and Gopal, 1986; Hillenbrand and Gayvert, 1993; Hillenbrand *et al.*, 1995, 2000b); and (6) probably of greatest importance, the widely cited evidence from Klatt (1982a) showing that formant frequencies are easily the most important acoustic parameters affecting vowel quality, with other manipulations of spectral shape (e.g., high- and low-pass filtering, spectral tilt, formant amplitudes, notches, formant bandwidths, etc.) resulting in little or no change in phonetic quality (see also related evidence from Assmann and Summerfield, 1989). There are, however, quite substantial problems with formant theory (see Bladon *et al.*, 1982; Zahorian and Jagharghi, 1993; Ito *et al.*, 2001 for reviews). These problems include (1) the unresolved and quite possibly unresolvable problem of tracking formants in natural speech; (2) evidence showing that some spectral details other than formant frequencies can affect vowel quality (e.g., Chistovich and Lublinskaya, 1979; Bladon, 1982); and (3) the observation that errors made by human listeners do not appear to be consistent with the idea that vowel quality is perceived on the basis of labeled formants (Klatt, 1982b; Ito *et al.*, 2001). The primary weakness of spectral shape approaches is generally thought to be the difficulty of this class of models to accommodate what would appear to be rather compelling evidence showing that large changes can be made in detailed spectral shape without affecting phonetic quality, as long as the formant frequency pattern is preserved (e.g., Remez *et al.*, 1981; Klatt, 1982a).

A feature that is common to nearly all spectral shape models is the derivation of the spectral envelope through some kind of smoothing operation. The motivation underlying the smoothing is straightforward. The left panel of Fig. 1 shows the vowel /ɑ/ spoken at three fundamental frequencies, but with the same vocal tract frequency response curve and glottal source shape. Vowel quality will be similar in the three cases (but not identical—e.g., Miller, 1953; Fujisaki and Kawashima, 1968) despite the obvious differences in their harmonic spectra. Smoothing is intended to remove the largely irrelevant harmonic detail. The smooth curves to the left in this figure represent the idealized smooth shape for this vowel. This idealized shape corresponds to a hypothetical filter representing the combined effects of the vocal tract frequency response curve, the shape of the glottal source spectrum, and the radiation characteristic. (Hereafter, to avoid this awkwardly long phrase, we will refer simply to the *vocal tract filter*, although it should be understood that the detailed shape of this hypothetical equivalent filter is controlled by the vocal tract transfer function, the spectrum of the source signal, and the radiation characteristic. Easily the most important of these three functions in signaling differences in vowel identity is the vocal tract frequency response curve.) As de Cheveigné and Kawahara (1999) note, for the case of phonated vowels, the vocal tract filter function is effectively sampled at discrete frequencies corresponding to the harmonics of the voice source. As such, the ability to estimate the shape of the vocal tract filter based on the amplitudes of the voice-source harmonics is subject to the well-known constraints of sampling theorem. de Cheveigné and Kawahara describe a series of simple tests that clearly demonstrate that the vocal tract transfer function is severely undersampled at higher fundamental frequencies. Of direct relevance to the present discussion, de Cheveigné and Kawahara go on to show that the smoothing that is so commonly used in spectral shape models results in aliasing-induced distortion of the estimated vocal tract filter function at even moderately high fundamental frequencies (see especially Fig. 4 of de Cheveigné and Kawahara, which provides a very concise summary of the aliasing problem). This aliasing effect is illustrated in the right half of Fig. 1, which shows smoothed spectra derived by a cepstral smoothing operation. Note that the smoothing operation recovers the vocal tract filter shape well at 100 Hz, reasonably well at 200 Hz, but very poorly at 300 Hz.

In response to this undersampling problem, de Cheveigné and Kawahara proposed a novel *missing data model* of vowel identification in which the *unsmoothed* harmonic spectrum is directly compared to a set of smoothed templates, with spectral differences computed *only at frequencies corresponding to voice-source harmonics*. Small-scale simulations of two versions of the *missing data model* showed good recognition of five synthetic vowels generated at a wide range of fundamental frequencies (20–300 Hz) when compared against templates consisting of known spectral envelopes for each vowel. An otherwise identical model using smoothed input spectra showed extreme sensitivity to fundamental frequency.

The purpose of the present article is to address two important limitations of the innovative work described by de Cheveigné and Kawahara (1999), one theoretical, the other experimental. A central feature of the *missing data model* is the notion that the input spectrum is compared to the tem-

plates only at harmonic frequencies. This is a reasonable restriction, for exactly the reasons discussed by de Cheveigné and Kawahara, but it is one which imposes some rather strict signal processing demands on the pattern recognizer. The instantaneous fundamental frequency must be estimated, presumably with considerable precision since it would not take much of a pitch estimation error to result in a situation in which the pattern recognizer ended up comparing the input spectrum with the template exclusively at frequencies that are irrelevant to vowel identity. Further, as the authors note, modifications would have to be made to the *missing data model* to handle aperiodic or marginally periodic signals, such as whispered or breathy vowels or vowels spoken with rapid changes in the fundamental frequency. de Cheveigné and Kawahara offer some reasonable speculations about methods that might be used to address these kinds of cases, but in the present article we will propose a vowel identification model which removes the harmonics-only restriction entirely. In common with the *missing data model*, our *narrow band pattern matching model* compares unsmoothed, high-resolution input spectra (i.e., spectra computed over a sufficiently long window to resolve voice-source harmonics) directly with a set of smooth vowel templates. However, unlike the *missing data model*, which computes spectral distances only at harmonic frequencies, spectral distances in our *narrow band model* are computed at all frequencies. Results will be presented which we believe demonstrate that the harmonics-only restriction needlessly complicates the pattern matching and that a recognition algorithm that gives no special treatment to harmonics accurately recognizes vowels spoken with a wide range of fundamental frequencies.

The second limitation which we wish to address is that the *missing data model* was not evaluated on naturally spoken utterances. The test signals that were classified by the model consisted of synthetic vowels generated with perfectly periodic source signals and static spectral envelopes. Further, the templates for the five vowel types that were used consisted of the known transfer functions that were used in generating the test signals. While these tests were quite reasonable in light of the goals of de Cheveigné and Kawahara's paper (demonstrating the aliasing effects discussed above), it remains unclear how well a model that is driven by unsmoothed harmonic spectra would classify natural spoken vowels. In the present work, our narrow band model will be tested using naturally spoken utterances comprising 12 American English vowels produced by a large group of men, women, and children. Further, the vowel templates will be derived empirically based on an analysis of those naturally spoken utterances.

II. THE NARROW BAND PATTERN-MATCHING MODEL

A. Preliminary comments

Before describing the details of the narrow band model, we should note that the model does not represent an attempt at a faithful simulation of the physiological response properties of the auditory system. Indeed, we believe that the clearest lesson of Klatt's (1982a) study is that a model of peripheral auditory processing in and of itself is inherently incapable of accounting for the recognition of vowel quality. Klatt's findings show that there are many aspects of spectral shape that are quite audible (and therefore must be preserved in any faithful auditory model) but which contribute very little to judgments of vowel timbre. A complete model of vowel recognition would need to incorporate not only a precise simulation of the low-level auditory representations from which generic judgments of timbre might be derived, but also of the (presumably) decision-level psychological mechanisms that must be involved in translating these low-level auditory representations into vowel percepts (i.e., ignoring, or largely ignoring, features such spectral tilt, formant amplitude relations, spectral notches, etc.). In the present work we have adopted a simpler approach in which a few signal processing steps are intended to model the most psychologically important aspects of vowel recognition, with both low-level analysis and decision-level mechanisms rolled into single process. In general, our strategy in constructing the model was primarily one of working backward from key perceptual data such as those of Klatt (1982a) and Ito *et al.* (2001) toward a psychologically plausible processing scheme rather than one of working forward from peripheral auditory processing principles.

B. Template creation

Our model assumes that the reference patterns defining each vowel category consist of *sequences* of smooth spectra. We assume a sequence of spectra rather than a single spectrum sampled at steady state because of the large body of evidence implicating a strong role for spectral movement in vowel recognition (e.g., Strange *et al.*, 1983; Nearey and Assmann, 1986; Jenkins *et al.*, 1983; Parker and Diehl, 1984; Andruski and Nearey, 1992; Jenkins and Strange, 1999; Hillenbrand and Gayvert, 1993; Hillenbrand and Nearey, 1999; Assmann and Katz, 2000, 2001). We assume that the individual spectral shape templates in the sequence are derived simply by averaging the narrow band spectra of like vowels sampled at comparable time points throughout the course of the vowel (followed by light smoothing—see below). Figure 2 shows a sequence of templates for /æ/ sampled at 15%, 30%, 45%, 60%, and 75% of vowel duration derived from adult male talkers using tokens from a large /hVd/ database (described below). Figure 3 shows the signal processing steps that are used to generate the individual spectra that are averaged to produce smooth templates such as those illustrated in Fig. 2. The signal processing steps, which are detailed below, consisted of (1) calculation of a narrow band (i.e., long time window) Fourier spectrum, (2) spectrum-level normalization, (3) enhancement of spectral peaks by a thresholding procedure, and (4) overall amplitude normalization. The initial Fourier spectrum was computed over a relatively long time window. In the experiments reported below, we used a 512-point (64 ms) Hamming-windowed FFT with 8 kHz sampled waveforms and 6 dB per octave high-frequency preemphasis. Linear frequency and amplitude scales were used [see panel (a) of Fig. 3]. The next step involved the application of a broadband spectrum-level normalization (SLN) function. The motivation behind this

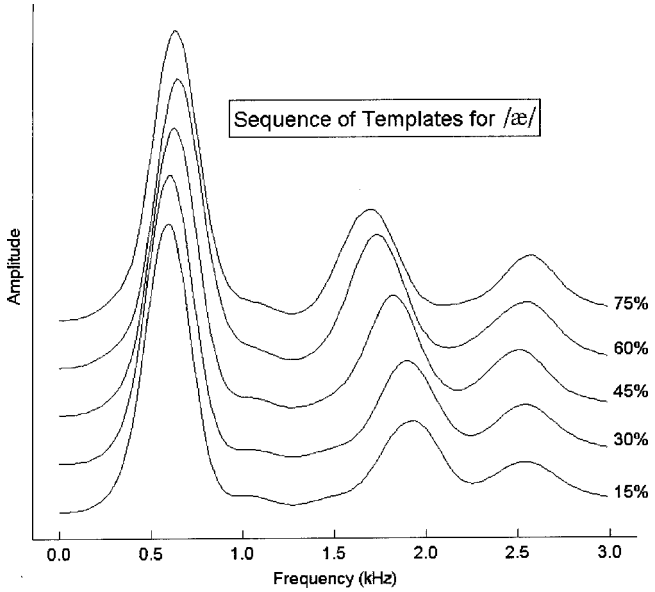


FIG. 2. Sequence of five vowel templates for /æ/ computed at 15%, 30%, 45%, 60%, and 75% of vowel duration. Note that successive templates have been offset on the amplitude scale so that the change in spectral shape over time can be seen more clearly.

step was to reduce as much as possible within-vowel-category differences in formant amplitude relations, following data such as Klatt (1982a) indicating that formant-amplitude variation, while quite audible to listeners, contributes little to perceived vowel color. The idea of the SLN operation, then, was simply to attenuate spectral regions of relatively high amplitude and amplify regions of relatively low amplitude, reducing the magnitude of amplitude differences among broad spectral peaks. The SLN operation was implemented by computing a gain function that

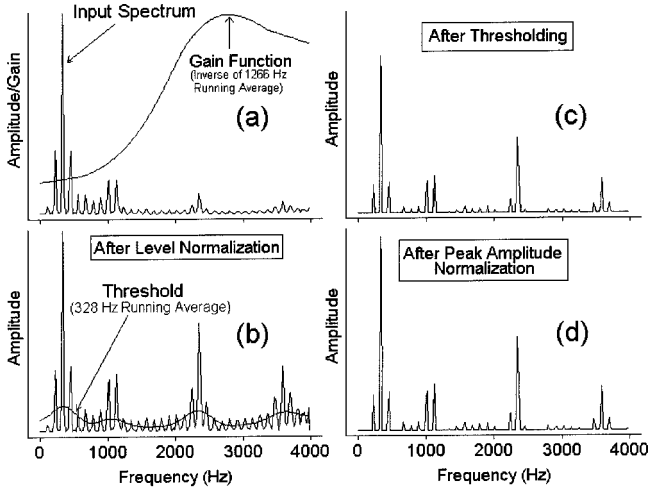


FIG. 3. Signal processing steps used in the narrow band pattern matching model: (a) FFT computed over a 64-ms Hamming-windowed segment and broadband spectrum-level normalization (SLN) function computed as the inverse of a 1266-Hz Gaussian-weighted running average of spectral amplitudes; (b) spectrum after processing by the SLN operation and a threshold function computed as a 328-Hz Gaussian-weighted running average of spectral amplitudes; (c) spectrum after thresholding, i.e., after zeroing all spectral values that lie below the threshold function; and (d) amplitude normalization, implemented by scaling the largest peak in the spectrum to a constant.

was relatively low in spectral regions with high average amplitude and, conversely, was relatively high in spectral regions with low average amplitude. The SLN function that is shown in panel (a) of Fig. 3 is simply the inverse of the Gaussian-weighted running average[1] of spectral amplitudes computed over an 81-channel (1265.6 Hz) spectral window. Panel (b) of Fig. 3 shows the spectrum after application of the broadband SLN operation. It can be seen that the variation in spectral peak amplitudes has been considerably reduced, although by no means entirely eliminated. The size of the smoothing window is a compromise, determined by inspecting a large number of individual normalized and unnormalized spectra. The rather large window size that was selected represents a compromise between two competing considerations. Very large window sizes produce rather limited benefit with respect to the goal of minimizing the importance of formant amplitude differences but have the advantage that they seldom amplify minor spectral peaks that are of little or no perceptual relevance. Smaller window sizes, on the other hand, do an excellent job of reducing the range of formant amplitude variation but can sometimes have the undesirable effect of amplifying minor spectral peaks. All of the informal experimentation involved in selecting a smoothing window size was carried out using a CVC database (Hillenbrand *et al.*, 2000b) other than the one used to evaluate the model.

The next signal processing step consisted of a thresholding procedure. The idea here was simply to emphasize the spectral peak regions (both narrow and broad band) that are known to have the greatest influence on vowel identity and to suppress the largely irrelevant spectral components in between harmonic peaks and in the less perceptually significant valleys that lie in between broad spectral peaks. This step was implemented by defining a threshold function as the Gaussian-weighted running average of spectral amplitudes computed over a 21-channel (328.1 Hz) spectral window. The running average is then subtracted from the spectrum, with all negative values (i.e., values below the threshold) set to zero.[2] As with the gain function described above, the size of the averaging window used for the threshold operation was determined through extensive informal experimentation using a vowel database other than the one used to evaluate the model. The process involved examination of a large number of individual cases of spectra with and without the thresholding operation and trying to find a smoothing window size that appeared to do the best job of enhancing the information-bearing aspects of the spectra (i.e., harmonics, especially those defining formant peaks). The final signal processing step involved amplitude normalization, implemented by scaling the largest peak in the spectrum to a constant [Fig. 3(d)].

In the tests reported below, separate template sequences were created for men, women, and children. These templates were created by averaging like vowels at like time points throughout the course of the vowel. For most of the tests reported below, we represent each vowel as a sequence of five templates, centered at 15%, 30%, 45%, 60%, and 75% of vowel duration, based on hand-measured values of vowel start and stop times from Hillenbrand *et al.* (1995). (The is-

sue of how many spectral slices are required will be considered below.) For example, the first spectrum of the /i/ template sequence for men was created by averaging all adult male tokens of /i/ sampled at 15% of vowel duration. Following the averaging, we apply light smoothing to each of the averaged spectra with a 171.9-Hz Gaussian-weighted running average. We assume that this final smoothing step has no counterpart in human perception and is a simple concession to the fact that we were forced by practicalities to create the templates from a limited number of tokens of each vowel. As will be explained below, each template was constructed by averaging roughly 40 examples of each vowel for each talker group. Although this is a relatively large database by experimental standards, it is a small fraction of the amount of speech a listener is likely to hear during the course of even a single day.

We attach no special importance to the choice of a sequence of five spectra to represent each vowel, as opposed to three or seven or any number of other reasonable alternatives. We will, however, report results demonstrating the simple point, in keeping with a good deal of evidence from human listeners, that a template sequence produces far better recognition performance than a pattern-matching scheme based on templates computed at a single time slice. Similarly, the choice of equally spaced samples between 15% and 75% is somewhat arbitrary. For illustration, Fig. 4 shows templates sampled at 30% of vowel duration for men (panel a), women (panel b), and children (panel c). Note that the formant frequency patterns that are traditionally associated with these vowels are well preserved in most but not all of these templates. For example, note the merger of F_1 and F_2 in the adult female /ɔ/ template (F_2 does not quite merge with F_1 in the child /ɔ/ template, but instead shows up as a soft shoulder rather than a peak) and the merger of F_2 and F_3 in all three /ɝ/ templates.

C. Computing distances between input spectra and template sequences

The distance between the input spectrum and a smoothed template for any given time slice is computed as the sum of the channel-by-channel absolute differences between the two spectra, divided by the total energy in the template, computed as the sum of template amplitudes across all 256 channels (i.e., a city-block distance, normalized for overall template amplitude). Unlike the method described by de Cheveigné and Kawahara (1999), which computes distances at voice-source harmonics only, distances are computed for all channels. Spectral differences between the narrow band input spectra and the smooth templates will obviously be large at frequencies remote from the harmonics, especially at high F_0, and especially in the deep valleys between the harmonics. A key assumption underlying our model is that these interharmonic differences should be approximately equally large to all vowel templates, presumably resulting in a more-or-less constant source of noise across all templates.

At each individual time slice, the spectral-distance algorithm produces a 12-element vector containing distances between the input spectrum and templates for each of the 12 vowel types derived from that same time slice (see Fig. 5). For illustration, Table I shows a set of distance vectors for a single token of /hæd/ using a method incorporating distances sampled at five time slices (15%, 30%, 45%, 60%, and 75% of vowel duration). The final distance vector used for recognition is created from the five individual vectors simply by computing a weighted average of the distances computed at each of the five time points. The recognition algorithm chooses the vowel corresponding to the smallest token-to-template distance. A weighted average was used based on an expectation that recognition performance would be improved if somewhat more weight were assigned to distances computed early in the vowel (typically corresponding to ''steady-state'' times measured in studies such as Peterson and Bar-

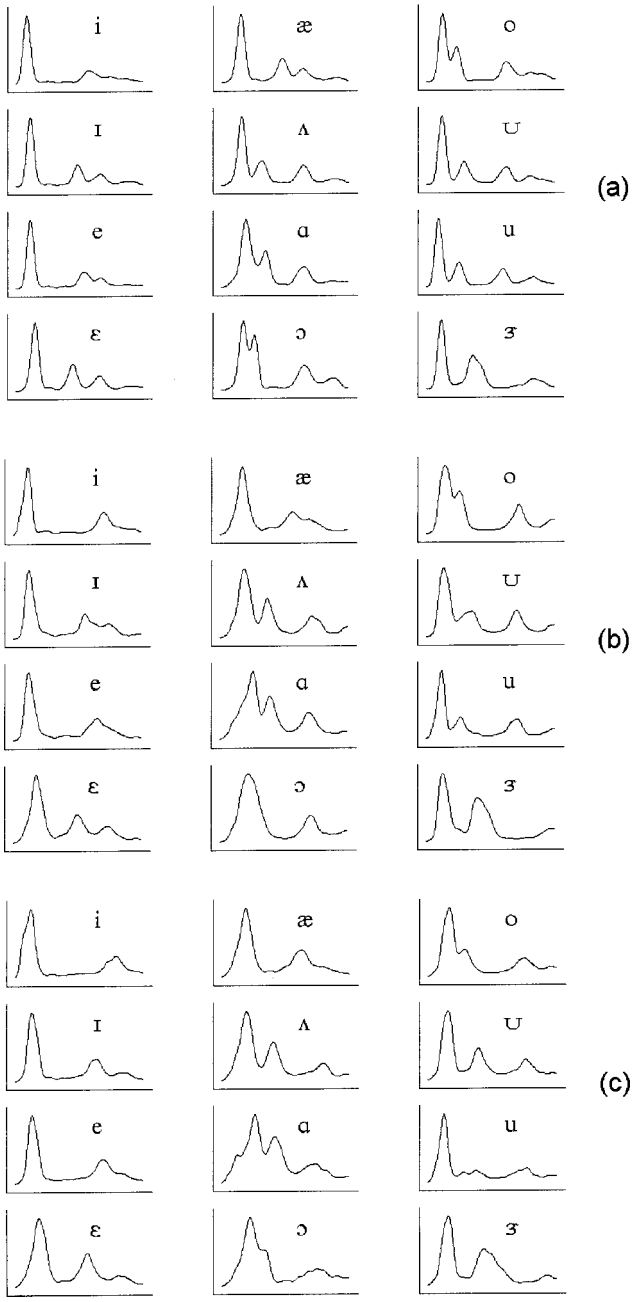


FIG. 4. Vowel templates computed at 30% of vowel duration for men (panel a), women (panel b), and children (panel c). The frequency scales are linear with a range of 0 to 4 kHz.

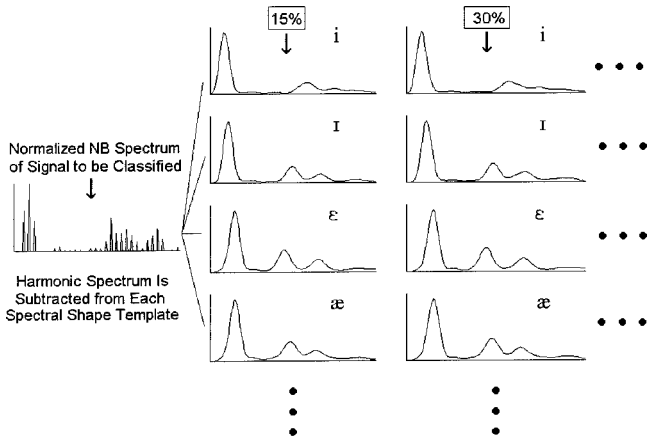


FIG. 5. Illustration of the recognition algorithm used in the narrow band model. The narrow band spectrum computed at 15% of vowel duration is compared to the 12 vowel templates computed at the same time point (only 4 of which are shown here); the narrow band spectrum at 30% of vowel duration (not shown) is then compared to the 12 vowel templates computed at 30% of vowel duration, and so on.

ney, 1952) than to offglide portions of the vowel which show the influence of the final consonant. Testing with a wide variety of weighting schemes showed that this was, indeed, the case. However, the advantage of a weighted over an unweighted average proved to be very slight. The five-slice results reported below used a scheme that assigned a weight of 1.0 to the first four slices and a weight of 0.6 to the last slice. The minimum distance in each row of Table I is shown in bold. Note that for the first three time slices, the minimum distance is to the (incorrect) /ɛ/ template. However, the minimum distance in the weighted-average vector, which takes the offglide into account as well, corresponds to the /æ/ category that was intended by the talker.

III. EVALUATION

The test signals consisted of 1668 naturally spoken /hVd/ utterances recorded by Hillenbrand *et al.* (1995). This database consists of 12 vowels (/i,ɪ,e,ɛ,æ,ɑ,ɔ,o, U,u,ʌ,ɝ/) spoken by 45 men, 48 women, and 46 10- to 12-year-old children. The stimuli, which were originally digitized at 16 kHz, were digitally low-pass filtered at 3.7 kHz, down-sampled to 8 kHz, and scaled to maximum peak amplitude. Omitted from the database *for the purposes of template creation only* were 192 utterances which had shown identification error rates of 15% or greater in the listening study described in the original 1995 article. The 1476-syllable stimulus set used for template construction consisted of roughly equal numbers of tokens spoken by men (482), women (522), and children (472). The full database of 1668 utterances was used during the recognition phase of the evaluation.

IV. RESULTS

Overall classification accuracy for the five-slice method was 92.0% for men, 92.2% for women, and 90.0% for children, with a mean 91.4% across the three talker groups. The corresponding intelligibility figures for human listeners, as measured in Hillenbrand *et al.* (1995), are 94.6%, 95.6%, and 93.7%, with a mean 94.6% across the three groups. Although the accuracy of the narrow band recognizer was ~3% poorer than the listeners, it is important to note that the listeners had access to duration cues, and there is clear evidence that duration has a modest but significant influence on vowel identification by human listeners. For example, Hillenbrand *et al.* (2000a), using resynthesized versions of a 300-utterance subset of the /hVd/ signals tested here, found a 2% decrease in vowel intelligibility when vowel duration was eliminated as a cue by setting the duration of all vowels to a neutral value. Further, overall vowel intelligibility was reduced by 4.5% to 5.0% when vowels were either shortened (fixed at 144 ms, 2 standard deviations below the grand mean of all vowel durations) or lengthened (fixed at 400 ms, 2 standard deviations above the grand mean). The current version of the narrow band model uses spectral information only.

Table II shows a confusion matrix relating the vowel intended by the talker to the vowel recognized by the model. The matrix combines results from all three talker groups. The overall look of the confusion matrix is similar in many respects to comparable matrices derived from human listeners. As with human listeners, nearly all of the confusions involve vowels that are quite close to one another in phonetic space; e.g., /ɑ/ is confused mainly with /ɔ/ and /ʌ/, /ɛ/ is confused mainly with /ɪ/ and /æ/, etc. There is also clear evidence that individual stimuli that were less intelligible to human listeners were far more likely to be misclassified by the narrow band recognizer. The error rate for the narrow band model for the 192 tokens with listener error rates of 15% or higher

TABLE I. City-block spectral distances between a sequence of spectra for a sample input signal (/hæd/, spoken by a man) and templates for each of 12 vowels. The first five rows show spectral distances sampled at equally spaced points from 15% to 75% to of vowel duration; the last row shows the weighted mean of these five distances, using weights of 1.0, 1.0, 1.0, 1.0, and 0.6 for the distances computed at 15%, 30%, 45%, 60%, and 75% of vowel duration, respectively. The minimum distance in each row is shown in bold.

	Vowel											
Time slice	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
Slice 1 (15%)	1.071	0.970	0.962	**0.871**	0.872	0.963	0.950	0.950	0.973	1.028	0.929	0.946
Slice 2 (30%)	1.079	0.964	1.012	**0.858**	0.872	0.953	0.948	0.968	0.986	1.034	0.926	0.973
Slice 3 (45%)	1.068	0.945	1.032	**0.879**	0.881	0.963	0.963	0.990	0.989	1.028	0.945	0.978
Slice 4 (60%)	1.091	0.950	1.063	0.885	**0.868**	0.964	0.960	1.003	1.015	1.056	0.958	0.998
Slice 5 (75%)	1.094	0.986	1.079	0.930	**0.881**	0.911	0.928	1.006	0.990	1.055	0.942	0.994
Weighted mean	1.079	0.961	1.025	0.880	**0.874**	0.954	0.952	0.982	0.991	1.039	0.940	0.977

TABLE II. Confusion matrix relating the vowel intended by the talker to the vowel recognized by the narrow band classification model. Results are summed across the three talker groups. Values on the main diagonal, indicating correctly recognized stimuli, are shown in boldface.

	Vowel as classified by the narrow band pattern recognition model											
	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
/i/	**93.5**	0.7	2.9						0.7	2.2		
/ɪ/		**93.5**	3.6	0.7					1.4	0.7		
/e/	3.5	2.1	**92.9**					0.7	0.7			
/ɛ/		4.3		**88.4**	4.4		0.7		0.7		1.4	
/æ/				13.8	**85.5**	0.7						
/ɑ/						**91.3**	6.5				2.2	
/ɔ/					0.7	11.4	**86.4**				1.5	
/o/			0.7				0.7	**94.3**	2.8	1.4		
/ʊ/			0.7					0.7	**94.2**	0.7	2.1	1.4
/u/	5.6	0.7				0.7		3.5	2.2	**87.2**		
/ʌ/						1.4	2.9		5.1		**90.6**	
/ɝ/		0.7		0.7					0.7			**97.8**

was 30.2%, more than five times greater than the error rate for the remaining well identified tokens (5.9%).[3]

Along with many similarities between the labeling patterns of humans and those of the narrow band model, there are some differences which reveal limitations of the current version of the model to simulate human vowel perception in detail. The most important difference is the occasional tendency of the model to confuse high front vowels with high back vowels. For example, note the ~3% of /i/ tokens that were incorrectly recognized as /u/ or /U/, the ~2% of /ɪ/ tokens that were incorrectly recognized as /u/ or /U/ and, most important, the ~6% of /u/ tokens that were incorrectly recognized as /i/ or /ɪ/. While these front–back confusions occurred on a small fraction of the utterances, this type of error is nearly nonexistent in human listener data (e.g., see Table VII in Hillenbrand *et al.*, 1995).

In keeping with a large body of evidence from human listeners, the narrow band model is better at recognizing vowels when spectral change is incorporated into the distance measure than when recognition is based on a single spectral slice. The five columns to the left in Table III show recognition rates by the model based on a single spectral slice sampled at either 15%, 30%, 45%, 60%, or 75% of vowel duration. These single-slice recognition rates, which typically vary from 75% to 80% correct, are considerably lower than the ~91% correct performance of the model that uses a sequence of five spectral slices.

The last three groups of columns in Table III compare versions of the recognition model incorporating two, three, and five spectral slices. The two-slice model used slices 1 and 5 with weights of 1 and 0.6, respectively, while the three-slice model used slices 1, 3, and 5 (15%, 45% and 75%) with weights of 1.0, 1.0, and 0.6. As the table shows, all of the multi-slice models performed better than the single-slice models, and performance differences among the two-, three-, and five-slice models are marginal at best. For the two-slice model, extensive testing with different combinations of slice locations showed a consistent advantage for more widely spaced slices; e.g., locations such as 1-5, 1-4, 2-5, etc., produced better recognition accuracy than locations such as 2-3, 2-4, etc. Variation in weighting scheme produced only slight differences in recognition accuracy. Similarly, the three-slice model performed better with more widely spread slice locations (e.g., 1-3-5, 1-3-4, 2-3-5, etc.) than tightly spaced locations (e.g., 1-2-3, 2-3-4, etc.), with little effect of weighting scheme. Taken as a whole, the results in Table III are consistent with data from both human listening studies and studies using pattern recognition methods indicating that spectral change patterns play a secondary but quite important role in vowel identification (for reviews, see Strange, 1989; Nearey, 1989; Hillenbrand and Nearey, 1999).

Table IV was designed to provide some insight into the benefit that is derived from the two major signal processing steps that are used to generate the test spectra and templates. The table shows recognition accuracy by the narrow band pattern matching model with and without thresholding and/or SLN. Results are for the five-slice version of the model de-

TABLE III. The five columns to the left of the table show the percent correct recognition accuracy by the narrow band model based on a single spectral slice, taken at 15% (slice 1), 30% (slice 2), 45% (slice 3), 60% (slice 4), or 75% (slice 5) of vowel duration. The three columns to the right of the table show percent correct figures for two-, three- and five-slice versions of the recognition model. Results are shown separately for syllables spoken by men, women, and children, with the bottom row showing a mean computed across the three talker groups.

	Slice no.					No. of slices		
	1	2	3	4	5	2	3	5
Men	79.8	83.7	85.4	82.2	74.4	94.3	93.1	92.0
Women	74.5	79.7	78.5	76.6	66.3	92.0	92.4	92.2
Children	72.1	75.2	77.2	72.6	65.9	85.5	89.3	90.0
Mean	75.5	79.5	80.4	77.1	68.9	90.6	91.6	91.4

TABLE IV. Recognition accuracy by the narrow band pattern-recognition model with and without thresholding and spectrum-level normalization (SLN).

	Thresholding and SLN	Thresholding, no SLN	No thresholding, SLN	No thresholding, No SLN
Men	92.0	92.0	60.0	50.9
Women	92.2	90.1	58.7	60.8
Children	90.0	86.4	61.1	57.8
Mean	91.4	89.5	59.9	56.5

scribed above. It can be seen that the broadband SLN operation, which was intended to reduce the importance of variation in formant amplitude relationships, provides at best a very small benefit. We retain the SLN operation because it does no harm, and it is also quite possible that the SLN operation would prove its usefulness in recognizing less well behaved test signals, such as those contrived by Klatt (1982a) in which spectral-shape features such as formant amplitude relations and spectral tilt have been deliberately altered from their typical values.

The most striking finding in Table IV is that the thresholding operation, implemented by zeroing out all spectral values below a 328-Hz Gaussian running average of spectral amplitudes, has a dramatic effect on the performance of the recognizer. When the thresholding operation is removed, the performance of the recognizer drops by over 30 percentage points. We assume that this operation improves recognition accuracy by suppressing spectral components in between harmonic peaks and in the valleys between formants, emphasizing those aspects of the spectrum that are most closely associated with vowel identity.

V. DISCUSSION

Our primary conclusion from these findings is that it is possible in principle to recognize vowels using a pattern-matching scheme involving the direct comparison of unsmoothed harmonic spectra with a set of empirically derived vowel templates. The overall recognition accuracy for the model approached but did not quite equal that of human listeners, and it is clear that at least part of the performance advantage for listeners can be attributed to listeners' use of duration cues. A goal of future work with this model is to develop some method to incorporate duration information along with the spectral distances that form the basis of the current method. On first glance this would seem to be a straightforward problem, but our earlier work on the use of duration cues by human listeners (Hillenbrand *et al.*, 2000a) showed that listeners are quite smart and flexible in their use of vowel duration. For example, listeners made virtually no use of duration cues in distinguishing pairs such as /i/−/ɪ/, /e/−/ɛ/, and /u/−/U/, in spite of large and systematic differences in duration separating these vowel pairs. Modeling work suggested that listeners assign little or no weight to duration for these vowel pairs because the vowels can be separated reliably on the basis of spectral cues alone. On the other hand, listeners make considerable use of duration cues in distinguishing vowels such as /æ/-/ɛ/ and the /ɑ/-/ɔ/-/ʌ/ cluster because these vowels show a greater degree of overlap in their spectral properties. We have experimented with some simple schemes for incorporating duration measures and have met with only modest success. We suspect that a humanlike scheme will be needed that assigns considerable weight to duration for some vowels and little or none for others.

It is of some interest to note that the recognition accuracy for the model was at most very slightly reduced at higher fundamental frequencies. For example, the model was as accurate in recognizing tokens spoken by women (92.2%) as men (92.0%), in spite of the roughly three-quarters of an octave difference in average fundamental frequency between the vowels of men ($\bar{\chi}=131$ Hz) and women ($\bar{\chi}=220$ Hz) in this database. This finding is consistent with labeling results from human listeners, who actually recognized syllables spoken by the women slightly (but significantly) better than those of the men (Hillenbrand *et al.*, 1995). There was a roughly 2 percentage point drop in recognition accuracy by the model for tokens produced by the children, which is consistent with a similar-size drop of 1–2 percentage points shown by our listeners. It is not entirely clear, for either the listeners or the model, whether the slightly lower intelligibility of the children's tokens has anything to do with their higher average fundamental frequencies ($\bar{\chi}=237$ Hz). It seems likely that the children's vowels were simply produced more variably and with a bit less articulatory precision than those of the adults (see Kent, 1976, for a review). Literature on the apparently simple question of whether vowel intelligibility degrades with increasing F_0 is surprisingly mixed. As noted, our study of naturally spoken /hVd/ syllables found no simple relationship between F_0 and vowel intelligibility. The same was true in a later study in which listeners identified formant synthesized versions of a 300-utterance subset of the same /hVd/ database (Hillenbrand and Nearey, 1999). The syllables were generated with a formant synthesizer driven by the original F_0 contours and either the original formant contours or flattened formant contours. Labeling results showed no evidence of a simple drop in intelligibility with increasing F_0. On the other hand, in Hillenbrand and Gayvert (1993), 300-ms signals with static formant patterns were synthesized based on F_0 and formant measurements of each of the 1520 signals in the Peterson and Barney (1952) database. There was a small but highly reliable drop in intelligibility with increasing F_0, whether the signals were generated with monotone pitch (men: 74.4%, women: 72.2%, children: 70.0%) or with falling pitch (men: 76.9%, women: 73.8%, children: 72.1%). However, even here the relationship between F_0 and intelligibility was

hardly simple since the roughly three-quarters of an octave difference in F_0 between men and women was accompanied by about the same small drop in intelligibility as the roughly one-quarter octave difference in F_0 between women and children. Ryalls and Liberman (1982) found that vowels with formants appropriate for a male talker were more intelligible at a 135-Hz F_0 than at 100 Hz, and that vowels with formants appropriate for a female talker were more intelligible at 185 Hz than at 250 Hz. Similarly, Sundberg and Gauffin (1982) found a decrease in intelligibility for vowels synthesized at F_0's between 260 and 700 Hz. For a nice summary and discussion of this issue, see de Cheveigné and Kawahara (1999).

This larger issue aside, for the test signals used here it is clear—for both listeners and the model—that vowel identity is conveyed quite well at both low and moderately high F_0, in spite of the fact that the template shape is much more sparsely sampled at higher fundamental frequencies. For the simple city-block distance method used by our model, the similarity that is measured between an input spectrum and a template will be dominated by spectral differences that are clearly irrelevant to vowel identity; that is, most of these channel-by-channel differences do not correspond to the voice-source harmonics which effectively sample the idealized envelope shape. Further, this problem will clearly be exacerbated at higher fundamental frequencies. A guiding assumption of the model is that these irrelevant differences will represent a more-or-less constant source of noise when comparing the input spectrum to all templates, meaning that the *variation* in the distance measure from one template to the next will be controlled mainly by spectral distances at the harmonics—this despite the fact that, unlike the *missing data model*, harmonics receive no special treatment. The relatively high classification accuracy by the model across a ~1-octave range of average fundamental frequencies and ~2-octave range of fundamental frequencies across individual tokens suggests that this assumption is valid, at least for this range of fundamental frequencies. Further, the fact that harmonics are not isolated for special treatment should in principle mean that the method ought to work without modification in the classification of whispered, breathy, or otherwise marginally periodic vowels. This remains to be tested.

While the narrow band model makes no use of fundamental frequency when computing token-to-template distance, pitch does, in fact, figure into the recognition process indirectly as a result of the strategy of using separate templates for signals spoken by men, women, and children. In our view, this method is defensible since the approach is consistent with a large body of psychological evidence suggesting that listener judgments of vowel timbre are, in fact, affected in an orderly way by F_0. For example, a large number of studies have demonstrated that when F_0 is increased, vowel quality can be maintained only by increasing formant frequencies (e.g., Potter and Steinberg, 1950; Miller, 1953; Fujisaki and Kawashima, 1968; Slawson, 1967; Carlson *et al.*, 1975; Ainsworth, 1975; Nearey, 1989). This finding implies that different standards are employed in evaluating the spectrum envelope at different fundamental frequencies. Further, formant-based modeling studies have shown that vowels can be classified with greater accuracy when F_0 is included among the classification features (e.g., Assmann *et al.*, 1982; Hirahara and Kato, 1992; Hillenbrand and Gayvert, 1993). Finally, a recent study by Scott *et al.* (2001) presented listeners with synthetic utterances generated by a source-filter vocoder in which the spectrum envelopes, the fundamental frequencies, or both the spectrum envelopes and the fundamental frequencies were shifted upward in frequency by varying amounts. The authors reported that high intelligibility could be maintained for shifted envelopes, but only if F_0 was also shifted up in frequency. Conversely, upward shifts in F_0 degraded intelligibility unless the F_0 shifts were accompanied by upward shifts in the spectrum envelope. Scott *et al.* argued that their results were best explained by assuming that listener judgments of phonetic quality were influenced by learned associations between F_0 and the spectrum envelope. Exactly how this kind of F_0-dependency is realized in the human system is unclear, and it is unlikely that it is anything as mechanical as the three-template method that we adopted. However, we would argue that our approach is broadly compatible with well-established findings on the interaction between F_0 and the spectral envelope in the perception of vowel quality.

There are several aspects of the labeling behavior of the narrow band model which appear to match well-known characteristics of vowel classification by human listeners. For example, implementations of the model which incorporated spectral change classified vowels with substantially greater accuracy than single-slice implementations. This finding is consistent with a substantial body of evidence implicating a role for spectral change (e.g., Strange *et al.*, 1983; Nearey and Assmann, 1986; Jenkins *et al.*, 1983; Parker and Diehl, 1984; Andruski and Nearey, 1992; Jenkins and Strange, 1999; Hillenbrand and Gayvert, 1993; Hillenbrand and Nearey, 1999; Assmann and Katz, 2000, 2001). The single-slice recognition rates produced by the model vary from about 70% to 80%, depending on the location of the slice. These figures are quite similar to recognition rates reported in several studies in which human listeners were asked to identify either synthetic or naturally spoken vowels with static spectral patterns. For example, Hillenbrand and Gayvert (1993) reported a 74.8% identification rate for 300-ms steady-state synthetic vowels that were synthesized using the F_0 and formant measurements from the 1520-stimulus vowel database recorded by Peterson and Barney (1952). Similarly, Hillenbrand and Nearey reported a 73.8% identification rate for flat-formant resynthesized versions of 300 /hVd/ utterances drawn from the Hillenbrand *et al.* (1995) database (see also Assmann and Katz, 2000, 2001). Finally, Fairbanks and Grubb (1961) reported a 72.1% identification rate for naturally spoken static vowels, demonstrating that the relatively low intelligibility of static vowels is not an artifact of the synthesis methods used in Hillenbrand–Gayvert, Hillenbrand–Nearey, and Assmann–Katz studies.

Implementations of the model incorporating multiple slices of the spectrum produced substantially greater recognition accuracy than the static model, with average recognition rates of 90.6%, 91.6%, and 91.4% for two-, three-, and five-slice models, respectively. It is of some interest to note

that two samples of the spectrum work nearly as well as three or five, suggesting that it is the gross trajectory of spectral movement that is critical and not the fine details of spectral change. A similar conclusion was reached in our earlier modeling study (Hillenbrand *et al.*, 1995) using quadratic discriminant analysis for classification and formants (with or without F_0) as features. The results showed nearly identical classification rates for a discriminant classifier trained and tested on three samples of the formant pattern (20%–50%–80% of vowel duration) as compared to a two-sample model (20%–80% of vowel duration). Both sets of findings are consistent with the dual target model of vowel recognition proposed by Nearey and Assmann (1986). The present results, of course, are based on vowels in a fixed /hVd/ environment. It remains to be determined whether a simple two-sample specification of spectral movement is adequate for more complex utterances involving variation in the phonetic environment surrounding the vowel.

Beyond the straightforward calculation of a high resolution Fourier spectrum, the key signal processing operations in the narrow band model are quite simple, consisting of just two steps: (1) a SLN operation designed to flatten the spectrum and reduce the importance of formant amplitude differences, and (2) a thresholding operation designed to emphasize information-bearing spectral peaks at the expense of perceptually less relevant spectral regions in the narrow valleys between harmonics, and in the broad valleys between formant peaks. One of the more striking aspects of the findings was the dramatic effect that the thresholding operation had on the performance of the recognizer, with recognition accuracy falling from near-human levels of ~90%–92% to about 60% with the removal of the thresholding operation. Recall that the simple city-block method that is used to measure token-template distance treats the information-bearing harmonics in the same way as the perceptually irrelevant nonharmonic spectral values. In all cases, the sum of the token-template differences for nonharmonic spectral values will greatly exceed the sum of the distances for the perceptually relevant spectral values at harmonic frequencies for the simple reason that the great majority of the 256 spectral values do not correspond to harmonics. As noted above, the simplified similarity measure relies on the assumption that the irrelevant spectral distances will represent a more-or-less constant source of noise across all templates, thereby allowing differences at harmonic frequencies to account for most of the variation in the distance measure across templates. The dramatic improvement in recognition accuracy attributable to the thresholding operation indicates that this assumption is valid only if steps are taken to emphasize harmonics at the expense of nonharmonic values, and formants over the valleys between formants. A similar, independently developed peak-enhancement method developed by Liénard and Di Benedetto (2000) has been used with some success to recognize French vowels from smoothed (as opposed to narrow band) ''bump vectors''—smooth spectra resulting from a thresholding operation similar to the one described here. In keeping with the present findings, recognition experiments showed a substantial advantage for the bump vector over a variety of alternative smoothed spectral representations that did not incorporate a thresholding operation (see also Aikawa *et al.*, 1993).

In sharp contrast to the near-essential status of the thresholding operation, the influence of the SLN operation on recognition performance was marginal at best. It is not entirely clear what should be concluded from this. There are several reasons why the SLN operation may have had little or no impact on the performance of the recognition model. First, as we indicated above, it is possible that the test signals, which were all recorded at comfortable vocal efforts and with the same recording equipment, were too well behaved to allow the SLN to have much effect. It is possible that the SLN operation would have had a more substantial effect if factors such as vocal effort (e.g., Liénard and Di Benedetto, 2000) or channel characteristics had varied across stimuli, or if steps had been taken to deliberately alter natural spectral-tilt or formant-level characteristics, as in Klatt's (1982a) study. However, there is some reason to believe that part of the problem is that we have simply not found the right way to implement the SLN operation. Recall that, along with the many similarities in vowel classification between listeners and the narrow band model, there was one aspect of model classification that did not resemble human listener behavior. The model showed some tendency to confuse high-front vowels such as /i/ and /ɪ/ with high back vowels such as /u/ and /U/. While this type of error was not common in the model output, it is almost never observed in human listeners. Examination of individual signals showed that these confusions tended to occur in cases in which the formant pattern of the input spectrum matched the correct template reasonably well, but differences in formant amplitude resulted in the incorrect template producing the smallest token-template distance. For example, there were tokens of /u/ with weak second formants, resulting in good matches to the incorrect /i/ template. This is exactly the situation for which the SLN operation was designed, and in examining individual cases we found that the effect of the SLN was in the right direction, but the change in amplitude relations was not large enough. We have found that simple modifications to the SLN function that are designed to produce a greater degree of flattening across the spectrum (e.g., using the inverse of the *squared* running average as the gain function) can virtually eliminate front–back confusions, but in rather limited experimentation we have not yet found a method that is free of undesirable side effects (e.g., increasing the number of confusions among back vowels). An appropriate solution to this problem may well require transformation to a nonlinear frequency scale, such as one based on the critical band or relationships between characteristic frequency and distance along the basilar membrane.[4]

As we indicated above, we regard the present findings as an existence proof, demonstrating that vowels can be recognized with a high degree of accuracy based on a distance metric which directly measures the similarity between an unsmoothed harmonic spectrum and a set of empirically derived smoothed vowel templates. The present work is, of course, a first step and the findings are necessarily preliminary. To cite just a few examples, it is presently unclear whether the method will provide satisfying results (1) with

stimuli representing a variety of syllable types as opposed to the constant /hVd/ stimuli used here (but see Hillenbrand *et al.*, 2000b, for promising results using formant-based modeling and a wide variety of consonant environments), (2) with whispered or breathy signals, (3) with stimuli such as those used by Klatt (1982a), which are designed to provide a systematic test of sensitivity to spectral shape details such as overall spectral tilt, spectral notches, formant bandwidth, and formant amplitude relationships, and (4) a range of fundamental frequencies exceeding the roughly 2-octave range represented by individual tokens in the present database. These and related questions are among the goals of future work with this method.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health Grant No. R01-DC01661 to Western Michigan University. We are grateful to Michael Clark for comments on earlier drafts.

[1]Here and elsewhere *Gaussian-weighted running average* refers to an approximation implemented with three passes of a rectangular (i.e., unweighted) running average. In this smoothing operation, each spectral amplitude is replaced by the weighted average of *n* neighbors of higher and lower frequency, with the *n* being determined by the width of the smoothing window. Greater weight is assigned to spectral values at the center of the averaging window than to values nearer to the edge of the window. In a true Gaussian-weighted average, the distribution of weights follows a Gaussian function. A simple-to-implement, close approximation to a Gaussian-weighted average can be achieved by running three passes of a rectangular average; i.e., the output of an initial running average operation becomes the input to a second running average, whose output in turn becomes the input to a third running average. A simple end-correction scheme is used in which the averaging window size is initially set to 1 point at either the left or right edge, and the window size is successively expanded until the running average has shifted far enough so that *n* points are available. The smoothing window sizes here and elsewhere refer to the width of the individual rectangular windows used in each of the three averaging passes.

[2]The thresholding operation that is being used here is closely related to simultaneous masking, that is, the tendency of highly active neurons in one frequency region to inhibit or suppress less active neurons in neighboring regions. The degree to which region *a* will mask region *b* depends on (1) the average level of activity in region *a* in relation to region *b*, and (2) the distance (i.e., difference in frequency) between the two regions. Both of these features are captured by a center-weighted average (implemented here with a Gaussian-weighted running average); i.e., the masking function clearly reflects average amplitude, but the masking influence exerted by a particular frequency region falls off systematically with increasing distance. In light of the *phonetically based* modeling goals of this work, no attempt was made to accurately simulate physiologically or psychophysically observed masking phenomena. As noted in Sec. II A, a physiologically accurate simulation of masking would of necessity retain spectral features that contribute to the detailed timbre percept of a stimulus but which may have little or no bearing on vowel identity. Consequently, the thresholding process used here was designed with the goal of maximizing the spectral similarity of vowels that are labeled identically by listeners. It is, in fact, for this reason that we have adopted the somewhat awkward term *thresholding* instead of the term *masking* that we used in some of our other writings describing this type of process (e.g., Hillenbrand and Houde, 2002).

[3]It might be argued that the lower recognition rate for the more poorly identified tokens is due to the simple and uninteresting fact that these tokens were not used to create the templates. We tested this possibility by repeating the tests described above, but with templates created from all 1668 tokens. As with the earlier tests, the model showed a much higher error rate for the 192 signals that were more poorly identified by human listeners (25.0%) than the signals that were well identified by listeners (6.6%).

[4]Limited experimentation with a theoretically reasonable basilar membrane distance scale based on Greenwood (1990) produced disappointing recognition results. Given the limited time we have devoted to this effort to date, we do not conclude much of anything from these results. However, we think it is quite possible that determining how much weight should be assigned to spectral differences in different frequency regions will involve more than simply transforming to a tonotopic scale. For example, the relative importance that is assigned to the F_1 and F_2 regions is likely to depend on the usefulness of these two regions in distinguishing one vowel from another rather than simply the relative amounts of space that are occupied along the basilar membrane.

Aikawa, K., Singer, H., Kawahara, H., and Tohkura, Y. (**1993**). ''A dynamic cepstrum incorporating time-frequency masking and its application to continuous speech recognition,'' ICASSP-93, 668–671.

Ainsworth, S. (**1975**). ''Intrinsic and extrinsic factors in vowel judgments,'' in Auditory Analysis and the Perception of Speech, edited by G. Fant and M. A. A. Tathum (Academic, London).

Andruski, J., 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., and Summerfield, Q. (**1989**). ''Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency,'' J. Acoust. Soc. Am. **85**, 327–338.

Assmann, P., and Katz, W. (**2000**). ''Time-varying spectral change in the vowels of children and adults,'' J. Acoust. Soc. Am. **108**, 1856–1866.

Assmann, P., and Katz, W. (**2001**). ''Effects of synthesis fidelity on vowel identification: Role of spectral change and voicing source,'' J. Acoust. Soc. Am. **110**, 2658(A).

Assmann, P., Nearey, T., and Hogan, J. (**1982**). ''Vowel identification: orthographic, perceptual, and acoustic aspects,'' J. Acoust. Soc. Am. **71**, 975–989.

Bladon, A. (**1982**). ''Arguments against formants in the auditory representation of speech,'' in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier Biomedical, Amsterdam), pp. 95–102.

Bladon, A., and Lindblom, B. (**1981**). ''Modeling the judgment of vowel quality differences,'' J. Acoust. Soc. Am. **69**, 1414–1422.

Carlson, R., Fant, G., and Granstrom, B. G. (**1975**). ''Two-formant models, pitch, and vowel perception,'' in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tathum (Academic, London), pp. 55–82.

Chistovich, L. A., and Lublinskaya, V. V. (**1979**). ''The 'center of gravity' effect in vowel spectra and critical distance between formants: Psychoacoustical study of the perception of vowel-like stimuli,'' Hear. Res. **1**, 185–195.

de Cheveigné, A., and Kawahara, H. (**1999**). ''A missing data model of vowel identification,'' J. Acoust. Soc. Am. **105**, 3497–3508.

Disner, S. F. (**1980**). ''Evaluation of vowel normalization procedures,'' J. Acoust. Soc. Am. **76**, 253–261.

Fairbanks, G., and Grubb, P. (**1961**). ''A psychophysical investigation of vowel formants,'' J. Speech Hear. Res. **4**, 203–219.

Fujisaki, H., and Kawashima, T. (**1968**). ''The roles of pitch and higher formants in the perception of vowels,'' IEEE Trans. Audio Electroacoust. **AU-16**, 73–77.

Greenwood, D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Hillenbrand, J., and Gayvert, R. T. (**1993**). ''Vowel classification based on fundamental frequency and formant frequencies,'' J. Speech Hear. Res. **36**, 694–700.

Hillenbrand, J. M., and Nearey, T. N. (**1999**). ''Identification of resynthesized /hVd/ syllables: Effects of formant contour,'' J. Acoust. Soc. Am. **105**, 3509–3523.

Hillenbrand, J. M., and Houde, R. A. (**2002**). ''Speech synthesis using damped sinusoids,'' J. Speech Hear. Res. **45**, 639–650.

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

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2000b**). ''Effects of consonant environment on vowel formant patterns,'' J. Acoust. Soc. Am. **109**, 748–763.

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.

Hirahara, T., and Kato, H. (**1992**). ''The effect of *F*0 on vowel identification,'' in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 89–112.

Ito, M., Tsuchida, J., and Yano, M. (**2001**). ''On the effectiveness of whole spectral shape for vowel perception,'' J. Acoust. Soc. Am. **110**, 1141–1149.

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.

Kent, R. D. (**1976**). ''Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies,'' J. Speech Hear. Res. **19**, 421–447.

Klatt, D. H. (**1982a**). ''Prediction of perceived phonetic distance from critical-band spectra: A first step,'' IEEE ICASSP, 1278–1281.

Klatt, D. H. (**1982b**). ''Speech processing strategies based on auditory models,'' in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier Biomedical, Amsterdam), pp. 181–196.

Liénard, J-S., and Di Benedetto, M-G. (**2000**). ''Extracting vowel characteristics from smoothed spectra,'' J. Acoust. Soc. Am. Suppl. 1 **108**, 2602(A).

Miller, G. A. (**1956**). ''The perception of speech,'' in *For Roman Jakobsen: Essays on the Occasion of his Sixtieth Birthday*, edited by M. Halle ('s-Gravenhage, Mouton, The Netherlands), pp. 353–359.

Miller, J. D. (**1989**). ''Auditory-perceptual interpretation of the vowel,'' J. Acoust. Soc. Am. **85**, 2114–2134.

Miller, R. L. (**1953**). ''Auditory tests with synthetic vowels,'' J. Acoust. Soc. Am. **18**, 114–121.

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, AB), 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.

Nearey, T. M., Hogan, J., and Rozsypal, A. (**1979**). ''Speech signals, cues and features,'' in *Perspectives in Experimental Linguistics*, edited by G. Prideaux (Benjamin, Amsterdam), pp. 73–96.

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., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Pols, L. C. W., van der Kamp, L. J., and Plomp, R. (**1969**). ''Perceptual and physical space of vowel sounds,'' J. Acoust. Soc. Am. **46**, 458–467.

Potter, R. K., and Steinberg, J. C. (**1950**). ''Toward the specification of speech,'' J. Acoust. Soc. Am. **22**, 807–820.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. E. (**1981**). ''Speech perception without traditional speech cues,'' Science **212**, 947–950.

Rosner, B. S., and Pickering, J. B. (**1994**). *Vowel Perception and Production* (Oxford U.P., Oxford).

Ryalls, J. H., and Liberman, A. M. (**1982**). ''Fundamental frequency and vowel perception,'' J. Acoust. Soc. Am. **72**, 1631–1634.

Scott, J. M., Assmann, P. F., and Nearey, T. N. (**2001**). ''Intelligibility of frequency-shifted speech,'' J. Acoust. Soc. Am. **109**, 2316(A).

Slawson, A. W. (**1967**). ''Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency,'' J. Acoust. Soc. Am. **43**, 87–101.

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.

Sundberg, J., and Gauffin, J. (**1982**). ''Amplitude of the fundamental and the intelligibility of super pitch sung vowels,'' in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier Biomedical, Amsterdam), pp. 223–238.

Syrdal, A. K., and Gopal, H. S. (**1986**). ''A perceptual model of vowel recognition based on the auditory representation of American English vowels,'' J. Acoust. Soc. Am. **79**, 1086–1100.

Zahorian, S., and Jagharghi, A. (**1986**). ''Matching of 'physical' and 'perceptual' spaces for vowels,'' J. Acoust. Soc. Am. Suppl. 1 **79**, S8.

Zahorian, S., and Jagharghi, A. (**1993**). ''Spectral shape features versus formants as acoustic correlates for vowels,'' J. Acoust. Soc. Am. **94**, 1966–1982.

Evaluating the function of phonetic perceptual phenomena within speech recognition: An examination of the perception of /d/–/t/ by adult cochlear implant users

Paul Iverson[a)]
Department of Phonetics and Linguistics, University College London, London NW1 2HE, England and Department of Otolaryngology—Head and Neck Surgery, University of Iowa Hospitals and Clinics, Iowa City, Iowa 52242-1078

(Received 5 July 2002; revised 28 September 2002; accepted 4 November 2002)

This study examined whether cochlear implant users must perceive differences along phonetic continua in the same way as do normal hearing listeners (i.e., sharp identification functions, poor within-category sensitivity, high between-category sensitivity) in order to recognize speech accurately. Adult postlingually deafened cochlear implant users, who were heterogeneous in terms of their implants and processing strategies, were tested on two phonetic perception tasks using a synthetic /dɑ/–/tɑ/ continuum (phoneme identification and discrimination) and two speech recognition tasks using natural recordings from ten talkers (open-set word recognition and forced-choice /d/–/t/ recognition). Cochlear implant users tended to have identification boundaries and sensitivity peaks at voice onset times (VOT) that were longer than found for normal-hearing individuals. Sensitivity peak locations were significantly correlated with individual differences in cochlear implant performance; individuals who had a /d/–/t/ sensitivity peak near normal-hearing peak locations were most accurate at recognizing natural recordings of words and syllables. However, speech recognition was not strongly related to identification boundary locations or to overall levels of discrimination performance. The results suggest that perceptual sensitivity affects speech recognition accuracy, but that many cochlear implant users are able to accurately recognize speech without having typical normal-hearing patterns of phonetic perception. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1531985]

PACS numbers: 43.71.Es, 43.71.Ky [KG]

I. INTRODUCTION

The ability of individuals to recognize speech via cochlear implants calls for a reconsideration of what types of phonetic information and perceptual processing are necessary for human speech recognition. Cochlear implants bypass much of the auditory periphery, such that the neural firing patterns resulting from cochlear implant stimulation differ from normal neural firing patterns (e.g., Rubenstein *et al.*, 1999). The functional number of frequency channels is fewer than for normal hearing (e.g., Dorman *et al.*, 2000; Fishman *et al.*, 1997), but temporal resolution can be about the same (e.g., Busby *et al.*, 1993; Shannon, 1989, 1992; see Shannon, 1993 for a review). Despite the facts that cochlear implant stimulation is quite different from normal hearing in many respects, and that the standard frequency-related phonetic cues (e.g., formant and burst frequencies) may be difficult to discern given the poor spectral resolution of cochlear implants (e.g., Dorman, 1991; Shannon *et al.*, 1995), the best postlingually deafened users of current cochlear implants are able to recognize more than 90% words correct in clinical tests of open-set sentence recognition (e.g., Parkinson *et al.*, 1998).

The aim of the present study was to determine whether cochlear implant users must perceive differences along phonetic continua in the same way as do normal hearing listeners (e.g., having sharp identification boundaries, low within-category sensitivity, and high between-category sensitivity; Liberman *et al.*, 1957; Studdert-Kennedy *et al.*, 1970; Repp, 1984) in order to recognize speech accurately. Dorman and colleagues (Dorman *et al.*, 1991) found that, among a group of six Symbion cochlear implant users who had above-average word recognition accuracy, four had phoneme labeling functions for a synthetic voice-onset-time (VOT) continuum that were like those of normal-hearing individuals. It is thus clear that at least a subset of cochlear implant users is similar to normal-hearing individuals, but it is unlikely that all cochlear implant users perceive phonetic differences in this way (cf. Hedrick and Carney, 1997), particularly given their large range of individual differences in speech recognition (e.g., Parkinson *et al.*, 1998). It is unknown whether the cochlear implant users with normal phoneme identification functions are more accurate at recognizing speech, or whether it is possible for individuals to recognize speech accurately despite having unusual patterns of phonetic perception.

In the normal-hearing speech recognition literature, current evidence contradicts the early interpretation of categorical perception, that speech is perceived in terms of phoneme labels (e.g., Liberman *et al.*, 1967). For example, fine-grained phonetic variation (e.g., variation due to differences between talkers) has been shown to affect speech recognition accuracy and to be stored in memory (e.g., Goldinger, 1996; Nygaard and Pisoni, 1998; Pisoni, 1997); listeners have been

[a)]Electronic mail: paul@phon.ucl.ac.uk

shown to perceive differences in goodness among stimuli that are categorized the same (e.g., Allen and Miller, 2001; Iverson and Kuhl, 1995, 1996, 2000; Miller, 1994); and current cognitive models of word recognition have been able to account for experimental data without including a phoneme categorization stage (e.g., Connine *et al.*, 1994; Luce and Pisoni, 1998; Norris *et al.*, 2000). Despite the fact that phoneme encoding may not occur, the perceptual phenomena associated with the categorical perception of consonants (e.g., sensitivity peaks near identification boundaries) have remained robust and ubiquitous in the literature. There has been little direct evidence, however, to indicate what role these perceptual phenomena have in speech recognition.

It is difficult to address this issue by testing normal-hearing individuals listening to their native language, because few individuals have unusual patterns of phonetic perception but are normal in other respects. However, language experience can produce these types of individual differences, and cross-language studies have consistently linked individual differences in phonetic perception and word recognition. For example, Japanese adults who have difficulty identifying synthetic /r/–/l/ syllables also have difficulty recognizing words with those phonemes (Yamada, 1995), and non-native speakers of English have a marked difficulty recognizing English words that require more phonetic information to be distinguished from lexical competitors (Bradlow and Pisoni, 1999). One drawback of cross-language research is that the origin of these speech recognition difficulties cannot be definitively isolated to any one level, because language experience affects many levels of neural processing simultaneously. The present study examined postlingually deafened adults with cochlear implants, because their speech recognition difficulties have a clearer sensory origin, and it can be assumed that these individuals have normal native-language linguistic processing due to their hearing during childhood.

The experiments measured phonetic perception (identification and discrimination) along a /dɑ/–/tɑ/ synthetic continuum, and speech recognition (open-set word recognition and forced-choice phoneme identification) for recordings of natural speech from multiple talkers. Phonetic perception experiments have traditionally used fixed-interval designs (e.g., 10-ms VOT differences between all pairs of stimuli in a discrimination task). Such designs are likely inappropriate for cochlear implant users given their wide range of individual differences (i.e., for the same interval sizes, better subjects would reach ceiling performance in discrimination tasks and poorer subjects would be at chance). Instead, the present experiments used adaptive procedures (Levitt, 1971) to adjust the interval size for individual subjects. Measures of phonetic perception along the /dɑ/–/tɑ/ synthetic continuum were compared to those of normal-hearing individuals, to assess whether the normality of phonetic perception for these stimuli is predictive of individual differences in speech recognition performance.

II. METHOD

A. Subjects

Twenty-five postlingually deafened cochlear implant users were tested. The subjects were not selected based on implant type or processing strategy, to increase the potential individual differences among subjects; eight used the Clarion implant with a CIS processing strategy, one used an Ineraid implant with a Med-El processor and a CIS processing strategy, six used a Nucleus-22 implant with a SPEAK processing strategy, four used a Nucleus-24 implant with an ACE processing strategy, five used a Nucleus-24 implant with a SPEAK processing strategy, and one had binaural Nucleus-24 implants, one with SPEAK and the other with ACE. The age of the subjects had a range of 40.8–80.3 years, with a mean of 58.6 years. Their duration of implant use had a range of 0.5–12.2 years with a mean of 5.9 years. Fourteen cochlear implant subjects were male and 11 were female. All were native speakers of American English.

Fourteen normal-hearing subjects were tested to provide comparison data on the phonetic perception tasks. Two subjects were dropped from this study because of unusual data; one subject had no clear sensitivity peak in the discrimination task (discrimination was accurate in a broad region near the identification boundary), and the other had levels of discrimination performance that were more than 2 standard deviations poorer than the average. These unusual data were omitted because they were not consistent with the aim of estimating typical normal-hearing performance. The age of the 12 remaining normal-hearing subjects had a range of 21.1–56.0 years, with a mean of 33.3 years. Four of these subjects were male and eight were female. All were native speakers of American English.

B. Apparatus

The subjects were tested in a double-walled booth. The stimuli were presented at a comfortable level via a computer sound card connected to two loudspeakers, positioned to the front-left and front-right of the subjects. Subjects entered their responses by clicking on buttons displayed on a computer screen, using a computer mouse. One subject was blind, and used a modified testing interface that collected responses via a button box.

C. Stimuli

1. Natural recordings

A list of 120 monosyllabic words and 80 /dɑ/ and /tɑ/ syllables was recorded by ten adult native speakers of American English who lived in Iowa. Five talkers were male and five were female. The word corpus comprised 20 /d/–/t/ minimal pairs (i.e., 40 words) with the target phonemes in syllable-initial position (*target-initial words*), 20 /d/–/t/ minimal pairs with the target phonemes in syllable-final position (*target-final words*), and 40 words that did not contain either /t/ or /d/ and were randomly selected from The Celex Lexical Database (1995; *nontarget words*). Minimal pairs were used for the target words so that the lexicon could not be used to distinguish /d/ and /t/ during the word recognition experi-

ment. The nontarget words were included in the corpus so that responses in the word recognition experiment would be less likely to be biased toward words containing /d/ or /t/.

During recording, the words and syllables were displayed one at a time on a computer screen, in a random order. The words were recorded using 16-bit samples and a 44.1-kHz sampling rate.

The recordings were screened for intelligibility and recording quality. The amplitude of each recording was scaled to make all recordings equal in rms amplitude. The final word corpus was selected to include 4 target-initial words (2 /d/ and 2 /t/), 4 target-final words (2 /d/ and 2 /t/), and 4 nontarget words from each of the ten talkers. Each of the 120 words occurred once in the final corpus. The final syllable corpus included 4 /dɑ/ and 4 /tɑ/ syllables from each of the ten talkers.

VOT was measured for the initial-target phonemes, quantified here as the latency between burst onset and the onset of voicing energy in the $F2$ range (i.e., onset of regular voicing). In words, the /d/ phonemes had an average VOT of 27 ms and a range of 10–51 ms, excluding two prevoiced stimuli; the /t/ phonemes had an average VOT of 104 ms and a range of 56–148 ms. In syllables, the /d/ phonemes had an average VOT of 23 ms and a range of 13–40 ms, excluding one prevoiced stimulus; the /t/ phonemes had an average VOT of 94 ms and a range of 48–136 ms.

2. Synthetic continuum

The stimulus continuum was created using the Klatt synthesizer controlled by higher-level articulatory parameters within the HLSYN computer program (1997; Stevens and Bickley, 1991). The synthesis parameters (e.g., formant frequencies and fundamental frequency contour) were modeled from recordings of /dɑ/ and /tɑ/ by a male speaker. The duration of voicing was 350 ms for every stimulus (i.e., the aspirated portion of each stimulus was added to the total stimulus length, rather than subtracted from the duration of the voiced portion). The formant frequencies for $F1-F4$ at the consonant release were 200, 1762, 2889, and 2972 Hz. The frequencies of $F1-F4$ at the vowel target were 781, 1501, 2532, and 3029 Hz. $F0$ fell from 120 to 80 Hz during the voiced portion of the stimuli. An articulatory parameter representing the cross-sectional area of a constriction formed at the tongue blade (ab) was set to 0 mm^2 during the consonant closure and reached 100 mm^2 (i.e., no constriction) 10 ms after the release of the closure.

VOT ranged from 0–150 ms, with a step size of 1 ms (i.e., a total of 151 stimuli). This variation in VOT was created by manipulating an articulatory parameter for the area of glottal opening (ag), relative to the release of the consonant closure (i.e., the start of the transition of ab from 0 to 100 mm^2). For example, a stimulus with a 0-ms VOT had modal voicing ($ag=5$ mm^2) beginning at the same time as the closure release. A stimulus with a 100-ms VOT had aspiration at the closure release ($ag=30$ mm^2) and modal voicing ($ag=5$ mm^2) 100 ms after the closure release. As a consequence of manipulating these articulatory parameters (i.e., ag and ab), multiple acoustic cues, such as the latency between the burst and voicing, the burst amplitude, and the $F1$ onset, were all varied according to HLsyn's (1997) articulatory model. The acoustic cues for VOT thus were designed to vary naturally along the stimulus continuum, and they were not directly controlled to equate acoustic differences among stimuli.

D. Procedure

The four experimental tasks were run in a single session for each subject, in the order listed below. Subjects were allowed to take breaks between experimental tasks.

1. Open-set word recognition

Subjects heard one word on each trial and identified what they thought they heard. Subjects were given the option to either type their response into the computer or tell the experimenter the word that they heard. Subjects were instructed that all of the stimuli would be real monosyllabic words, and that they needed to type their best guess for the word even if they were not certain. Subjects were not told that the word corpus had a high percentage of words containing /t/ or /d/. Moreover, this was the first condition that was run for each subject, and subjects had yet to be told that the later conditions would involve /t/ and /d/ identification. Postexperiment comments by the subjects suggested that they were unaware that there were a large number of /t/ and /d/ words in the corpus, although some subjects noticed that some of the words rhymed.

Each of the 120 words was presented in an order that was randomized for each subject. There was no practice or feedback.

After the experiment was completed, each response was corrected for spelling and transcribed phonemically. The responses were scored in terms of whether the word response was correct and whether the target phoneme was correct.

2. Phoneme identification: Natural syllables

Subjects heard one syllable on each trial and judged whether it began with /d/ or /t/. Subjects began with a short practice session composed of randomly selected trials with no feedback; the practice was ended as soon as the experimenter and the subject were confident that the subject was comfortable with the task. Subjects then completed an experimental session composed of the full corpus of 80 syllables presented in a random order for each subject.

3. Phoneme identification: Synthetic continuum

As with natural syllables, subjects were presented one stimulus on each trial and judged whether it began with /d/ or /t/. Subjects began with a short practice session composed of randomly selected trials (from 0 to 120-ms VOT) with no feedback. The practice was ended as soon as the experimenter and the subject were confident that the subject was comfortable with the task.

In the experimental session, the trials were set by an interleaved double-staircase adaptive procedure that was designed to find the identification boundary location and width for each subject. Specifically, one-up/two-down Levitt procedures (1971) were used to find two locations along the

stimulus continuum: The point where stimuli were identified as /d/ on 71% of trials (found by the */d/ series* of the adaptive procedure), and the point where stimuli were identified as /t/ on 71% of trials (found by the */t/ series* of the adaptive procedure). The midpoint between these locations was defined as the identification boundary location. The difference between these locations was defined as the identification boundary width.

The adaptive procedure had four stages. In the first stage, the /d/ series began with a 16-ms VOT and the /t/ series began with a 54-ms VOT. The step size was 16 ms, and the first stage was completed after both adaptive series completed three reversals. The second stage had a step size of 8 ms and was finished when both series completed seven reversals. The third stage began by resetting the values of /d/ and /t/ to the average of their reversals in the second stage, had a step size equal to half the difference between the /d/ and /t/ series (estimated from stage 2), and was finished when both series completed 11 reversals. The fourth stage began by resetting the values of /d/ and /t/ to the average of their reversals in the third stage, had a step size equal to half the difference between the /d/ and /t/ series (estimated from stage 3), and was finished after both series completed 15 reversals. The average of the reversals in stages 2–4 were used to calculate the boundary locations.

Half of the presented trials were from neither adaptive series. On these trials subjects were presented a stimulus that was randomly selected from the series, to prevent the responses from being affected in some way by having stimuli concentrated only at the phoneme boundary. The results from these trials were not included in the estimation of boundary locations.

The order of all trials (i.e., /d/ series, /t/ series, and the other trials) was randomized for each subject.

4. Discrimination

Subjects heard three stimuli on each trial with an interstimulus interval of 250 ms. Two stimuli were the same and one was different, and the different stimulus was either the first or last that they heard. Subjects gave a two-alternative forced-choice response to indicate which stimulus, the first or the last, they thought was different.

Discrimination was tested at 14 different anchor points along the synthetic stimulus continuum: The locations of the identification boundary, the best[1] /d/, and the best /t/ for each subject; and at the points 0, 10, 20, 30, 40, 50, 60, 70, 80, 100, and 120-ms VOT. A one-up/two-down Levitt procedure (1971) was used for each anchor point to find the amount of VOT difference between stimuli that was required to perform the task at 71% correct (the 14 adaptive series were run within the same blocks of trials).

The stimuli were centered around each anchor point. For example, if the anchor point was 50 ms and the difference between stimuli was 10 ms, then subjects were tested with 45- and 55-ms stimuli. The stimulus selection was altered when an end of the range (0 or 150 ms) was reached. For example, if an anchor point was 10 ms and the VOT difference was 30 ms, subjects were tested with 0- and 30-ms stimuli.

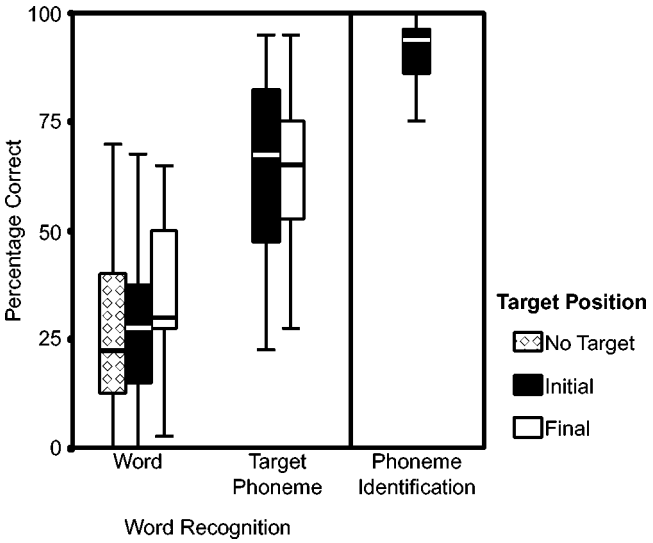


FIG. 1. Boxplots of results for the open-set word recognition and phoneme identification tasks using natural recordings of speech. Boxplots display the interquartile range of scores. The box shows the 25th to 75th percentiles, with a line at the median value. The lower and upper ''whiskers,'' respectively, show the first and last quartiles. Nontarget words were those that did not have either /t/ or /d/. Target-initial words had either /t/ or /d/ in syllable initial position. Target-final words had either /t/ or /d/ in syllable final position.

The adaptive procedure varied VOT multiplicatively. For example, if the step size was 2, the VOT difference between stimuli was doubled after an incorrect response and halved after two correct responses. The VOT difference was limited so that it was never less than 1 ms or greater than 100 ms.

Subjects completed a practice, without feedback, in which they heard a randomly selected anchor point and a randomly selected VOT difference on each trial. The practice was terminated as soon as the subject and the experimenter were confident that the subject understood the task.

The experimental session had seven stages. In stage 1, the VOT differences began at 16 ms, the adaptive step size was 2, and there were two reversals for each anchor point. In stages 2–4, the VOT difference for each anchor point was reset at the beginning of the stage to be equal to the average reversals at that anchor point and at the adjacent anchor points along the series (the adjacent anchor points entered into this calculation because they provided additional information about sensitivity at that general location in the VOT continuum). The adaptive step size was $2^{0.5}$ and there were two reversals at each stage. Stages 5–7 had an adaptive step size of $2^{0.25}$, but were the same as stages 2–4 in all other respects. Subjects were permitted to take a short break between stages.

The difference limen (DL) at each anchor point for each subject was calculated by averaging the two median reversals in stages 2–7. The location of the sensitivity peak (i.e., minimum DL) for each subject was estimated using parabolic interpolation (Press *et al.*, 1992). Specifically, the sensitivity peak location was defined to be the minimum of a parabola that was found by the equation

$$\min = \frac{b - 0.5*\{[b-a]^2*[f(b)-f(c)] - [b-c]^2*[f(b)-f(a)]\}}{[b-a]*[f(b)-f(c)] - [b-c]*[f(b)-f(a)]}, \quad (1)$$

where b was the anchor point with the lowest measured DL, a and c were the anchor points adjacent to b, and $f(a)$, $f(b)$, and $f(c)$ were the DL values at these anchor points. Interpolation was used so that sensitivity peaks were based on the data from three points rather than one, and so that the location estimates had a higher resolution than did the anchor point locations.

III. RESULTS

A. Word recognition and phoneme identification for natural recordings

Figure 1 displays the ranges of word recognition and phoneme identification results for cochlear implant subjects. As is typical of cochlear implant users, there was substantial individual variability in percentage-correct scores for entire words and for target phonemes within words. Their average word recognition accuracy was poor (average of 29%), but this likely was due to the difficulty of this particular word corpus; these subjects had averaged 59% correct for CNC words, in tests conducted during their clinical visits. The percentage-correct scores in the forced-choice syllable identification task approached ceiling levels of performance (i.e., 100%), which reduced the range of scores.

B. Phoneme identification and discrimination: Synthetic stimulus continuum

1. Sensitivity functions

Figure 2 displays *sensitivity functions* (i.e., DL values along the VOT continuum) and identification boundary locations. The normal-hearing subjects were relatively homogeneous. Sensitivity was best (i.e., lowest DL) in the region of 30–40 ms along the stimulus series, near the average normal-hearing category boundary (37 ms). Sensitivity was poorest within phoneme categories.

The sensitivity functions from cochlear implant subjects were highly variable, to an extent that would make the presentation of group sensitivity functions meaningless. Instead, sensitivity functions from three individual subjects are displayed in Fig. 2. Subject 1 is an example of a cochlear implant user who had results that were similar to those of normal-hearing subjects. There was a clear sensitivity peak near the category boundary (at a longer VOT than was found for normal-hearing subjects), and poorer sensitivity within phoneme categories. At the sensitivity peak, the level of sensitivity was within the normal-hearing range. Within phoneme categories, sensitivity was somewhat poorer than was found for normal-hearing subjects.

However, many subjects had data that were markedly different from that of normal-hearing individuals. For example, Subject 2 did not have an identification boundary and a sensitivity peak at the same location. This subject had an identification boundary that was near that found for normal-hearing individuals, but had a sensitivity peak at a lower VOT (14 ms) and perhaps a second sensitivity peak at a higher VOT (80 ms). The subject also had much poorer sensitivity overall compared to normal-hearing subjects, and approached the 100-ms maximum difference at several points along the continuum. This makes the sensitivity peak difficult to interpret, because large DLs should lead to more

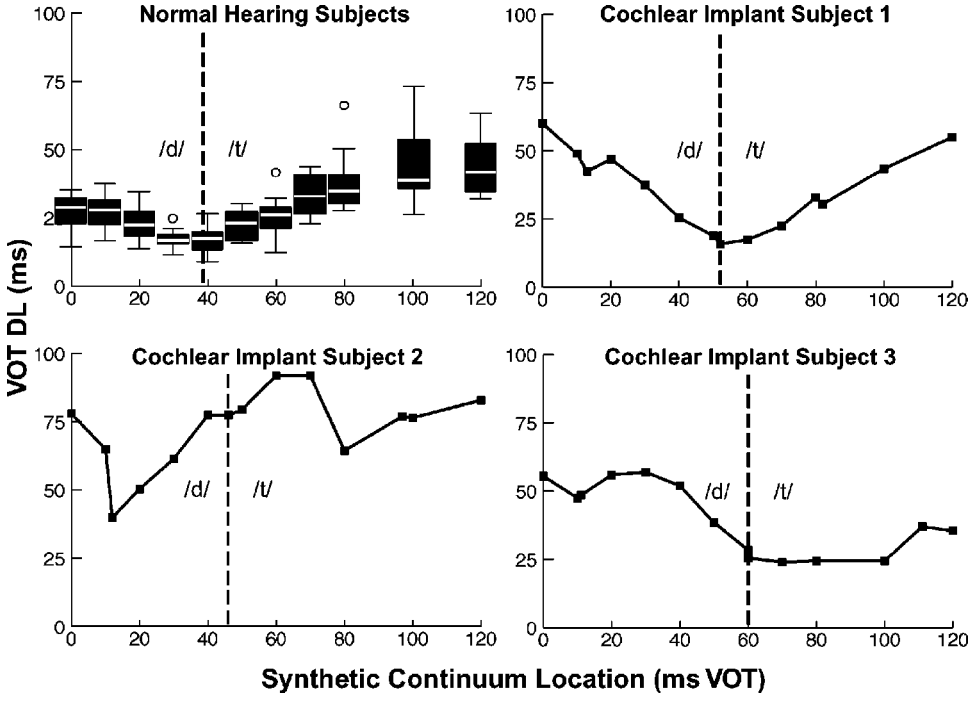


FIG. 2. Boxplots of sensitivity functions for normal-hearing subjects and individual sensitivity functions for three example cochlear implant subjects. Boxplots display the interquartile range of scores, with outliers marked with circles. The vertical dashed lines in each plot indicate the location of the phoneme identification boundary. The normal-hearing subjects were fairly homogeneous and had results consistent with categorical perception (i.e., high sensitivity at the category boundary, low sensitivity within phoneme categories). The data from the cochlear implant subjects were highly variable; there is evidence of categorical perception for Subject 1, but not for Subjects 2 and 3.

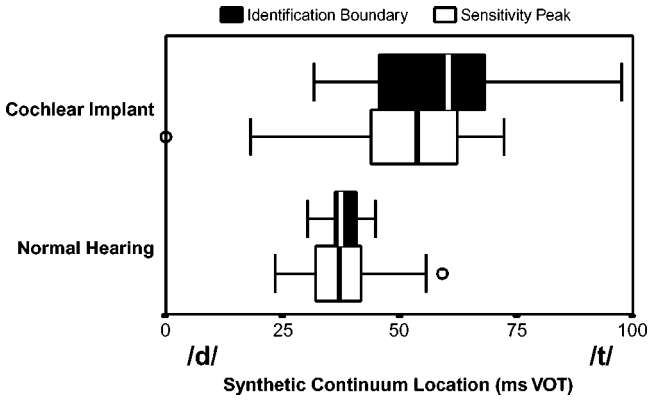


FIG. 3. Boxplots of the locations of identification boundaries and sensitivity peaks along the synthesized stimulus continuum. The distributions of both measures were shifted to longer VOT values for cochlear implant users, compared to those of normal-hearing individuals. Moreover, the individual differences were greater for cochlear implant users than for normal-hearing individuals, on both location measures.

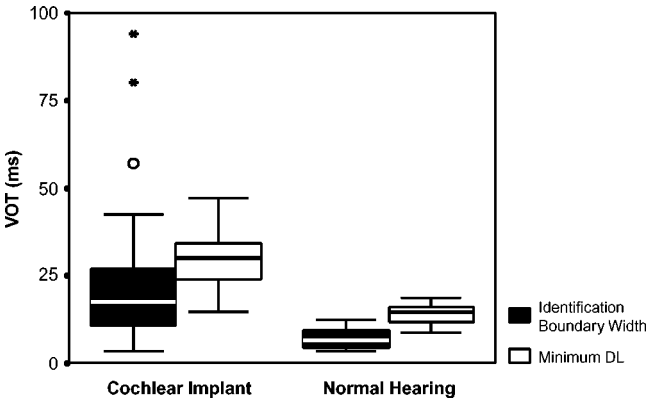


FIG. 4. Boxplots of identification boundary widths and minimum DL values. Although there is overlap between the distributions for individuals with cochlear implants and normal hearing, the cochlear implant users generally had greater widths and DL minima, demonstrating that they were less sensitive to VOT differences near their /d/–/t/ phoneme boundary.

gradual changes along the continuum (because neighboring stimulus pairs have more overlap); this individual had very sharp changes in the sensitivity function near the peak.

Subject 3 is an example of an intermediate case. Sensitivity was poor within the /d/ category and increased near the category boundary, but sensitivity within the /t/ category remained as high as at the category boundary, forming a broad region of high sensitivity rather than a peak. In fact, this individual had higher sensitivity for stimuli within the /t/ category (i.e., 60–120 ms on the continuum) than did any of the normal-hearing subjects.

2. Location measures: Sensitivity peaks and identification boundaries

As displayed in Fig. 3, cochlear implant subjects tended to have sensitivity peaks and identification boundaries at longer VOT values than did normal-hearing subjects. Cochlear implant subjects also had a wider range of VOT locations for both measures, such that some cochlear implant users had identification boundary and sensitivity peak locations that were within, or below, the normal-hearing range.

The correlation between the locations of sensitivity peaks and category boundaries for cochlear implant users was significant, $r=0.49$, $p<0.01$. However, few cochlear implant users had their sensitivity peak and identification boundary at exactly the same location; the difference between the two location measures was as large as 49.9 ms for one subject, and there was a median difference among subjects of 15.3 ms. There appeared to be continuous variation among subjects in the extent to which the locations of sensitivity peaks and identification boundaries differed.

3. Sensitivity measures: Minimum DLs and identification boundary widths

As displayed in Fig. 4, cochlear implant subjects had larger identification boundary widths and larger minimum DL values than did normal-hearing subjects. Although there was some overlap between these distributions, it appears that, as a group, cochlear implant users are less sensitive, compared to normal-hearing individuals, to changes in VOT near identification boundaries and sensitivity peaks. These two measures were correlated for cochlear implant users, $r=0.47$, $p<0.01$.

C. Relationships among experimental measures

Initial analysis of the data suggested an inverted U-shaped relationship between the location measures (identification boundary and sensitivity peak) and speech recognition measures for cochlear implant users. In Fig. 5, for example, there is a significant curvilinear relationship, measured using polynomial regression, between sensitivity peak location and initial target phoneme recognition within words, $R=0.65$, $F(21)=5.16$, $p<0.01$. This shows that subjects who had sensitivity peaks near 45–50 ms along the stimulus series had the highest phoneme recognition accuracy, and accuracy declined for subjects who had sensitivity peaks at longer or shorter VOT values. To allow for simpler linear statistical comparisons with the speech recognition scores, the location measures were recalculated in terms of their distance from the peak location of the inverted U-shaped function (i.e., 47.5 ms). These recalculated mea-

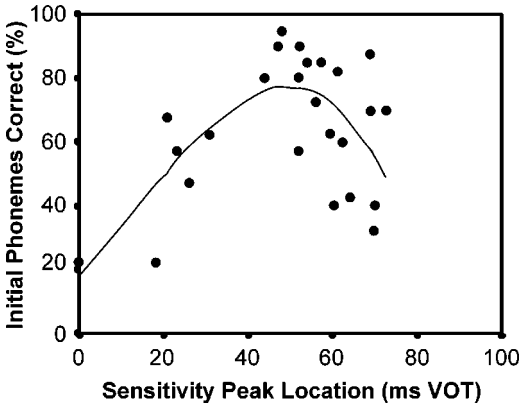


FIG. 5. Scatterplot of the relationship between sensitivity peak locations and the percentage-correct initial target phoneme recognition within words. There was a significant inverted-U-shaped relationship, indicated by the best-fit polynomial regression line, between sensitivity peak location and phoneme recognition; subjects with a sensitivity peak at a 45–50-ms VOT tended to have higher phoneme recognition scores.

TABLE I. Correlations (r) of measures of word recognition and phonetic perception for cochlear implant subjects.

	Words			Phonemes in words		Forced-choice identification
	Nontarget	Initial	Final	Initial	Final	
Optimality of identification boundary location	−0.27	−0.26	−0.21	−0.31	−0.09	−0.24
Optimality of sensitivity peak location	−0.38[a]	−0.47[a]	−0.47[a]	−0.70[a]	−0.45[a]	−0.53[a]
Identification boundary width	−0.32	−0.29	−0.45[a]	−0.28	−0.37[a]	−0.16
Minimum DL	−0.29	−0.16	−0.10	−0.06	−0.18	−0.11

[a]$p<0.05$.

sures thus quantified how far each listeners' identification boundary and sensitivity peak locations were from the optimal location for word recognition.[2]

Pearson correlation coefficients between these phonetic perception and speech recognition measures are displayed in Table I. The results demonstrated that there was a clear and consistent relationship between the optimality of sensitivity peak locations and all speech recognition measures. There was a particularly strong tendency ($r=-0.70$) for individuals with sensitivity peaks near 47.5 ms to correctly recognize initial target phonemes within words, and there was even a tendency ($r=-0.38$) for these individuals to correctly recognize words that did not contain /t/ or /d/. The relationship between identification boundary optimality and the speech recognition measures was weaker; although all correlations were in the expected negative direction, none was significant.

The correlations (Table I) revealed that there was a weak inverse relationship between identification boundary width and the speech recognition measures, reaching significance only for target-final words and phonemes; individuals with a broader phoneme identification boundary tended to have more difficulty recognizing these phonemes within natural speech. In contrast, minimum DL was not significantly correlated with any of the speech recognition measures. It is somewhat surprising that identification boundary width was more strongly related to word recognition than was minimum DL, because both are sensitivity measures (i.e., sharper identification boundaries indicate higher sensitivity, as do smaller DLs). However, the identification boundary width can also be interpreted as an indirect measure of the optimality of the identification boundary location, because identification boundary widths can be expected to be sharper when the identification boundaries and sensitivity peaks are at the same location. The minimum DL is more strongly related to the overall sensitivity to acoustic differences along the series.

To further test the contribution of all of the phonetic measures to word recognition accuracy, an ANCOVA analysis was conducted with non-, initial-, and final-target words coded as a repeated measure. Sensitivity peak location optimality was significant, $F(1,20)=5.348$, $p<0.05$. Identification boundary optimality, $F(1,20)=1.039$, identification boundary width, $F(1,20)=1.535$, and minimum DL, $F(1,20)=0.078$, were not significant. Likewise, an ANCOVA was conducted for the phoneme recognition measures, with syllable identification, initial-target, and final-target coded as a repeated measure. Again, sensitivity peak location optimality was significant, $F(1,20)=13.112$, $p<0.01$. Identification boundary optimality, $F(1,20)=0.671$, identification boundary width, $F(1,20)=1.756$, and minimum DL, $F(1,20)=0.017$, were not significant. Together, these analyses confirm that sensitivity peak location optimality was the best predictor of speech recognition accuracy in this study.

The relationships between phonetic perception measures and speech recognition can be further illustrated by inspecting the example data presented in Fig. 3. Among these three subjects, word recognition performance was related to the shape of their sensitivity function; Subject 1 had high word recognition performance along with a normally shaped sensitivity function (e.g., 67.5% correct initial target words), and Subjects 2 and 3 had poorer word recognition performance (e.g., 30.0 and 32.5% correct initial-target words, respectively). These examples also illustrate why overall levels of sensitivity did not correlate with word recognition performance. Subject 3 had sensitivity levels that surpassed those of normal-hearing individuals within the /t/ category, but this did not lead to exceptional word recognition accuracy. Moreover, Subject 2 had levels of word recognition accuracy that were similar to those of Subject 3, despite Subject 2's much poorer levels of sensitivity.

IV. DISCUSSION

There were two main findings. First, cochlear implant users do not, as a group, perceive phonetic differences along a VOT continuum in the same way as do normal-hearing individuals; cochlear implant subjects tend to have sensitivity peaks and identification boundaries at longer VOT locations, identification boundaries that are less sharp, higher minimum DLs, and more intersubject variability on all of these measures. Second, speech recognition accuracy by cochlear implant users is related to the shape of the phonetic sensitivity function, at least in terms of the location of the peak, but is not strongly related to other aspects of phonetic perception, such as the level of sensitivity at the peak or to the phoneme identification boundary.

From the standpoint of normal-hearing speech perception theories, it is particularly notable that many cochlear implant subjects were able to accurately categorize voicing in natural speech (e.g., median forced-choice /d/–/t/ identification was 94%), despite the fact that their phonetic identi-

fication and sensitivity functions were markedly different from those of normal-hearing individuals. It is clearly not necessary to have normal categorical perception in order to recognize speech accurately. However, it is beneficial to have favorable sensitivity functions. When a listener has a sensitivity peak at an advantageous location (e.g., near normal-hearing identification boundaries), words likely become more distinct perceptually from potential lexical competitors, thereby facilitating recognition. Other types of sensitivity functions likely impair performance by making the perceptual differences between lexical competitors less salient than within-category variation. The patterns of sensitivity measured in phoneme discrimination experiments thus affect recognition accuracy, even though phoneme labeling may have little functional importance.

It was surprising that word recognition accuracy was unrelated to the overall level of sensitivity. Previous cochlear implant research has focused on the levels of spectral (e.g., Dorman *et al.*, 1996) or temporal (e.g., Cazals *et al.*, 1994; Hochmair-Desoyer *et al.*, 1985) resolution available to users (see also Svirsky, 2000). The present results provide a conflicting view; it seems more important for cochlear implant users to have relatively high sensitivity to critical VOT differences than it is for listeners to have high sensitivity to VOT differences throughout the continuum.

This conclusion is limited by the fact that it is based only on VOT data. It is logically necessary that word recognition performance must be affected by sensitivity levels to some extent, because accurate auditory word recognition would be impossible for individuals who were unable to hear any differences between sounds. Cochlear implant users, as a group, may have sensitivity levels for VOT that are above this lower limit, such that increases in phonetic sensitivity do not further improve recognition performance for voicing contrasts (see also Tyler *et al.*, 1989). The levels of sensitivity to VOT could prove important under more difficult listening conditions, such as when speech is combined with noise. Furthermore, the effects of sensitivity level could be stronger for phonetic dimensions that are more dependent on frequency cues (e.g., consonant place or vowel height), given that spectral sensitivity by cochlear implant users is generally poor.

It is unknown what caused the observed shifts in sensitivity peak locations. It would be straightforward to hypothesize that shifts of sensitivity peaks to longer VOTs are a result of temporal processing deficits. That is, normal-hearing research has suggested that VOT boundary locations could be a result of an auditory threshold for detecting the temporal order of a burst and the onset of voicing (e.g., Pastore and Farrington, 1996), so it would be reasonable to predict that individuals with poorer temporal resolution would have a higher threshold for detecting this difference, causing sensitivity peaks to occur at longer VOTs. However, there was no evidence that individuals with sensitivity peak locations at longer VOTs had unusually poor levels of sensitivity. Furthermore, this explanation does not account for why some individuals have sensitivity peaks at shorter-than-normal VOTs.

Sinex and colleagues (Sinex, McDonald, and Mott, 1991; cf. Soli, 1983) have suggested that spectral cues, such as the first formant frequency ($F1$), are more responsible for sensitivity peaks along VOT continua than are temporal cues. The locations of the sensitivity peaks along the continuum could thus have been affected by individual differences in $F1$ perception. For example, listeners may have been more sensitive to $F1$ transition differences when it spanned more than one electrode frequency band, or when it exceeded the low-frequency cutoff of the implant processor. The shifts in sensitivity peak locations along the VOT continuum could therefore have been caused by complex interactions between the characteristics of the cochlear implant processors, electrode locations, and the frequencies of the $F1$ transitions.

It is plausible too that cochlear implant users differ in their use of acoustic cues. Variability in cue weightings has been shown to occur among normal-hearing individuals (Hazan and Rosen, 1991), and the functional importance of these differences may be magnified when the available phonetic information is reduced. For example, individuals who attend to spectral cues for VOT (e.g., $F1$ onset) may have more difficulty discerning voicing through their cochlear implant than do individuals who attend to temporal cues for VOT (e.g., duration of aspiration), which tend to be better represented via cochlear implants. Individual differences in cue weightings may be particularly large for cochlear implant users, arising from changes to speech recognition strategies following prolonged periods of deafness and a subsequent accommodation to electric hearing.

ACKNOWLEDGMENTS

This work was supported by research Grants Nos. 1 R03 DC03999 and 2 P50 CD 00242 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health; and Grant No. RR00059 from the General Clinical Research Centers Program, Division of Research Resources, National Institutes of Health. I am grateful to Gina Hart and Annie Vranesic for their assistance with data collection; to Lynne E. Bernstein for initial comments on the research plan; and to Richard S. Tyler, Andrew Faulkner, and Stuart Rosen for their comments on this manuscript.

[1]In a separate experiment, cochlear implant users were asked to rate the subjective goodness of the synthetic stimuli (see Iverson and Kuhl, 1995, 1996), and the stimuli with the highest goodness ratings for each category were used as anchor points in the discrimination task. The results from the goodness rating task are not discussed further, because of concerns over their reliability. Subjects mostly reported following the test that the stimuli ''all sounded the same'' or that they performed the goodness task on the basis of some idiosyncratic perceptual detail of the stimuli, such as loudness or ''number of overtones.''

[2]Although the optimal location was operationally defined here as 47.5 ms, the average normal-hearing sensitivity peak location—which would have been expected to be optimal—occurred at a shorter VOT (37.7 ms). This discrepancy between potentially optimal locations could be due to not having enough statistical power to determine the exact shape of the relationship between location and recognition accuracy (e.g., there were no subjects who had sensitivity peaks between 33 and 43 ms). A function with a 37.7-ms optimal location could have fit the data nearly as well, if the slope of the function had been fit to be steeper toward the left than to the right.

Allen, J. S., and Miller, J. L. (**2001**). ''Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate,'' Percept. Psychophys. **63**, 798–810.

Bradlow, A. R., and Pisoni, D. B. (**1999**). ''Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors,'' J. Acoust. Soc. Am. **106**, 2074–2085.
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.
Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). ''Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants,'' J. Acoust. Soc. Am. **96**, 2048–2054.
Celex Lexical Database. (**1995**). (Version 2.5). Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics.
Connine, C. M., Blasko, D. G., and Wang, J. (**1994**). ''Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context,'' Percept. Psychophys. **56**, 624–636.
Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., and Smith, L. (**1991**). ''Vowel and consonant recognition with the aid of a multichannel cochlear implant,'' Q. J. Exp. Psychol. A **43**, 585–601.
Dorman, M. F., Loizou, P. C., Kemp, L. L., and Kirk, K. I. (**2000**). ''Word recognition by children listening to speech processed into a small number of channels: Data from normal-hearing children and children with cochlear implants,'' Ear Hear. **21**, 590–596.
Dorman, M. F., Smith, L. M., Smith, M., and Parkin, J. L. (**1996**). ''Frequency discrimination and speech recognition by patients who use the Ineraid and continuous interleaved sampling cochlear-implant signal processors,'' J. Acoust. Soc. Am. **99**, 1174–1184.
Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,'' J. Speech Lang. Hear. Res. **40**, 1201–1215.
Goldinger, S. D. (**1996**). ''Words and voices: Episodic traces in spoken word identification and recognition memory,'' J. Exp. Psychol. Learn Mem. Cogn **22**, 1166–1183.
Hazan, V., and Rosen, S. (**1991**). ''Individual variability in the perception of cues to place contrasts in initial stops,'' Percept. Psychophys. **49**, 187–200.
Hedrick, M. S., and Carney, A. E. (**1997**). ''Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants,'' J. Speech Lang. Hear. Res. **40**, 1445–1457.
HLSYN High-Level Parameter Speech Synthesis System. (**1997**). (Version 2.2). Sensimetics Corporation, Sommerville, MA.
Hochmair-Desoyer, I. J., Hochmair, E. S., and Stiglbrunner, H. K. (**1985**). ''Psychoacoustic temporal processing and speech understanding in cochlear implant patients,'' in *Cochlear Implants*, edited by R. A. Schindler and M. M. Merzenich (Raven, New York), pp. 291–304.
Iverson, P., and Kuhl, P. K. (**1995**). ''Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling,'' J. Acoust. Soc. Am. **49**, 467–471 **97**, 553–562.
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.
Iverson, P., and Kuhl, P. K. (**2000**). ''Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?'' Percept. Psychophys. **62**, 874–886.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–471.
Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). ''The discrimination of speech sounds within and across phoneme boundaries,'' J. Exp. Psychol. **54**, 358–368.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.
Luce, P. A., and Pisoni, D. B. (**1998**). ''Recognizing spoken words: The neighborhood activation model,'' Ear Hear. **19**, 1–36.
Miller, J. L. (**1994**). ''On the internal structure of phonetic categories: A progress report,'' Cognition **50**, 271–285.
Norris, D. G., McQueen, J. M., and Cutler, A. (**2000**). ''Merging information in speech recognition: Feedback is never necessary,'' Behav. Brain Sci. **23**, 299–325.
Nygaard, L. C., and Pisoni, D. B. (**1998**). ''Talker-specific learning in speech perception,'' Percept. Psychophys. **60**, 355–376.
Parkinson, A. J., Parkinson, W. S., Tyler, R. S., Lowder, M. W., and Gantz, B. J. (**1998**). ''Speech perception performance in experienced cochlear-implant patients receiving the SPEAK processing strategy in the Nucleus Spectra-22 cochlear implant,'' J. Speech Lang. Hear. Res. **41**, 1073–1087.
Pastore, R. E., and Farrington, S. M. (**1996**). ''Measuring the difference limen for identification of order of onset for complex auditory stimuli,'' Percept. Psychophys. **58**, 510–526.
Pisoni, D. B. (**1997**). ''Some thoughts on 'normalization' in speech perception,'' in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego).
Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. H. (**1992**). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge).
Repp, B. (**1984**). ''Categorical perception: Issues, methods, findings,'' in *Speech and Language*, edited by N. J. Lass (Academic, New York), Vol. 10, pp. 243–335.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). ''Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation,'' Hear. Res. **127**, 108–118.
Shannon, R. V. (**1989**). ''Detection of gaps in sinusoids and pulse trains by patients with cochlear implants,'' J. Acoust. Soc. Am. **85**, 2587–2592.
Shannon, R. V. (**1992**). ''Temporal modulation transfer functions in patients with cochlear implants,'' J. Acoust. Soc. Am. **91**, 2156–2164.
Shannon, R. V. (**1993**). ''Psychophysics,'' in *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular, San Diego), pp. 357–388.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Sinex, D. G., McDonald, L. P., and Mott, J. B. (**1991**). ''Neural correlates of nonmonotonic temporal acuity for voice onset time,'' J. Acoust. Soc. Am. **90**, 2441–2449.
Soli, S. D. (**1983**). ''The role of spectral cues in discrimination of voice onset time differences,'' J. Acoust. Soc. Am. **73**, 2150–2165.
Stevens, K. N., and Bickley, C. A. (**1991**). ''Constraints among parameters simplify control of Klatt formant synthesizer,'' J. Phonetics **19**, 161–174.
Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., and Cooper, F. S. (**1970**). ''Theoretical notes. Motor theory of speech perception: A reply to Lane's critical review,'' Psychol. Rev. **77**, 234–249.
Svirsky, M. A. (**2000**). ''Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues,'' J. Acoust. Soc. Am. **107**, 1521–1529.
Tyler, R. S., Moore, B. C., and Kuk, F. K. (**1989**). ''Performance of some of the better cochlear-implant patients,'' J. Speech Hear. Res. **32**, 887–911.
Yamada, R. A. (**1995**). ''Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 305–320.

The effects of short-term training for spectrally mismatched noise-band speech

Qian-Jie Fu[a] and John J. Galvin III
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 2 May 2002; accepted for publication 18 November 2002)

The present study examined the effects of short-term perceptual training on normal-hearing listeners' ability to adapt to spectrally altered speech patterns. Using noise-band vocoder processing, acoustic information was spectrally distorted by shifting speech information from one frequency region to another. Six subjects were tested with spectrally shifted sentences after five days of practice with upwardly shifted training sentences. Training with upwardly shifted sentences significantly improved recognition of upwardly shifted speech; recognition of downwardly shifted speech was nearly unchanged. Three subjects were later trained with downwardly shifted speech. Results showed that the mean improvement was comparable to that observed with the upwardly shifted training. In this retrain and retest condition, performance was largely unchanged for upwardly shifted sentence recognition, suggesting that these listeners had retained some of the improved speech perception resulting from the previous training. The results suggest that listeners are able to partially adapt to a spectral shift in acoustic speech patterns over the short-term, given sufficient training. However, the improvement was localized to where the spectral shift was trained, as no change in performance was observed for spectrally altered speech outside of the trained regions. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1537708]

PACS numbers: 43.71.Ky, 43.71.An, 43.66.Ts [CWT]

I. INTRODUCTION

Cochlear implants transform acoustic sounds into electrical signals that directly stimulate remaining auditory nerve fibers, thereby partially restoring hearing sensation to profoundly deaf patients. Multi-channel cochlear implant speech processors divide acoustic signals into several frequency bands, extract the temporal envelope information from each band, convert the acoustic amplitudes into electric currents, and deliver the electric currents to appropriate electrodes situated within the cochlea. To recreate the tonotopic distribution of activity within the normal cochlea, the amplitude envelope from low-frequency bands is delivered to apical electrodes (on an appropriate carrier) and the amplitude envelope from high-frequency bands is delivered to basal electrodes. The frequency-to-electrode mapping provides spectral cues for speech recognition by cochlear implant listeners. The frequency-to-electrode mapping necessarily compresses a wide acoustic frequency range onto the limited cochlear extent of the implanted electrode array. The acoustic frequency ranges mapped to the stimulating electrodes will be shifted as well as compressed. Implant patients may vary in terms of electrode insertion depths and neural populations, and the degree of spectral shifting and compression may be different for individual implant users.

The spectral shift via the frequency-to-electrode mapping can acutely affect implant listeners' speech recognition abilities. Several experiments have explored implant listeners' ability to recognize spectrally shifted speech patterns, given little or no time for adaptation. For example, Fu and Shannon (1999a,b) investigated the interaction between acoustic frequency allocation and electrode location in Nucleus-22 cochlear implant listeners using experimental four-channel continuous interleaved sampling speech processors. In these studies, vowel recognition was measured as a function of ten different frequency allocations and two sets of four-electrode configurations. Each frequency allocation represented the same cochlear extent but different cochlear locations based on Greenwood's frequency-to-place formula (Greenwood, 1990). Results showed that for a given electrode configuration, the best vowel score was obtained within only a narrow range of frequency allocations. When the location of the stimulating electrodes was shifted by 3 mm, the frequency allocation that produced the best vowel recognition also shifted by 3 mm. These results suggest that speech recognition with cochlear implants is highly sensitive to the mapping between frequency allocation and the location of the stimulating electrodes. A severe mismatch between frequency allocation and electrode location could result in a dramatic and immediate deterioration in speech performance. The effect of spectral mismatch was also explored in normal-hearing subjects listening to noise-band simulations of an implant speech processor in the same study. The carrier frequency bands were fixed while the analysis filter bands were systematically shifted. The results showed that the best performance was achieved when the analysis filter bands and the carrier filter bands were matched or closely matched. Performance was unchanged as long as the spectral mismatch between the analysis filters and carrier filters was 3 mm or less (in terms of Greenwood's function). However, the performance dropped steeply for both apically and upwardly shifted speech when the spectral mismatch between

[a]Electronic mail: qfu@hei.org

analysis and carrier filters exceeded 3 mm. In conclusion, while some degree of spectral mismatch can be tolerated (~3 mm), speech recognition can suffer acutely if the spectral mismatch is too severe.

While a severe spectral mismatch may cause a significant performance drop under acute testing conditions, it is difficult to gauge the significance of the frequency-to-electrode assignments tested in such studies because subjects were given no time to adapt to the new patterns of electrical stimulation. Several studies have noted improved speech performance by cochlear implant patients after long-term exposure to new electrical stimulation patterns provided by updated speech processors, speech processing strategies, and/or clinical fitting systems (e.g., Wilson *et al.*, 1991; Pelizzone *et al.*, 1999). For example, Pelizzone *et al.* (1999) reported that, initially, vowel identification scores were unchanged for a group of Ineraid users who switched from the compressed analog (CA) to the continuous interleaved sampler (CIS) speech processing strategy. However, after 6 months of experience with the new CIS strategy, these subjects' vowel identification scores were significantly better than those obtained with the previous CA processor. When the subjects were switched back to their previous CA processor, vowel perception scores returned to the same levels measured before switching to the CIS strategy. This example illustrates that ''acute'' experiments may provide different results than those of long-term studies. As cochlear implant users become more experienced with the implant device and parametric changes to the speech processor, they may be able to largely overcome initial difficulties associated with spectral distortions.

To further investigate the importance of frequency-to-electrode assignment on speech performance by cochlear implant users, the longer-term effects of frequency-to-electrode assignment have also been explored in cochlear implant listeners (Fu *et al.*, 2002; McKay and Henshall, 2002; Skinner *et al.*, 1995). Skinner *et al.* (1995) found that six out of seven Nucleus-22 implant patients using the SPEAK strategy performed better in vowel tests with frequency allocation Table 7 (120–8658 Hz) than with frequency Table 9 (150–10 832 Hz). In the SPEAK strategy, the frequency allocation tables determine the frequency-to-electrode assignment. Skinner *et al.* argued that Table 7 assigned more electrodes to important frequency regions below 1 kHz, thereby improving vowel recognition. Similarly, McKay and Henshall (2002) investigated whether selectively increasing the discrimination of low-frequency information by altering the frequency-to-electrode allocation would improve speech perception by cochlear implant patients. Results showed that some subjects were able to adapt to frequency shifts up to ratio changes of 1.33, as well as changes in the distribution of stimulating electrodes, given 2 weeks' experience. Fu *et al.* (2002) measured speech performance over time in three Nucleus-22 cochlear implant subjects who, for a 3-month period, continuously wore experimental speech processors that were altered in terms of the frequency-to-electrode assignment. A large frequency shift was employed in the study in which the frequency boundary assigned to electrodes was lowered by 1 oct in two subjects and 0.68 oct in one subject. Baseline speech performance using each subject's clinically assigned speech processor was measured just prior to implementation of the experimental processor. Results showed that the experimental processor produced significantly lower performance on all measures of speech recognition immediately following implementation, consistent with the results from the previous ''acute'' experiments (Fu and Shannon, 1999a, b). Over the 3-month test period, all measures were significantly higher than those measured immediately post-fitting. These results indicated that long-term exposure to the frequency-shifted speech processor significantly reduced the initial performance deficit. However, even after 3 months' exposure, speech recognition with the experimental processors remained significantly lower than baseline levels measured with the clinically assigned processors, suggesting the detrimental effects of a severe spectral mismatch were indeed long-standing, though not as damaging as were first observed. Similar results have also been reported in normal-hearing subjects listening to an acoustic simulation of a cochlear implant. Rosen *et al.* (1999) examined the effect of short-term training on normal-hearing listeners' perception of spectrally shifted four-channel noise-band speech. Subjects were able to improve from nearly no recognition to identifying correctly nearly 30% of words in sentences in only 3–4 h. However, recognition of spectrally shifted speech remained significantly lower than that of unshifted speech.

These short- and long-term studies suggest that subjects may be able to completely adapt to the new speech patterns as long as the spectral mismatch is moderate (e.g., frequency shifts up to ratio changes of 1.3). However, if the spectral mismatch is severe (e.g., 1-oct frequency shift), complete adaptation may not occur within a somewhat long observation period (e.g., 3 months), though partial adaptation to the new speech patterns may occur. It is also unclear in these studies whether adaptation was restricted to the trained speech patterns or due to subjects' reshaping of internal speech representations. The present study investigated the effects of short-term learning on spectrally altered speech in six normal-hearing subjects listening to noise-band simulations of a cochlear implant speech processor. Sentence recognition scores were measured as a function of the amount of spectral distortion, which was achieved by restricting cochlear stimulation to simulate various implant electrode insertion depths, ranging from a very deep to a very shallow insertion. To simulate shallow electrode insertions, acoustic speech signals were delivered to basal cochlear locations, resulting in upwardly shifted (basally shifted) speech patterns; to simulate deep electrode insertions, acoustic speech signals were delivered to apical cochlear locations, resulting in downwardly shifted (apically shifted) speech patterns. After baseline measures for all simulated electrode locations, subjects were trained using the most upwardly shifted speech and then retested at all simulated electrode locations after the training. Subjects were also trained later using the most downwardly shifted speech and retested at all simulated electrode locations. Several hypotheses concerning training with spectrally shifted speech were explored.

The first hypothesis predicted that training with the most upwardly shifted speech might indeed improve recognition

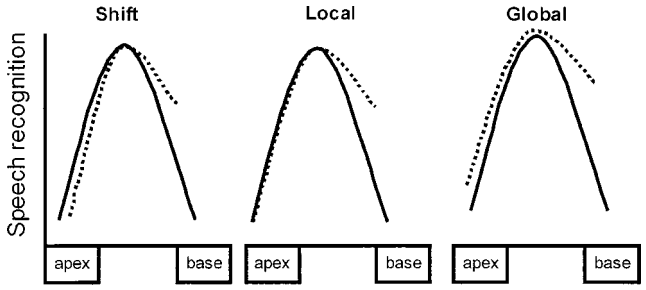


FIG. 1. Three hypothetical outcomes of training with spectrally altered speech. The solid lines represent the baseline data prior to training and the dashed lines show the predicted performance after training.

of upwardly shifted speech, but that recognition of downwardly shifted speech might be reduced; recognition of close-matched speech would not be affected. The underlying mechanism for this ''shift hypothesis'' is a slight adjustment to the internal speech representations due to training with upwardly shifted speech patterns, resulting in a larger spectral mismatch between downwardly shifted speech and the now-reshaped internal representations of speech; subjects would become ''biased'' toward the upwardly shifted speech patterns. However, because subjects would continue to make use of the normal speech patterns in listening conditions beyond the experiment, there would be no change in performance for experimental conditions in which the spectral distortions were closely matched to the long-established, normal speech patterns. The second hypothesis predicted that improved recognition might be restricted to the (upwardly shifted) trained speech patterns, while performance with the untrained (downwardly shifted) speech would not be affected. The underlying mechanism for this ''local adaptation'' hypothesis is that the subjects, in adapting to the (upwardly shifted) trained speech, develop alternative representations of speech while preserving the previous ''internal'' representations. The third hypothesis predicted that training with upwardly shifted speech would improve the recognition of both upwardly and downwardly shifted speech. The underlying mechanism of this ''global adaptation'' hypothesis is that as subjects gradually adapt to the (upwardly shifted) trained speech patterns, they develop an aptitude for adapting to spectrally distorted speech in general, regardless of the shift direction. Figure 1 illustrates the potential outcomes of short-term training based on these three hypotheses.

II. METHODS

A. Subjects

Six normal-hearing listeners aged 25 to 35 participated in the present experiment. All subjects had thresholds better than 15 dB HL at audiometric test frequencies from 250 to 8000 Hz and all were native speakers of American English. All subjects were paid for their efforts.

B. Signal processing

Twenty-channel noise-band speech processors were used to simulate cochlear implant speech processing, and were implemented as follows. Speech signals were band-pass filtered into 20 frequency bands using eighth-order Butterworth filters. To evaluate the effect of acoustic input range, two groups of analysis bands were used: frequency allocation Table 9 used in the Nucleus-22 implant (150–10 831 Hz) and frequency allocation Table 6 used in the Nucleus-24M implant (116–7871 Hz). The temporal envelope of each band was extracted by half-wave rectification and low-pass filtering at 160 Hz. The envelope was used to modulate a wideband noise that was then spectrally limited by a band-pass filter (carrier band). The corner frequencies and bandwidths of the carrier frequency bands were dependent on the simulated electrode insertion depth. The frequency range of the carrier bands was determined by the following equation:

$$p(i) = P_0 + 0.75*i, \quad i = 0,1,...,20, \tag{1}$$

where P_0 is the most apical carrier band location for a given frequency allocation in mm (from the apex). The corner frequencies of the carrier bands were determined by the following equation, from Greenwood (1990):

$$f(i) = 165.4*(10^{P(i)*0.06} - 0.88). \tag{2}$$

Note that Eq. (2) assumes a 35-mm-long cochlea; actual cochlea lengths can vary by several mm. Combining Eqs. (1) and (2), all corner frequencies of carrier frequency bands were determined for a given insertion depth. Between adjacent carrier bands, the crossover attenuation was -3 dB. The most apical carrier band location (P_0) varied between 7.75 and 15.25 mm from the apex of the cochlea (or 27.25 to 19.75 mm from the base) to simulate a range of deep to shallow electrode insertion depths. Six carrier band frequency ranges were generated between these endpoints. Table I lists the corner frequencies of the analysis and carrier bands used in the experiment. Schematic diagrams of experimental conditions are shown in Fig. 2.

The outputs of the carrier bands were then summed and presented to listeners seated in a sound-treated booth via one loudspeaker (Tannoy Reveal) at 70 dBA. Figure 3 shows the spectral envelope of the phoneme /i/; spectral envelopes are shown for both the unprocessed speech and for the 20-channel noise-band processed speech at three cochlear locations (simulated insertion depths). The top panel shows the spectral envelopes when NU-22 Table 9 was used for the analysis filters; the bottom panel shows the spectral envelopes when the NU-24 Table 6 was used. Note the different degrees of spectral mismatch caused by the different frequency allocation tables.

C. Speech materials and procedures

Recognition of words in sentences was measured using novel sentences from the IEEE sentence corpus (IEEE, 1969). The sentences were digitized recordings spoken by one male talker (recorded at House Ear Institute). The IEEE sentences were of easy to moderate difficulty. For testing, a list was chosen pseudo-randomly from among 72 lists, and sentences were chosen randomly, without replacement, from the ten sentences within that list; two lists were used for each testing session. Subjects responded by repeating the sentence

TABLE I. The cutoff frequencies of two analysis filters and six carrier filters. CF0 represents the lower cutoff and CF1 represents the upper cutoff for the lowest filter band. Numbers in the table for CF2–CF20 represents the upper cutoff of the frequency band assigned to successively higher frequency bands.

	Analysis filters		Carrier filters					
	NU22	NU24	$P_0=7.75$	$P_0=9.25$	$P_0=10.75$	$P_0=12.25$	$P_0=13.75$	$P_0=15.25$
CF0	150	116	337	448	585	753	960	1214
CF1	350	243	390	513	665	851	1081	1363
CF2	550	393	448	585	753	960	1214	1528
CF3	750	540	513	665	851	1081	1363	1710
CF4	950	687	585	753	960	1214	1528	1913
CF5	1150	833	665	851	1081	1363	1710	2138
CF6	1350	978	753	960	1214	1528	1913	2387
CF7	1550	1125	851	1081	1363	1710	2138	2663
CF8	1768	1285	960	1214	1528	1913	2387	2970
CF9	2031	1477	1081	1363	1710	2138	2663	3310
CF10	2333	1696	1214	1528	1913	2387	2970	3687
CF11	2680	1949	1363	1710	2138	2663	3310	4106
CF12	3079	2238	1528	1913	2387	2970	3687	4570
CF13	3571	2597	1710	2138	2663	3310	4106	5085
CF14	4184	3043	1913	2387	2970	3687	4570	5656
CF15	4903	3565	2138	2663	3310	4106	5085	6289
CF16	5744	4177	2387	2970	3687	4570	5656	6992
CF17	6730	4894	2663	3310	4106	5085	6289	7771
CF18	7885	5734	2970	3687	4570	5656	6992	8635
CF19	9238	6718	3310	4106	5085	6289	7771	9594
CF20	10 823	7871	3687	4570	5656	6992	8635	10 657

as accurately as possible; the experimenter tabulated correctly identified words and sentences. Subjects were trained daily using the DARPA/TIMIT acoustic-phonetic continuous speech corpus (Garofolo *et al.*, 1993). The multi-talker TIMIT sentences were of moderate to extreme difficulty. For training, subjects auditioned a set of 300 sentences processed by the most upwardly shifted noise-band processor (15.25 mm from the apex). Subjects viewed the text of the sentence as the sentence was played, and were allowed to repeat the sentence as often as they liked.

The six subjects were divided into two groups. Three subjects were given speech processors using the Nucleus-22 Table 9 analysis filters and three were given speech processors using the Nucleus-24 Table 6 analysis filters. The testing and training was conducted as follows. Baseline (pretraining) data for all simulated insertion depths were collected on day 1, using the IEEE sentences. Over the next four days, subjects trained with TIMIT sentences processed by the most upwardly shifted noise-band processor (15.25 mm from the apex). On day 5, after a final training session, subjects were retested for all simulated insertion depths. The three subjects who were assigned NU-22 Table 9 were retested and trained using the most downwardly shifted speech (7.75 mm from apex) at a later date (3–10 days after completing testing and training with the most upwardly shifted speech).

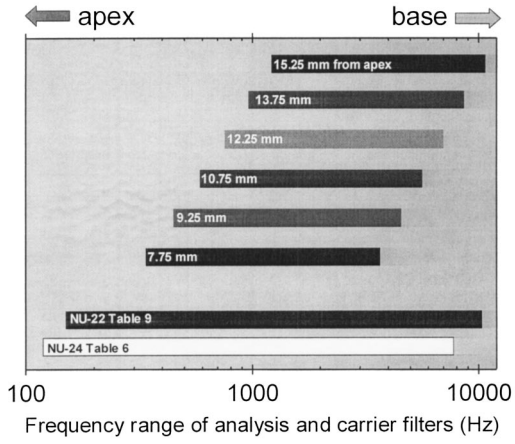


FIG. 2. Frequency ranges for the analysis and carrier filter bands. The bottom two bars show the analysis filter range for frequency allocation Tables 9 (Nucleus-22) and 6 (Nucleus-24). The upper six bars show the carrier band frequency ranges for six simulated electrode insertion depths, ranging from deep (7.75 mm from the apex) to shallow (15.25 mm from apex). The frequency range of the carrier bands was fixed to simulate the cochlear extent of an inserted electrode array.

III. RESULTS

Figure 4 shows the mean IEEE sentence recognition scores for all speech processor conditions before and after subjects were trained with spectrally shifted TIMIT sentences. Figure 4(a) shows data for the three subjects assigned the Nucleus-22 Table 9 analysis filters, before and after training with TIMIT sentences processed by the most upwardly shifted noise-band processor. Figure 4(b) shows data for the three subjects assigned the Nucleus-24 Table 6 analysis filters, before and after training with TIMIT sentences processed by the most upwardly shifted noise-band processor. Figure 4(c) shows data for the three subjects assigned the Nucleus-22 Table 9 analysis filters, before and after training with TIMIT sentences processed by the most downwardly shifted noise-band processor; this additional train and retest condition was performed 3–10 days after subjects had completed the earlier train and test condition. The solid line shows baseline data before the training while the dashed line

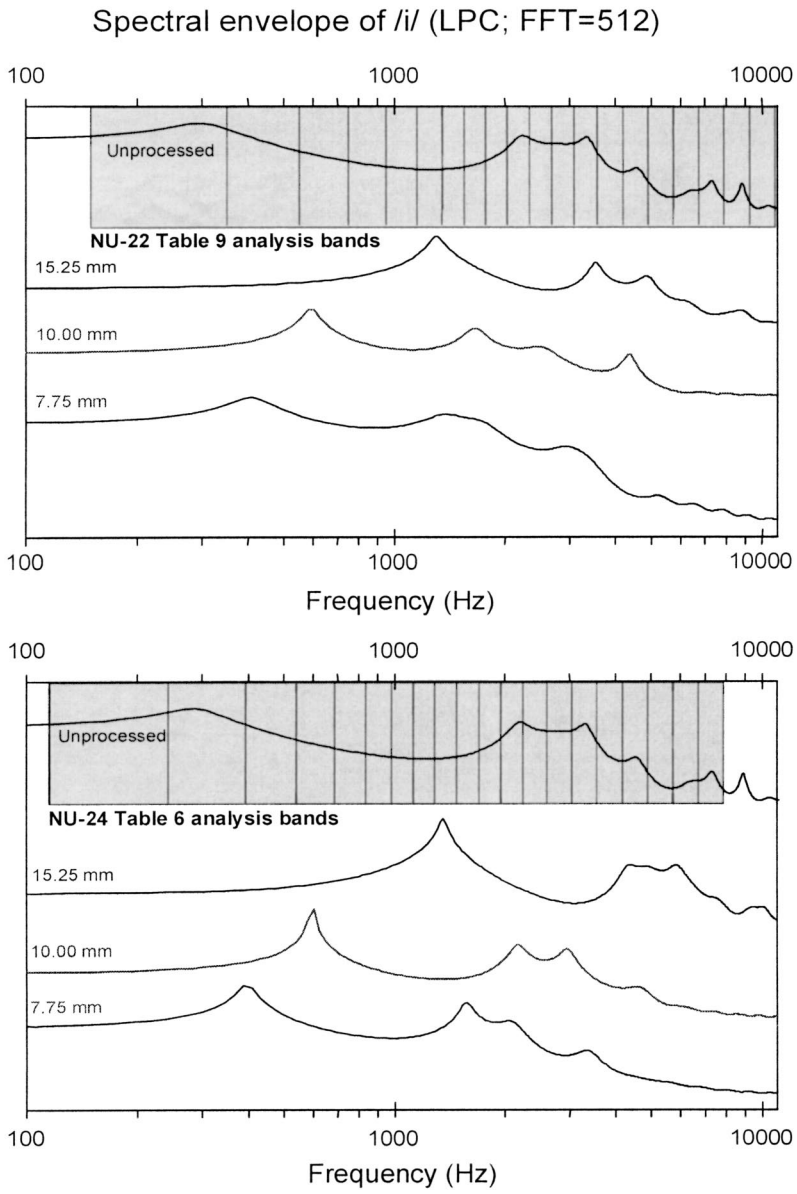


FIG. 3. The effects of analysis and carrier frequency ranges on the spectral envelope of the vowel /i/. The top panel shows the frequency analysis by Nucleus-22 Table 9. The unprocessed spectral envelope is shown in the shaded area, and the individual analysis bands are shown in the shaded area. The signal was analyzed by Table 9 and the amplitude envelope from each filter band was used to modulate a corresponding carrier band, depending on the simulated electrode location. The spectral envelope of /i/ is shown for three carrier band locations, ranging from a simulated deep electrode insertion (7.75 mm from apex) to a simulated shallow insertion. The bottom panel shows the effect of frequency analysis by Nucleus-24 Table 6 on the spectral envelope for three simulated electrode insertion depths.

shows the data after the 5 days of training. The filled symbols show the simulated insertion depth at which subjects were trained.

For subjects assigned the Nucleus-22 Table 9 analysis filters [Fig. 4(a)], baseline measures showed that mean sentence recognition was best with a simulated insertion depth of 12.25 mm from the apex (slightly more basal than the typical insertion depth of 10 mm from apex). Baseline performance was largely unchanged for adjacent simulated electrode insertion depths (10.75 to 13.75 mm from the apex). A large performance drop (~50 percentage points) was observed for the most apically situated carrier bands ($P_0 = 7.75$ mm). A smaller drop in recognition (14 percentage points) was observed for the most basally situated carrier location ($P_0 = 15.25$ mm). After training with upwardly shifted speech, a Student *t*-test revealed no significant difference between pre- and posttraining performance for the apical and mid-cochlea carrier locations ($p > 0.05$). However, a Student *t*-test revealed a significant improvement at the most basally situated carrier location where training had been performed ($p = 0.02$).

For subjects assigned the Nucleus-24 Table 6 analysis filters [Fig. 4(b)], baseline measures showed that peak sentence recognition was found for a range of apical to mid-cochlea carrier band locations (close to the typical electrode insertion depth 10 mm from the apex). Baseline recognition scores were nearly unchanged for simulated electrode insertion depths ranging between 7.75 and 10.75 mm from the apex. However, there was a large performance drop (67 percentage points) for the most basally situated carrier location ($P_0 = 15.25$ mm). After training with upwardly shifted speech, performance at the most basal carrier location improved by more than 20 percentage points ($p = 0.03$). Slight improvement was also observed at nearby basal carrier locations after training. Improvement at apical and middle carrier locations was minimal.

Subjects who were assigned the Nucleus-22 Table 9 analysis filters were also trained with TIMIT sentences pro-

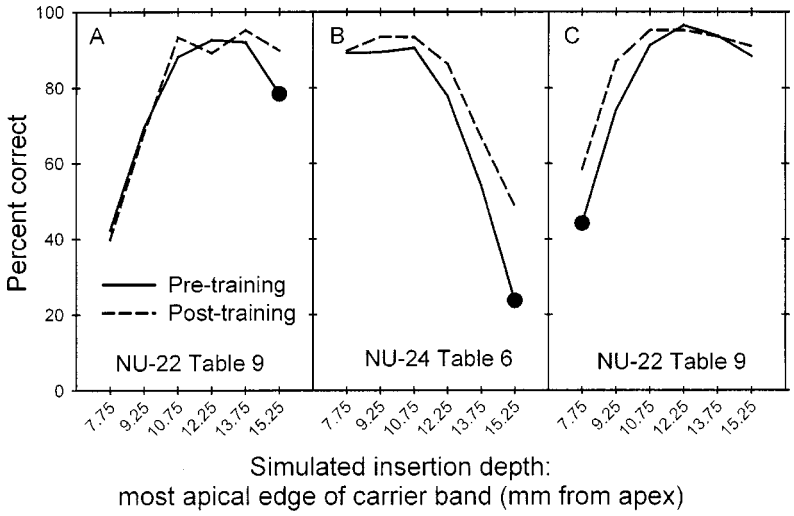


FIG. 4. Baseline pretraining and posttraining results for IEEE sentence recognition with 20-channel noise-band processors. Subjects were trained with either upwardly shifted or downwardly shifted 20-channel speech. Panel (a) shows results for three subjects assigned the frequency analysis filters from Nucleus-22 Table 9 (150–10 823 Hz), trained with upwardly shifted speech. Panel (b) shows results for three subjects assigned the frequency analysis filters from Nucleus-24 Table 6 (116-7871 Hz), trained with upwardly shifted speech. Panel (c) shows results for the three subjects assigned the frequency analysis filters from Nucleus-22 Table 9 [same subjects shown in panel (a)] who were retested and trained with downwardly shifted speech after completing the training with upwardly shifted speech. The *x* axis shows the carrier band locations for various simulated electrode insertion depths; the *y* axis shows percent correct. The solid lines show baseline performance; the dashed lines show performance after five training sessions. The filled circles show the carrier band location where training was performed.

cessed by the most downwardly shifted noise-band processor (7.75 mm from the apex) at a later date. This training and retesting was performed between 3 and 10 days after completing training with upwardly shifted speech. The additional train and retest condition was created because the baseline performance deficit observed at the most basal carrier location was much less than that observed at the most apical carrier location. In fact, the subjects using the Nucleus-22 Table 9 filters analysis filters had relatively high levels of sentence recognition at the most basal carrier location ($P_0 = 15.25$ mm). Note that subjects assigned the Nucleus-24 Table 6 analysis filters did not participate in this additional train and retest condition because near-peak baseline performance was observed at the most apical carrier location, meaning there was little room for improvement. Mean performance for this train and retest condition is shown in Fig. 4(c). Baseline performance for all simulated insertion depths was remeasured before training with downwardly shifted speech, and are shown by the solid line. Similar to the earlier baseline measures, peak performance in the retested baseline measures was found at the simulated electrode insertion depth of 12.25 mm from the apex. In the retested baseline measures, there remained a slight performance drop at the most basal carrier location ($P_0 = 15.25$ mm). However, the performance deficit at this carrier location was less than that observed in the original baseline measures. In fact, the retested baseline scores were quite similar to the posttraining performance of the previous experiment [as shown by the dashed line in the panel 4(a)]. After training with the most downwardly shifted speech, mean recognition scores remained relatively unchanged at the middle and basal carrier locations. However, significantly improved recognition was observed for the trained apical carrier locations ($p < 0.001$).

IV. DISCUSSION

The results demonstrate that short-term training with spectrally altered speech can significantly improve the recognition of the trained speech patterns. However, the improvement was generally restricted to the cochlear location where the training had taken place. The results strongly support the hypothesis of ''local adaptation,'' which asserted that subjects would only adapt to the specific spectral mismatch on which they were trained. The results also suggest that subjects may have developed alternate spectral patterns after training with spectrally altered speech, while retaining previous ''internal'' representations, at least over the short term.

Training with upwardly shifted speech indeed improved the recognition of upwardly shifted speech, while not affecting speech recognition at the apical and mid-cochlea carrier locations. Similarly, the follow-up training with downwardly shifted speech improved the recognition of downwardly shifted speech and did not affect speech recognition at the basal and mid-cochlea carrier locations. The subjects who were assigned the Nucleus-22 Table 9 analysis filters participated in both training experiments, with downwardly shifted speech training occurring at a later date (3–10 days after completion of upwardly shifted speech training). It was interesting to observe that the improved recognition of upwardly shifted speech (after training with upwardly shifted speech) was largely retained, at least over the short term. The recollected baseline data (before training with downwardly shifted speech) showed that recognition of upwardly shifted speech was similar to the levels of performance after training with upwardly shifted speech. After 5 days of training with and improved recognition of downwardly shifted speech, recognition of upwardly shifted speech remained at the recollected baseline levels (which were the same levels after the earlier training with upwardly shifted speech). These results suggest that subjects may have temporally preserved (''internalized'') the upwardly shifted speech patterns while accommodating the newly trained downwardly shifted speech patterns.

The ''local adaptation'' to spectrally altered speech suggests that listeners are able to accommodate alternate speech patterns, while preserving previously accommodated patterns. If a ''global adaptation'' had occurred, in which subjects' performance improved at all carrier band locations, subjects may have simply been adapting to the reduced spectral resolution of the noise-band processors. Given enough time and training at various carrier locations, the combined improvements from the localized training might cumula-

tively amount to a ''global adaptation.'' If a ''shifted adaptation'' adaptation had occurred, in which the improvements due to localized training also produced a deficit at other carrier locations, subjects would be unable to retain previously learned spectral patterns.

There was an interactive effect between the frequency analysis range and the place of stimulation in both baseline and posttraining measures. The baseline performance showed that, acutely measured, frequency allocation Nucleus-22 Table 9 (150–10 823 Hz) was better for basal carrier locations that simulated shallow electrode insertions, while frequency allocation Nucleus-24 Table 6 (116–7871 Hz) was better for middle and apical carrier locations that simulated typical and deep insertion depths. As insertion depths of 10 mm or more from the apex are typical for contemporary implant devices, the frequency allocation used in Nucleus-24 Table 6 should provide the best results, as the degree of spectral mismatch may somewhat reduced. Such reduced spectral mismatch result may have contributed to the improved speech recognition observed in a previous study (Skinner *et al.*, 1995) with Nucleus-22 patients who were reassigned frequency allocation Table 7 (120–8568 Hz) rather than the default frequency allocation Table 9. Skinner *et al.* (1995) argued that the improvement was due to the better spectral resolution in the low-frequency region. In contrast, the long-term studies conducted by Fu *et al.* (2002) revealed that improved spectral resolution in the low-frequency region may not, in and of itself, provide better recognition. In that study, cochlear implant patients continuously used a spectrally shifted speech processor for a 3-month period; subjects were assigned frequency allocation Table I (75–5411 Hz), which mapped the significantly more low-frequency speech information onto the available electrodes than the subjects' clinically assigned frequency allocation (Table 7 or 9). However, vowel recognition remained significantly lower with frequency allocation (Table I) than with the frequency allocations used in subjects' clinically assigned speech processor (Table 7 or 9), even after 3 months of continuous exposure. This suggests that increased spectral resolution of low-frequency speech information may not overcome a severe spectral mismatch, and that the degree of spectral mismatch may be the limiting factor in cochlear implant listeners' utilization of speech spectral cues. McKay and Henshall (2002) also found that, while most subjects were able to accommodate frequency shifts up to ratio changes of 1.3, it was unclear if subjects could have accommodated greater frequency shifts, given enough time or given gradual adaptation (learning small incremental shifts).

These and many other studies have examined listeners' ability to adapt to changes to or deterioration of representations of speech patterns. There would be very little adaptation needed if, for example, the number of perceptual channels and/or spectral resolution were to be increased or the spectral mismatch between the acoustic signal and electrode positions were to be reduced, as speech recognition will most likely be immediately improved. Skinner *et al.* (1995) found this sort of acute improvement by moving Nucleus-22 implant users from frequency allocation Table 9 (150–10 823 Hz) to Table 7 (116–7871 Hz). It is uncertain whether this improvement was due to increased spectral resolution or reduced spectral mismatch, or a combination of both effects, but subjects were able to accommodate a shift from their previously adapted electric map without much practice. McKay and Henshall (2002) found little improvement when configuring implant listeners' speech processors according to patients' electrode discriminability. Such a mapping might have improved the spectral resolution and/or increased the stimulation rate, either of which could have resulted in improved performance. The lack of improvement (indeed, the sometimes reduced performance with perceptual mapping) suggests that a mismatch between the acoustic signal and stimulating electrodes may have proved too difficult to accommodate. The subjects were able to accommodate small shifts in the frequency-to-electrode assignment, reemphasizing the importance of spectral mapping and learning in implant speech recognition.

V. CONCLUSION

The present study examined the effects of short-term learning on normal-hearing listeners' ability to accommodate spectrally altered speech patterns. Results showed that speech recognition with 20-channel noise-band processors was acutely affected by severe spectral mismatch. Initially, subjects could tolerate only a small amount of spectral mismatch. However, after short-term training with severely spectrally altered speech, subjects were able to significantly improve their speech recognition of spectrally shifted speech where the training had been performed. The improvement was restricted to the trained spectral shift, and did not generalize to other spectral shifts distant from the location where the training had occurred. These results strongly suggest that a ''local adaptation'' had occurred, in which listeners adapted only to the specific spectral alteration where training was performed. Such local adaptations were preserved, at least over the short term. Listeners may develop alternate spectral patterns given enough training, while preserving previous ''internal'' representations of speech sounds.

ACKNOWLEDGMENTS

The research was supported by Grants Nos. R03-DC-03861 and R01-DC04993 from NIDCD. We are grateful to Christopher Turner and two anonymous reviewers for useful suggestions on an earlier draft of this paper. We would like to also thank the research participants for their considerable time spent with this experiment.

Fu, Q.-J., and Shannon, R. V. (**1999a**). ''Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus 22 cochlear implant,'' Ear Hear. **20**, 332–344.

Fu, Q.-J., and Shannon, R. V. (**1999b**). ''Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing,'' J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q.-J., Shannon, R. V., and Galvin III, J. J. (**2002**). ''Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant,'' J. Acoust. Soc. Am. **112**, 1664–1674.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L. (**1993**). ''The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM,'' NTIS order number PB91-100354.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

IEEE (**1969**). *IEEE Recommended Practice for Speech Quality Measurements* (Institute of Electrical and Electronic Engineers, New York).
McKay, C. M., and Henshall, K. R. (**2002**). ''Frequency-to-electrode allocation and speech perception with cochlear implants,'' J. Acoust. Soc. Am. **111**, 1036–1044.
Pelizzone, M., Cosendai, G., and Tinembart, J. (**1999**). ''Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for Ineraid implanted subjects,'' Ear Hear. **20**, 228–237.
Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). ''Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants,'' J. Acoust. Soc. Am. **106**, 3629–3636.
Skinner, M. W., Holden, L. K., and Holden, T. A. (**1995**). ''Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 307–311.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''New levels of speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth

Andrew Faulkner,[a] Stuart Rosen, and Deborah Stanton
Department of Phonetics and Linguistics, UCL, Wolfson House, 4 Stephenson Way, London NW1 2HE, United Kingdom

(Received 17 September 2002; accepted for publication 19 November 2002)

It has been claimed that speech recognition with a cochlear implant is dependent on the frequency alignment of analysis bands in the speech processor with characteristic frequencies (CFs) at electrode locations. However, the most apical electrode location can often have a CF of 1 kHz or more. The use of filters aligned in frequency to relatively basal electrode arrays leads to the loss of lower frequency speech information. This study simulates a frequency-aligned speech processor and common array insertion depths to assess this significance of this loss. Noise-excited vocoders simulated processors driving eight electrodes 2 mm apart. Analysis filters always had center frequencies matching the CFs of the simulated stimulation sites. The simulated insertion depth of the most apical electrode was varied in 2-mm steps between 25 mm (CF 502 Hz) and 17 mm (CF 1851 Hz) from the cochlear base. Identification of consonants, vowels, and words in sentences all showed a significant decline between each of the three more basal simulated electrode configurations. Thus, if implant processors used analysis filters frequency-aligned to electrode CFs, patients whose most apical electrode is 19 mm (CF 1.3 kHz) or less from the cochlear base would suffer a significant loss of speech information. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1536928]

PACS numbers: 43.71.Ky, 43.71.Es, 43.66.Ts [CWT]

I. INTRODUCTION

It has been claimed that speech recognition with a cochlear implant is significantly impaired by a frequency mismatch of the analysis bands in the speech processor to the characteristic frequencies (CFs) at the implanted electrode locations when this mismatch is equivalent to basalward basilar membrane shifts of 3 mm or more (Shannon *et al.*, 1998). The support for this claim comes from Shannon *et al.*'s simulations of cochlear implant speech processing in normally hearing listeners using vocoder-based processing, reinforced by other studies using similar methods (Dorman *et al.*, 1997; Fu and Shannon, 1999a). For prelingually deafened patients, with minimal auditory experience of the distribution of auditory speech cues over frequency, and hence over cochlear place, it seems unlikely that such mismatch would have negative consequences. However, for implant users with many years of exposure to the frequency-to-place mapping of the normal acoustically stimulated ear, a frequency-to-place mismatch may represent a significant obstacle to speech perception.

In the simulations of frequency-to-place mismatch cited above, speech is presented as a series of band-limited carriers, each modulated by an amplitude envelope extracted from one of a series of band-pass analysis filters. When the band-limited carriers are shifted upwards in frequency relative to the analysis band that determines the carrier's amplitude envelope, performance in speech intelligibility tasks is substantially poorer than in an unshifted control condition. The center frequencies of the carrier bands may be assumed to simulate the positions of the electrodes of an array, with upward-shifted carrier bands representing less apical sets of electrode positions. One practical implication of this effect of upward spectral shifting is that the speech receptive performance of cochlear implant users would be improved by the matching of speech processor analysis filters to the characteristic frequencies at the implant electrode locations.

One caution in accepting this implication is that the studies cited above are all based on performance without any extended opportunity to adapt to the effects of spectral shifting. In marked contrast to the findings of these studies, when normal-hearing listeners are given a few hours of training with spectrally shifted speech, performance is substantially increased, indicating that the effect of spectral shifting is much reduced after some experience (Rosen *et al.*, 1999). Since implant users necessarily use the clinical mapping of speech processor filters to their electrode locations for extended periods of time, it is very likely that they too adapt to the frequency mapping provided by their implant. That such adaptation does occur in patients is supported by a study in which the processor filter to electrode mapping was varied (Fu and Shannon, 1999a). Here, the participating subjects performed better with a mapping similar to that they were used to than with alternative mappings to which they were acutely exposed. In a later study in which three implant users had 3 months experience of an experimental mapping, speech perception improved significantly over the first three weeks, although it did not reach the baseline levels observed with the familiar clinically fitted processor (Fu *et al.*, 2002). In this later study, the experimental processor map used filters up to one octave lower than the clinical processor, and the results confirm that implant users can at least partially adapt to a basalward shift of excitation. A further study in experienced implant users that would be expected to reveal a

[a]Electronic mail: andyf@phon.ucl.ac.uk

lack of adaptation to upward spectral shift was recently reported (Harnberger *et al.*, 2001). In this study implant users selected the tokens from a set of synthesized vowel stimuli that best matched their expectation of the sounds of a set of vowels. An incomplete adaptation to spectral shifting would be expected to lead to choices of stimuli with lower first and second formants than natural vowels. However, there was no evidence of such effects. Overall, these findings suggest that experience can at least partially if not fully outweigh any benefit that might arise from an improved frequency match between a speech processor and the CFs at electrode locations.

A second reason for caution in accepting the implication that speech processor filters should match electrode locations comes from a consideration of the range of electrode locations that are observed in implanted patients. In a study of 19 patients implanted with the Nucleus 22 channel electrode, spiral CT data showed that the most apical electrode position varied between 24 and 13.7 mm from the base of the cochlea, with a median distance of 20.3 mm (Ketten *et al.*, 1998). All these electrode arrays were reported at surgery as fully inserted. From the cochlear position to frequency map due to Greenwood (1990), the range of characteristic frequencies at the most apical electrode in this patient group can be estimated to be between 400 and 2600 Hz, with a median of 1000 Hz.[1] The use of a speech processor whose lowest frequency band is centered on the CF of a most apical electrode at a position 20 mm or less from the cochlear base must entail the loss of speech information at frequencies below 1 kHz. Additional higher frequency information would be introduced around the higher CFs of the more basal electrodes, but, for a typical 14 to 16 mm array length, these frequencies are above 5 kHz. Articulation index (AI) studies show that the loss of information below 1 kHz will significantly reduce the intelligibility of unprocessed speech, and that additional higher frequency information will be of slight importance (French and Steinberg, 1947; ANSI, 1997).

It is, however, possible that AI predictions are not appropriate for speech as represented by cochlear implant stimulation, where, amongst many factors, spectral resolution is substantially reduced in comparison to that of normal hearing. Vowel identification data through acoustic simulations that address this issue for insertion depths that extend from CFs of 290 to 960 Hz at the most apical electrode location have been described by Fu and Shannon (1999a). Over this range, simulated insertion depth had little effect, and indeed AI predictions are also affected little by the equivalent change of speech bandwidth. However, shallower insertions than these appear common, and when the lowest frequency band of a tonotopically mapped processor does not cover frequencies below about 1 kHz, greater effects on intelligibility can be expected.

In order to examine the effects of changes in the frequency span of speech information delivered by a cochlear implant, the present study simulates the effect of electrode insertion depth on the intelligibility of speech processed through an eight-band cochlear implant speech processor. Given the inconclusive outcomes of related studies in implant users, for whom the familiar mapping appears to have an inherent advantage over others (Fu and Shannon, 1999b), simulations in normal-hearing listeners are likely to provide the most tractable approach to this issue. Here, simulated electrode locations are varied to span a 14-mm cochlear region with the most apical electrode insertion depth varying from 17 to 25 mm from the cochlear base, positions that are representative of the range of electrode locations found by Ketten *et al.* (1998). Analysis filters are aligned with the CFs of the simulated stimulation sites. The frequency range represented for the simulated 25-mm insertion depth is 416 Hz to 5.14 kHz, while for the simulated 17-mm depth the frequency range is 1.59 to 15.85 kHz.

TABLE I. Simulated position of electrodes and filter center and cutoff frequencies. The rightmost five columns indicate the allocated center frequencies of bands 1–8 for each of the five processors. In the text, positions are given to the nearest mm.

Distance from base (mm)	Center frequency (Hz)	Cutoff (Hz)	Simulated insertion depth (mm)				
			24.9 Band	22.9 Band	20.9 Band	18.9 Band	16.9 Band
25.9		416					
24.9	502		1				
23.9		601					
22.9	715		2	1			
21.9		845					
20.9	995		3	2	1		
19.9		1167					
18.9	1364		4	3	2	1	
17.9		1591					
16.9	1851		5	4	3	2	1
15.9		2150					
14.9	2492		6	5	4	3	2
13.9		2886					
12.9	3338		7	6	5	4	3
11.9		3857					
10.9	4453		8	7	6	5	4
9.9		5138					
8.9	5923			8	7	6	5
7.9		6826					
6.9	7861				8	7	6
5.9		9050					
4.9	10 416					8	7
3.9		11 983					
2.9	13 783						8
1.9		15 850					

II. METHOD

A. Speech processing and equipment

Speech processing used an eight-band noise-excited vocoder similar to that used by Shannon *et al.* (1995). The channel filter center frequencies and −3 dB cutoff frequencies are shown in Table I. This series of center frequencies represents cochlear locations separated by a distance of 2 mm. Five insertion depths were simulated, also in 2-mm steps, with the most apical simulated electrode position varying between 17 and 25 mm from the cochlear base. Crossover and center frequencies for both the analysis and output filters were calculated using an equation (and its inverse) relating position on the basilar membrane to characteristic frequency, assuming a basilar membrane length of 35 mm (Greenwood, 1990):

$$\text{frequency} = 165.4(10^{0.06x} - 1),$$

$$x = \frac{1}{0.06}\log\left(\frac{\text{frequency}}{165.4} + 1\right).$$

The stages of processing in each band comprised an analysis filter, half-wave rectification, envelope smoothing with a 400-Hz low-pass filter, multiplication of white noise by the envelope, and an output filter that always matched the analysis filter. Finally, the outputs of each band were summed. Each channel of the processor received speech as input, without preemphasis.

Two implementations of this processing were employed. Training, which was through live-voice connected discourse, made use of real-time processing, while testing employed off-line processing implemented in MATLAB.

Off-line processing was executed at a 44.1-kHz sample rate. Prior to processing, all the recorded speech materials were band-limited to 11.05 kHz. Analysis filters in the off-line processing were sixth-order Butterworth IIR designs (with three orders per upper and lower side) having responses that crossed 3 dB down from the pass-band peak. Envelope smoothing used second-order low-pass Butterworth filters (400 Hz cutoff). A final low-pass filter was applied to the summed waveform from each of the eight bands at the upper cutoff frequency of the highest frequency channel (15.8 kHz) to limit the signal spectrum. This used a sixth-order low-pass elliptical filter forwards and backwards to obtain the equivalent of a 12th-order elliptical filter with zero phase characteristic.

Real-time processing ran at a 16-kHz sample rate on a DSP card (Loughborough Sound Images TMSC31), and was implemented using the Aladdin Interactive DSP Workbench (Hitech Development AB). To reduce the required computation, elliptical filter designs were used, with the same −3-dB crossover frequencies as those used for off-line processing, Analysis and output filters were fourth-order band-pass designs, while the envelope smoothing filters were third-order low-pass. Because of the limited 8-kHz bandwidth, the uppermost three bands of the total set used, those having center frequencies of 7.86, 10.4, and 13.8 kHz, could not be implemented. Hence, in training, the simulated insertion depth of 17 mm used only five bands, the 19-mm depth used six, and the 21-mm depth used seven bands. All testing was based on off-line processing in which all eight bands were always implemented.

B. Tests of speech reception

1. Consonant identification

The consonant set contained 20 intervocalic consonants /m, b, w, p, v, f, θ, n, d, z, t, s, r, l, j, ʃ, tʃ, dʒ, g, k/ with the vowels /i/, /ɑ/, and /u/. Stress placement was on the second syllable. Materials were from digital anechoic recordings of one female and one male talker. Both talkers had a standard Southern British English accent. Each run presented 40 consonants, with one consonant from each talker being selected at random from a set of six to ten tokens. Stimulus presentation was computer controlled. Subjects responded using the computer mouse to select 1 of 20 buttons on the computer screen that were orthographically labeled to represent each of the 20 consonants. To reinforce any learning that may have taken place, feedback was given by a visual display of the presented consonant after each response.

2. Vowel identification

Seventeen b-vowel-d words from the same male and female talkers were used, again from digital anechoic recordings. Presentation was computer controlled. Each test run presented one token of each word from each of the two talkers, selected at random from a total set of six to ten tokens of each word from each talker. The vowel set contained ten monophthongs, /æ/ (bad); /ɑː/ (bard); /iː/ (bead); /e/ (bed); /ɪ/ (bid); /ɜː/ (bird); /ɒ/ (bod); /ɔː/ (board); /uː/ (booed); /ʌ/ (bud)) and seven diphthongs, /eə/ (bared); /eɪ/ (bayed); /ɪə/ (beard); /aɪ/ (bide); /əʊ/ (bode); /aʊ/ (boughed); /ɔɪ/ (Boyd). The parenthesized spellings are those that appeared on the computer response buttons. As in consonant identification, feedback was given after each response, by the visual display of the presented word.

3. Sentence perception

BKB sentences from two different male and female talkers with the same British accent were used. The female speech was from a digital audio recording made simultaneously with an audio-visual recording (EPI Group, 1986; Foster *et al.*, 1993). The male speech was from an anechoic digital recording. Each test run used one list of 16 sentences with 50 scored key words per list. Sixteen sentences from the ASL sentence set (MacLeod and Summerfield, 1990) produced by the same male talker as the BKB sentences were also used in an initial practice session. No visual feedback was given for sentence testing.

C. Subjects

Eight adult native speakers of English took part. They were screened for normal hearings at 0.5, 1, 2, and 4 kHz, and paid for their services.

D. Procedure

All testing and training took place in a sound-isolated room. The subject received diotic presentation of the processed speech stimuli over headphones (Sennheiser HD475 headphones for testing, AKG K240DF for training). Presentation levels were approximately 70 dBA. Since the processing conditions using higher frequency filters led to lower level processed output, a level correction was applied to ensure that all conditions were presented at a similar SPL.

While the primary concern on this study is not with perceptual adaptation, we have found considerable training effects with some forms of simulated cochlear implant processing (Rosen *et al.*, 1999). We sought to reduce variability due to learning by the use of a 15-min training period prior to testing in each processor condition. Interactive training was performed with the talker and subject in adjacent sound-isolated rooms, without visual communication.

The processing condition was held constant throughout each of 11 sessions of approximately 1 h. Each session com-

menced with 15 min of connected discourse training (DeFilippo and Scott, 1978) with processed speech. The talker (author DS) was not visible to the subjects. In the testing that followed, the subject was presented with a sentence list (16 sentences) from each of the male and female talkers, followed by the consonant stimuli (120 items) and finally the vowel stimuli (two lists: 136 items). The first session was treated as practice, and employed the intermediate simulated insertion depth of 21 mm. In each of ten subsequent test sessions, the same sequence of training and testing was again administered. Each processing condition was used in two testing sessions, one early in the series (sessions 2 to 6) and one later in the series (sessions 7 to 11). Apart from this constraint, the condition presented in each test session was randomly ordered over sessions for each subject.

III. RESULTS

Analyses of each test dataset were performed using repeated-measures ANOVA, with factors of simulated insertion depth, talker, and test run (scores from the first five test sessions compared to scores from last five test sessions). Vowel context was an additional factor in the analyses of consonant identification accuracy. Hyunh-Feldt epsilon corrections were applied to all F tests of factors with more than 1 degree of freedom.

A. Sentences

The results are shown in Fig. 1. ANOVA (see Table II) showed significant main effects of simulated insertion depth and of talker, and a significant interaction of insertion depth and talker. The effect of test run was close to significance, and there was a marginally significant interaction of insertion depth and test run. This interaction represented an increase in scores in the two most basal simulated insertion depths on the second test run compared to the first. Scores were at or close to ceiling levels for the more apical simulated insertions, which would obscure any learning effects in these conditions. The increase in performance on the second test at the 19-mm simulated depth was slight, but at the 17-mm simulated insertion depth there was an increase from approximately 60% to 80% key words correct.

Because of the insertion depth by talker interaction, data for the two talkers were subjected to two separate ANOVAs (see again Table II). These showed main effects of insertion depth for both talkers. Further subanalyses of the male talker data within the first and second run showed highly significant main effects of insertion depth in each case. In every analysis and subanalysis, planned comparisons between insertion

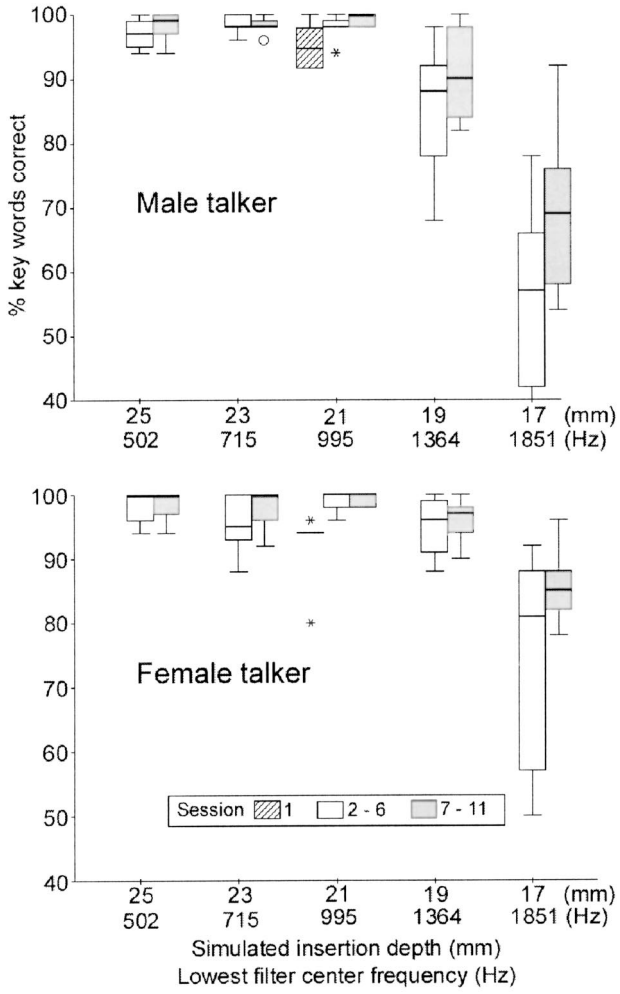


FIG. 1. Key words correct for BKB sentences as a function of simulated insertion depth. The center frequency of the lowest band is also indicated on the x axis. Scores are shown for each of the two talkers, and from the two test runs separately. Scores from the practice session at the 21-mm insertion depth are also shown. The box and whisker plots show the median score (bar), the interquartile range (box), and the range (whiskers). Outlying points are shown as unfilled circles or asterisks.

TABLE II. Significant terms in ANOVA of sentence data for both talkers and for the male and female talker separately. $\eta 2$ indicates the eta-squared statistic, which estimates the proportion of the variance in the data that can be attributed to the factor.

Talker	Factor	df	F	p	$\eta 2$
Both	Simulated insertion depth	2.2,13.1	88.2	<0.001	0.94
	Talker	1,6	55.4	<0.001	0.90
	Test run	1,6	5.81	0.052	0.49
	Simulated insertion depth by Talker	2.1,12.3	22.6	<0.001	0.79
	Simulated insertion depth by test run	2.3,13.8	3.66	0.048	0.38
Male	Simulated insertion depth	2.2,15.2	54.5	<0.001	0.89
	Test run	1,7	9.62	0.017	0.58
	Simulated insertion depth by test run	2.9,20.5	4.21	0.019	0.38
Female	Simulated insertion depth	1.5,9.6	36.1	<0.001	0.86

depths showed that scores at 19 mm were significantly lower than at greater insertion depths, while at 17 mm, scores were significantly lower still. Scores for the 25-, 23-, and 21-mm insertion depths were equivalent, these all being at or very close to ceiling levels.

Analyses of the individual talker data revealed no significant effect of test run for the female talker, but the male talker data did show a significant effect and also a significant insertion depth by test run interaction. Effects of test time are included in Fig. 1. This figure also includes data from the initial practice session during which only the 21-mm simulated insertion depth condition was employed.

B. Intervocalic consonants

Accuracy in consonant identification as a function of simulated insertion depth is shown for each talker in Fig. 2. Almost all of the errors in consonant identification involved confusions of place of articulation rather than errors of voicing or manner. A repeated measures ANOVA with factors of insertion depth, talker, vowel context, and test run showed main effects of insertion depth, context, and test run, but not of talker. There were also several significant interactions involving vowel context. The vowel by insertion depth interaction was highly significant [$F(6.65,46.6)=15.1$, $p<0.001$]. The talker by vowel interaction [$F(2,14)=14.6$, $p=0.026$] and the vowel by talker by insertion depth interaction [$F(7.05,49.3)=2.94$, $p=0.012$] also reached significance. Hence, separate ANOVAs were performed for each vowel context, using factors of insertion depth, talker, and test run.

Scores for each vowel context are displayed in Fig. 3. With the exception of a modestly significant interaction between insertion depth and talker for the /ɑ/ vowel context, only main effects were significant in these three subanalyses. The significant terms in each case are shown in Table III. Insertion depth was a highly significant factor for each vowel context. Talker significantly affected scores only for the /ɑ/ vowel. There was a small but significant effect of test run for the /ɑ/ and /i/ contexts (of 6.4% in each case), but this effect narrowly missed significance for the /u/ context ($p=0.058$, mean difference of 4.6%).

The insertion depth by talker interaction seen for the /ɑ/ vowel context is illustrated in Fig. 4. Insertion depth had a similar effect on accuracy for both talkers except at 25 mm. Here it seems that the loss of information from the highest band present for the 23-mm insertion depth (around 5923 Hz) led to a decline in performance with the female speech only.

The effects of simulated insertion depth, while strongly significant for each of the three vowel contexts, showed notable differences between vowel contexts (see Fig. 3). These have been examined in detail using *a priori* contrasts based on the separate ANOVAs for each vowel context. For the /i/ context, accuracy varied relatively little with insertion depth compared to the other contexts. There were nevertheless significant differences. Scores at the insertion depth of 25 mm were significantly lower than those at 23 mm, while the 23 and 21 mm scores were statistically equivalent. Scores at 19 and 17 mm were both significantly lower than at 21 mm,

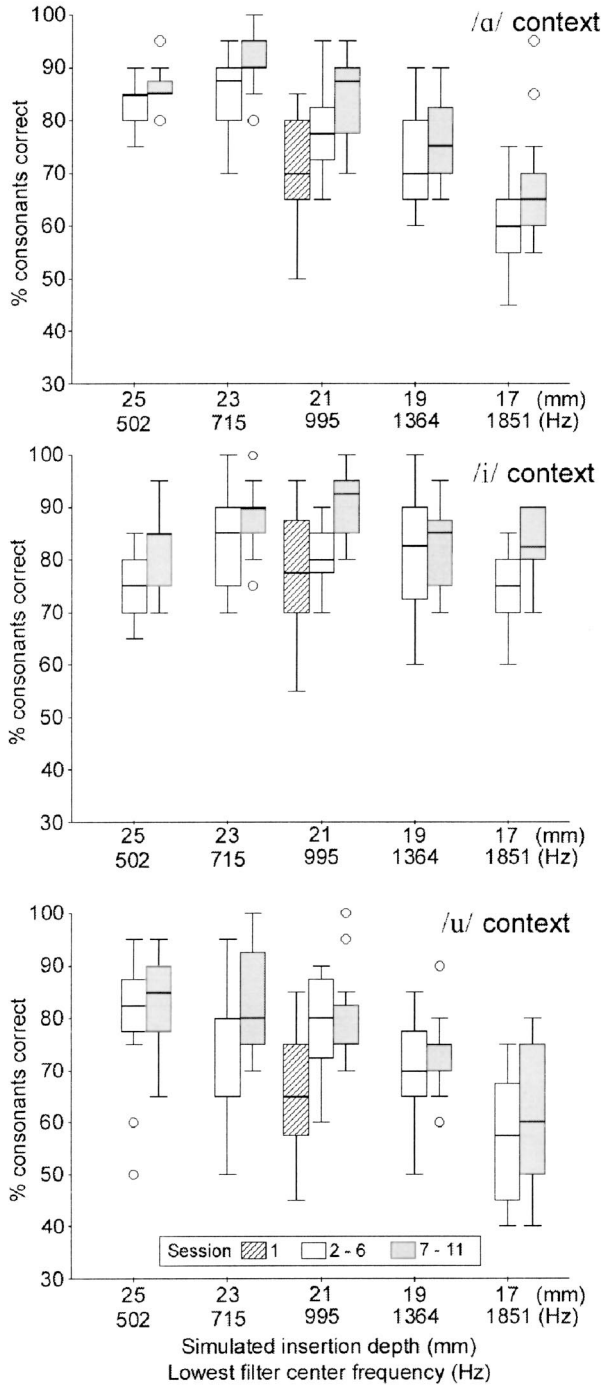


FIG. 2. Percentage correct consonant identification as a function of simulated insertion depth. Scores over both talkers are shown as box and whisker plots for each of the three vowel contexts in the three panels. Scores from the two test runs are shown separately. Scores from the practice session at the 21-mm insertion depth are also shown.

while 19- and 17-mm insertion depths showed equivalent scores. For the /ɑ/ vowel context, *a priori* contrasts showed significant differences between each successive pair of simulated insertion depths, with accuracy at the 23-mm insertion depth being highest. With the /u/ context, the 25-, 23-, and 21-mm insertion depth scores were statistically equivalent, while basal shifts to 19 and 17 mm each led to significant decreases in performance.

A general outcome that holds for each of the three vowel

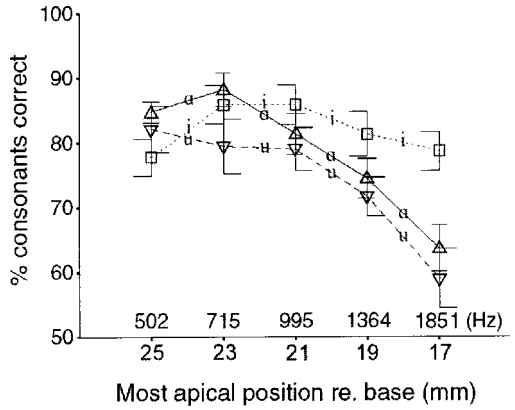


FIG. 3. Vowel context by insertion depth interaction for consonant identification. Average scores over both talkers are shown for each of the three vowel contexts. □ symbol and leftward error bars for /i/; △ symbol and centered error bars for /ɑ/; ▽ symbol and rightward error bars for /u/. Error bars show 95% confidence limits.

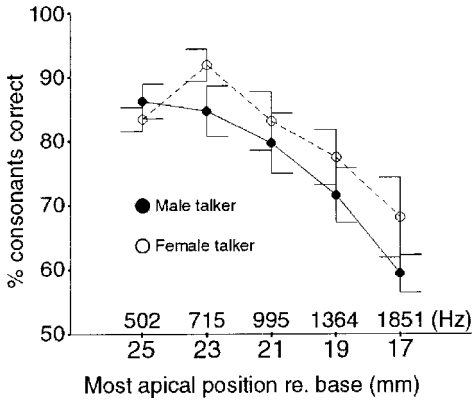


FIG. 4. Insertion depth by talker interaction for /ɑ/ vowel context. Malc talker scores are shown with solid symbols and right-facing error bars, and female talker scores with unfilled symbols and left-facing error bars. Error bars represent 95% confidence limits.

contexts was that consonant identification accuracy declined significantly at insertion depths of 19 mm or less from the cochlear base compared to more apical positions.

C. Vowel identification

Repeated measures ANOVA of accuracy in vowel identification showed all main effects (insertion depth, talker, and test run) to be significant. As in the sentence data, there were also strong interactions involving the talker factor, these being the talker by insertion depth term [$F(4,28)=34.5$, $p<0.001$] and the talker by insertion depth by test run term [$F(4,28)=6.31$, $p<0.001$]. Consequently, subanalyses were performed for each of the two talkers. The significant terms in these subanalyses are summarized in Table IV.

Performance across the simulated insertion depths for each talker is shown in Fig. 5. Planned contrasts based on the ANOVA of the male talker data showed that scores dropped significantly with each basal shift from the 23-mm insertion depth, while the 25 and 23 mm scores did not differ significantly. For the female talker, each successive basal shift in insertion depth produced a significant drop in performance.

Practice effects were significant for both talkers, although relatively small, 5.7% for the male talker and 8.2% for the female talker. The interaction found between insertion depth and test run for the female talker (see Table III) is primarily due to a larger test run effect (of 17%) at the 17-mm insertion depth than at other positions.

TABLE III. Significant factors in ANOVAs of consonant identification accuracy for each vowel context.

Vowel context	Factor	*df*	*F*	*p*	$\eta 2$
/ɑ/	Simulated insertion depth	4, 28	53.5	<0.001	0.88
	Talker	1,7	18.9	0.0034	0.73
	Test run	1,7	26.5	0.0013	0.79
	Simulated insertion depth by Talker	4, 28	3.12	0.0306	0.31
/i/	Simulated insertion depth	4, 28	8.42	<0.001	0.55
	Test run	1,7	46.7	<0.001	0.87
/u/	Simulated insertion depth	4, 28	43.9	<0.001	0.86

IV. DISCUSSION

These simulations of cochlear electrode insertion depth show clear general trends for all of the speech materials used, as illustrated in Fig. 6. If speech processors are set up so that analysis filters are matched to CF at electrode positions, it appears that insertion depths of 19 mm or less will give significantly poorer speech intelligibility than insertions that are 2 to 6 mm deeper. Further, this loss of intelligibility is likely to be greater for male than for female talkers. As insertion depth becomes more shallow, tonotopically mapped processors lose lower frequency information, and add higher frequency information. In terms of useful speech information, however, the added higher frequencies are generally of less value than the lost lower frequencies.

A specific exception to this overall trend of lower intelligibility with shallower insertion occurs where specific speech sounds carry critical cues to their identity in relatively high-frequency regions that remain within the frequency range covered by the insertion depths simulated here. The median F2 for /i/ was measured from the h-vowel-d stimuli used here as 2340 Hz for the male talker and 2624 Hz for the female. For the /ɑ/ vowel, the median F2 was 1130 Hz (male) and 1123 Hz (female), while for /u/, F2 was 1172 Hz (male) and 1437 Hz (female). For the /ɑ/ and /u/ vowels, the shallower insertion depths will remove consonant place cues around the vowel F2. However, the F2 of /i/ is covered by the lowest band of the processors even for the shallowest simulated insertion. Hence, consonant place cues carried by formant transitions involving an /i/ vowel are hardly affected by simulated insertion depth, whereas consonant place cues in /ɑ/ and /u/ vowel contexts are affected by the loss of lower frequencies.

The recent finding that 7 of 19 ''full'' insertions of the Nucleus array were to depths 19 mm or less from the base (Ketten *et al.*, 1998) suggests that relatively shallow insertions may be fairly common in implanted patients. Three of the cases studied by Ketten *et al.* showed the apical electrode to be 17 mm or less from the base of the cochlea. Our simu-

TABLE IV. Significant terms in ANOVAs of vowel accuracy for female and male talkers.

Talker	Source	*df*	*F*	*p*	$\eta 2$
Female	Simulated insertion depth	4, 28	99.4	<0.001	0.93
	Test run	1,7	119	<0.001	0.94
	Simulated insertion depth by test run	4, 28	3.74	0.0146	0.35
Male	Simulated insertion depth	4, 28	146	<0.001	0.95
	Test run	1,7	14.4	0.0067	0.67

lation of a processor that is tonotopically matched to an electrode 17 mm from the base shows a substantial loss of intelligibility compared to insertions that are 21 mm or more from the base. Here, sentence scores fell to just over 70% compared to 100% correct, and vowel scores were below 40% compared to 70% correct.

Vowel identification data based on similar speech processing and manipulation of the presented frequency bands as that used here has been reported by Fu and Shannon (1999a). That study, however, did not investigate conditions that simulate insertion depths of less than 21 mm. That study and the present one agree in both showing a comparable, and relatively modest, decline in vowel identification over a range of simulated insertion depths from 25 to 21 mm from the base.

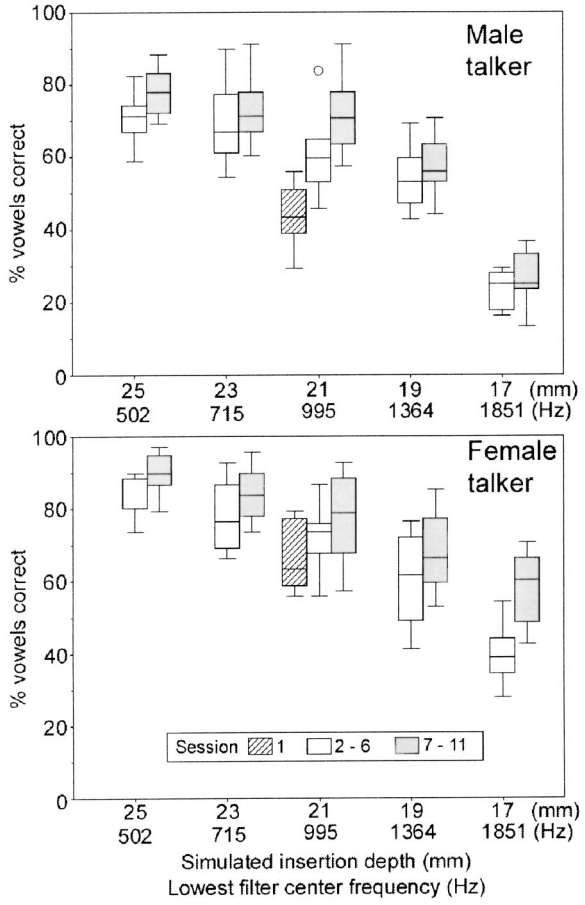


FIG. 5. Box and whisker plots of vowel identification accuracy as a function of simulated insertion depth, degree of test run, and talker. Session 1 was the initial familiarization session, in which only the 21-mm insertion depth was simulated. The first test run at each insertion depth occurred at one of sessions 2–6, while the second occurred at one of sessions 7–11.

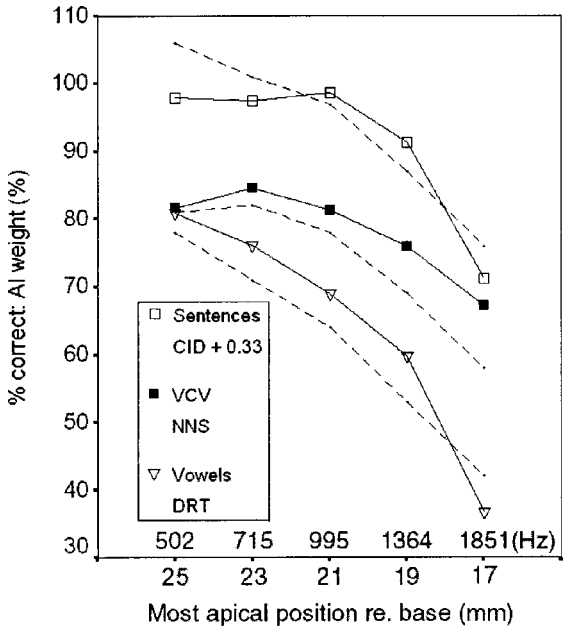


FIG. 6. Summary of effects of simulated insertion depth over speech materials. The symbols represent data: □ sentences, ■ consonants; and ▼ vowels. The dashed lines show predicted relative scores over insertion depth from AI weightings for comparable materials (see text for details).

A. Comparison with predictions from the Articulation Index

The effects of frequency range in each condition here fit fairly closely with predictions based on AI weightings for comparable material, which are included in Fig. 6. Predictions were derived from the ''critical band'' weights for CID sentences, from the NNS nonsense syllables for the consonant data, and from the Diagnostic Rhyme Test (DRT) for the vowel data (ANSI, 1997). A prediction of the relative intelligibility for each simulated electrode array location was derived by summing the critical band weights for those critical bands covered by each of the simulated processors. Where a critical band was not completely covered by one of the extreme processor filters, the critical band weight was reduced in proportion to the relative basilar membrane extent of the critical band that was covered by the processor band. The processor-weighted AI weightings for the NNS and DRT materials are directly displayed in Fig. 6. The processor-weighted AI weights for the CID sentences were increased by 0.33 to align these with the sentence scores that are below ceiling.

While AI weights are based on normal auditory frequency selectivity, the extension of AI proposed in the ANSI SII procedures makes the assumption that a broadening of frequency selectivity will lead to a proportional decrease in the AI weight in the affected frequency band (ANSI, 1997). Our processors represent frequency selectivity based on equal basilar membrane distance for each band, and normal auditory filter bandwidths are also closely related to basilar membrane distance (Greenwood, 1990). As a result the degree of broadening of selectivity compared to normal hearing

is approximately the same for each processor filter band and it would be expected that the AI weighting in each critical band covered by a processor would be equally affected by the broadening of selectivity represented by the processor filters. Hence the relative (but not the absolute) values of the processor-weighted AI weights over frequency should not depend on the degree of selectivity.

B. Effects of training and practice

The study was not designed to track changes in performance with increasing experience of the processed stimuli. There are, however, clear indications of performance increasing with experience. In the case of vowel and consonant identification, accuracy observed in the second half of testing was generally only slightly (5% to 8%) higher than in the first part. For the shallowest simulated insertion depth, however, larger (15% to 20%) increases in performance with experience were seen both in vowel identification for the female talker only, and in the identification of words in sentences. These practice effects suggest that there may be some potential for learning to increase the informativeness of higher frequency speech cues when lower frequencies are removed.

V. CONCLUSION

Although acute simulation studies (Dorman *et al.*, 1997; Fu and Shannon, 1999a; Shannon *et al.*, 1998) suggest that speech processor filters centered below the CFs of electrode locations may be less ideal than filters matched to CFs, the unshifted control conditions employed in those studies have represented simulations of relatively deeply inserted electrodes. Where a patient has an electrode array that does not reach a depth of more than 19 mm from the base, speech intelligibility is likely to be less than ideal whatever fitting approach is taken. If the processor filters are matched to the electrode position CFs, significant low-frequency information will be lost, while additional high-frequency information that is gained by the use of higher filter center frequencies is of little advantage. If, on the other hand, the processor filters are centered at frequencies below these CFs, upward shifting may cause difficulties. However, listeners are able to adapt at least to some extent to upward shifting (Rosen *et al.*, 1999). Further research is required to estimate the costs associated with spectral shifting alongside the possible benefits of making low-frequency information available after listeners have had sufficient experience to adapt to shifting. It may also be important to establish the potential benefits of learning to use higher frequency speech cues. Only then can conclusions be drawn on the expected outcomes of fitting speech processors using shifted and tonopically mapped filter frequency allocations for less than ideal electrode insertion depths.

ACKNOWLEDGMENTS

This work was supported by a Wellcome Trust Vacation Scholarship to Deborah Stanton (VS99/UCL/262) and the RNID. We are grateful to Bob Shannon and an anonymous reviewer for helpful comments on the manuscript, to Philip Loizou for access to MATLAB routines that formed the basis for off-line speech processing, and to Quentin Summerfield and John Foster for assistance in the recording of sentence materials.

[1]These estimated CFs are based on a cochlear length of 33 mm, which was the average of Ketten *et al.*'s (1998) measurements from CT data. Elsewhere in this paper, CFs are always estimated for a 35-mm cochlear length. The accuracy of these stimulation-site estimates is limited to the extent that the stimulated neural tissues may not be those that are normally excited by motion at a particular basilar membrane location. When there is dendritic degeneration, the stimulated auditory fibers may even include those in an adjacent cochlear turn (Frijns *et al.*, 2001).

ANSI (**1997**). American National Standard Methods of the Calculation of the Speech Intelligibility Index, ANSI S3.5-1997 American National Standards Institute, New York.

DeFilippo, C. L., and Scott, B. L. (**1978**). ''A method for training and evaluation of the reception of on-going speech,'' J. Acoust. Soc. Am. **63**, 1186–1192.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,'' J. Acoust. Soc. Am. **102**, 2993–2996.

EPI Group (**1986**). The BKB (Bamford-Kowal-Bench) standard sentence lists. Department of Phonetics and Linguistics, University College London.

Foster, J. R., Summerfield, A. Q., Marshall, D. H., Palmer, L., Ball, V., and Rosen, S. (**1993**). Lip-reading the BKB sentence lists; corrections for list and practice effects, Br. J. Audiol. **27**, 233–246.

French, N. R., and Steinberg, J. C. (**1947**). ''Factors governing the intelligibility of speech sound intelligibility,'' J. Acoust. Soc. Am. **19**, 90–119.

Frijns, J. H. M., Briaire, J. J., and Grote, J. J. (**2001**). ''The importance of human cochlear anatomy for the results of modiolus-hugging multichannel cochlear implants,'' Otol. Neurotolo. **22**, 340–349.

Fu, Q. J., and Shannon, R. V. (**1999a**). ''Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing,'' J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q. J., and Shannon, R. V. (**1999b**). ''Effects of electrode location and spacing on phoneme recognition with the nucleus-22 cochlear implant,'' Ear Hear. **20**, 321–331.

Fu, Q.-J., Shannon, R. V., and Galvin, J. J. (**2002**). ''Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant,'' J. Acoust. Soc. Am. **112**, 1664–1674.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Harnberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., and Meyer, T. A. (**2001**). ''Perceptual 'vowel spaces' of cochlear implant users: Implications for the study of auditory adaptation to spectral shift,'' J. Acoust. Soc. Am. **109**, 2135–2145.

Ketten, D. R., Vannier, M. W., Skinner, M. W., Gates, G. A., Wang, G., and Neely, J. G. (**1998**). ''In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays,'' Ann. Otol. Rhinol. Laryngol. **107**, S175,1–16.

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

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). ''Perceptual adaptation by normal listeners to upward shifts of spectral information in speech and its relevance for users of cochlear implants,'' J. Acoust. Soc. Am. **106**, 3629–3636.

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

Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.

Reed vibration in lingual organ pipes without the resonators

András Miklós,[a] Judit Angster,[b] and Stephan Pitsch
Fraunhofer-Institut für Bauphysik, Nobelstrasse 12, D-70569 Stuttgart, Germany

Thomas D. Rossing[c]
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

(Received 12 January 2002; revised 31 October 2002; accepted 5 November 2002)

Vibrations of plucked and blown reeds of lingual organ pipes without the resonators have been investigated. Three rather surprising phenomena are observed: the frequency of the reed plucked by hand is shifted upwards for large-amplitude plucking, the blown frequency is significantly higher than the plucked one, and peaks halfway between the harmonics of the fundamental frequency appear in the spectrum of the reed velocity. The dependence of the plucked frequency on the length of the reed reveals that the vibrating length at small vibrations is 3 mm shorter than the apparent free length. The frequency shift for large-amplitude plucking is explained by the periodic change of the vibrating length during the oscillation. Reed vibrations of the blown pipe can be described by a physical model based on the assumption of air flow between the reed and the shallot. Aerodynamic effects may generate and sustain the oscillation of the reed without acoustic feedback. The appearance of subharmonics is explained by taking into account the periodic modulation of the stress in the reed material by the sound field. Therefore, a parametric instability appears in the differential equation of vibration, leading to the appearance of subharmonics. [DOI: 10.1121/1.1534101]

PACS numbers: 43.75.Np, 43.75.Pq, 43.25.Ts, 43.28.Ra [NHF]

I. INTRODUCTION

The pipe organ has been called the ''king of musical instruments.'' No other instrument can match it in size, range of tone, loudness, or complexity. No two pipe organs in the world are exactly alike. There are two basic types of organ pipes: flue (labial) pipes and reed (lingual) pipes. Flue pipes produce sound by means of a vibrating air jet, in a manner similar to the flute or recorder, while reed pipes use a vibrating metal reed to modulate the air stream. Although sound production in labial pipes has been studied fairly extensively,[1–10] relatively little research has been reported on sound production in lingual pipes.[11–14] In these pipes, there is a strong and rather complicated interaction between the vibrating reed and the pipe resonator.

In most lingual pipes the reed (tongue) vibrates against the fixed shallot. The foot (boot) of a typical reed pipe is shown in Fig. 1. Air under pressure from the windchest flows through the bore (foot hole) into the boot and through the thin opening into the shallot. The wind sets the thin, flexible tongue into vibration; this in turn modulates the flow of air passing through the shallot into the resonator. The reed is pressed against the open face of the shallot by a tuning wire that can be adjusted up and down to tune the vibrating reed. The pipe voicer generally adjusts the reed and the resonator to produce the best sound, and tunes the pipe by adjusting the tuning wire in several subsequent steps.

According to the experience of organ building, the pitch and timbre of the pipe can be affected by many factors.[15,16] Although reed pipes speak without their resonator, certain partials or partial groups can be reinforced and the loudness can be increased by applying a proper resonator. The sound of the reed pipe is strikingly affected by the relationship between tongue length and resonator length. Natural ''full length'' cylindrical resonators correspond roughly in length to stopped flue pipes of the same pitch, while ''full length'' conical resonators are somewhat longer; the ''resonance length'' is about three-quarters of the length of a corresponding open flue pipe. Reed pipes with cylindrical resonators have mainly odd-numbered harmonics, while the sound of a reed pipe with a conical resonator contains all harmonics.

The timbre of reed pipes can also be affected by the tongue (reed) and shallot. The thinner the reed the richer is the sound in harmonics; the thicker the reed, the smoother and more fundamental the sound. Wider cylindrical resonators produce stronger tone, while the sound of pipes with narrow resonator tubes is weaker and has a character similar to that of reed woodwind instruments. Helmholtz resonators are used in some families of lingual organ pipes. A broad variation in shape and size of the resonator can be found in contemporary organ building.[15]

In order to facilitate the attack (speech), the reeds of lingual organ pipes are curved. During the prevoicing procedure the voicer, holding a metal rod, applies a compression force on the reed, which lies on a flat metal surface. He moves the rod slowly towards the end of the reed and gradually increases the force. By repeating this procedure (called ''polishing'' by organ builders) several times, the reed will be curved up. The resulting curvature is quite even, although it may increase towards the end of the reed. Because of this curvature there is a gap between the reed and the shallot, which cannot be closed completely when the reed is pressed against the shallot by the wind pressure in the foot. There-

[a]Electronic mail: andreas.miklos@urz.uni-heidelberg.de
[b]Electronic mail: angster@ibp.fhg.de
[c]Electronic mail: rossing@physics.niu.edu

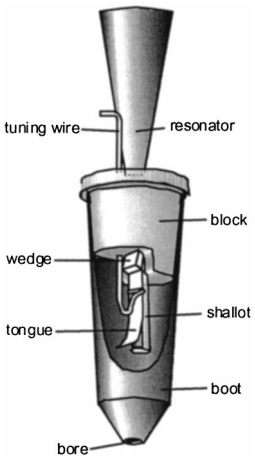


FIG. 1. The foot (boot) of a reed pipe.

fore, the air flow between foot and shallot is never interrupted entirely.

Lingual organ pipes have been recently investigated[17–19] and several features were observed. In order to understand better how sound is produced in reed organ pipes, more detailed experiments were carried out. The results of the measurements of reed vibrations without resonators and theoretical considerations are presented in this paper. Reed vibrations of pipes with resonators and sound generation in reed pipes will be discussed in a second paper.

II. EXPERIMENTAL METHOD

In the experiments reported in this paper, reed pipes were tested on a slider chest, with the wind supplied by an organ blower and pressure regulator (bellows). Wind pressure was measured on a water manometer. A window was installed on the regular pipe boot, so that the reed could be observed while it vibrated. The reed velocity was recorded by means of a laser vibrometer (Polytec, OFV 3000). A $\frac{1}{4}$-in. microphone (B&K 4135) was inserted into the shallot wall to record the sound pressure inside the shallot, and another microphone (B&K 4165) recorded the sound pressure near the open end of the pipe.

The fundamental frequency of the free vibration of the reed was measured by plucking the reed and measuring the velocity with the laser vibrometer. The reed did not strike against the shallot in this experiment; it was vibrating freely. The frequency and damping were determined from the decaying velocity signal.

The reed velocity and the sound pressure inside the shallot, as well as the sound pressure near the open end of the pipe, were recorded as the vibrating reed length was varied by means of the tuning wire. The wind pressure was set to the optimal value (80-mm water) for sounding the pipe, as recommended by an experienced organ voicer. Sound pressure and reed velocity waveforms were recorded both with and without the pipe resonator in place. Sound spectra, velocity spectra, and frequency were obtained with the help of a dual-channel FFT analyzer (HP 35670A).

A G_4 trumpet pipe was used for the study presented in this paper. The conical resonator of the pipe was 687 mm long with diameters 9.2 and 65 mm at the ends (conical angle: 4.7 deg). The shallot was also slightly conical (internal diameters: 6.5 mm at upper end and 8 mm at lower end), with a length of 56.5 mm and a wall thickness of 1.5 mm. The opening on the face of the shallot has a triangular shape with 4.8-mm maximal width and 40.4-mm height. The bronze tongue (reed) has a trapezoidal shape with length of 55.8 mm and widths of 7.2 and 5.5 mm at the ends. The thickness of the reed is 0.31 ± 0.005 mm. The smaller end was clamped by means of a wedge into the block (head, nut) of the pipe. The total length of the reed from the wedge to the free end was 45 mm. The vibrating length could be adjusted from 37.5 to 15 mm.

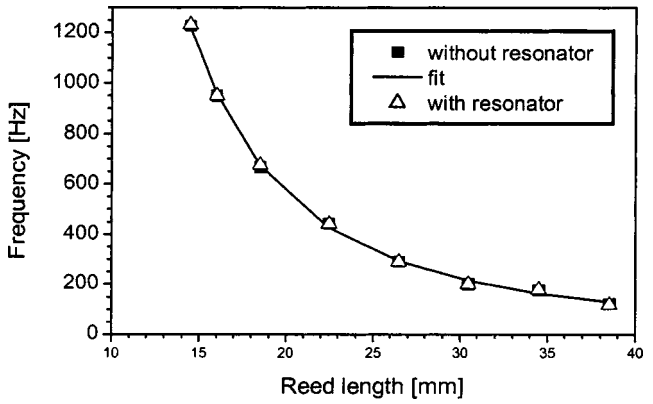


FIG. 2. Frequency versus reed length for a plucked reed with and without a resonator attached.

III. EXPERIMENTAL RESULTS

A. Plucked reed

In the first experiment the reed was plucked by hand and the resulting reed velocity waveforms were measured by the laser vibrometer. In case of plucking with a small amplitude, the velocity waveform, disregarding the first two-three periods, is an exponentially decaying sine signal. Frequencies and Q factors are determined from the exponentially decaying waveforms. Attaching the resonator does not influence the plucked frequency of the reed, but the Q factor decreases about 20% (for example, from $Q=16$ to $Q=12.4$ at 200.8 Hz), probably due to air mass loading. The frequency vs reed length curves for plucked reed with and without a resonator in Fig. 2 are practically the same.

If the reed is plucked with large initial amplitude, the velocity waveform is more complicated. The frequency increases as the amplitude of the vibration decreases. This effect was observed in an earlier experiment,[17] where the frequency determined from the period of the second cycle was 247 Hz, whereas it has increased to 272 Hz by the 16th cycle of the decaying velocity waveform.

B. Blown reed pipe without resonator

In this case the resonator is not attached to the boot of the pipe. The investigated system, as shown in Fig. 1, consists of the boot, the shallot inserted into the block (head, nut), and the reed, attached to the shallot. This system, when blown, produces quite a strong sound, whose frequency can

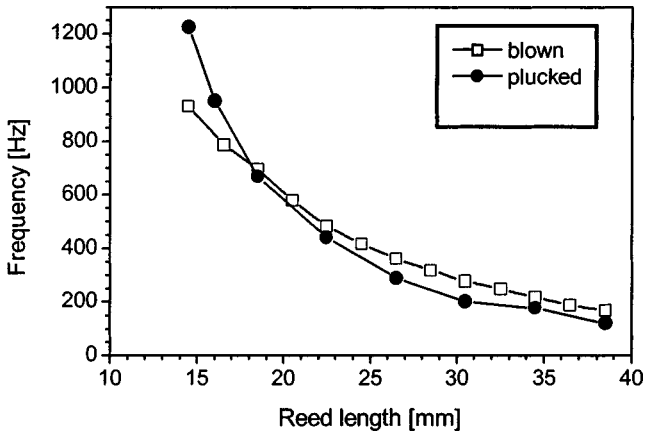


FIG. 3. Frequency versus reed length for plucked and blown reeds without a resonator.

be tuned continuously by the tuning wire. Reed vibration waveforms and spectra, as well as sound pressure in the shallot, are measured for several different positions of the tuning wire. The effect of the wind pressure on the frequency of the blown reed is also investigated.

Plucked and blown frequencies for different reed lengths are shown in Fig. 3. Blown frequencies are somewhat higher than that of the plucked reed for the longer lengths. For a vibrating length of 30.5 mm, for example, a frequency of 202 Hz is measured for the plucked reed, whereas the frequency of the blown reed is 278 Hz, almost 40% higher. Plucked and blown frequencies are equal at 18-mm length, while for shorter reeds the plucked frequency is larger than the blown one.

Waveforms of reed velocity and acceleration are shown in Fig. 4 for a reed length of 37.5 mm and a wind pressure of

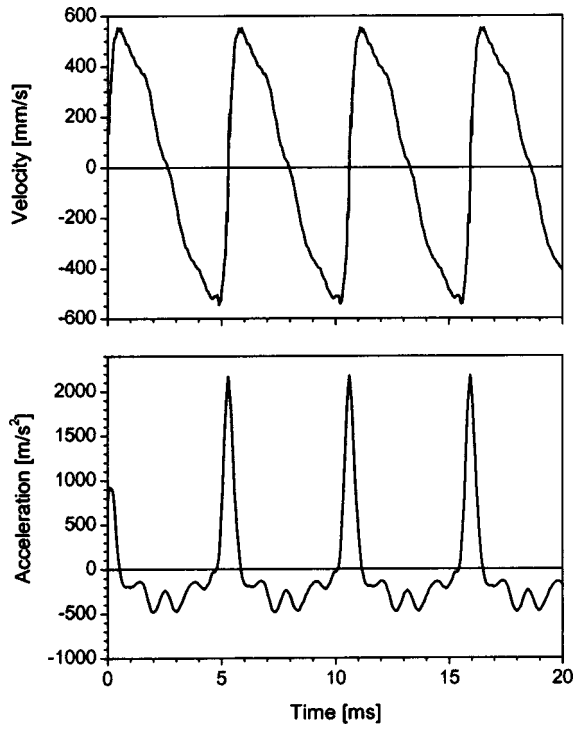


FIG. 4. Velocity and acceleration waveforms for a blown reed ($l = 30.5$ mm) without a resonator.

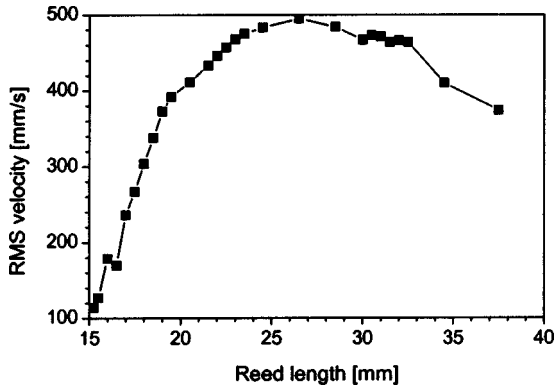


FIG. 5. rms value of reed velocity versus reed length for a blown reed without a resonator.

80-mm water without a resonator. The velocity waveform has a sawtooth-like profile. Positive velocity values correspond to a movement towards the laser vibrometer, i.e., outwards from the shallot. Velocity zeros show the turning points of the vibration. Zero crossing of the steep slope shows the inner turning point close to the shallot, while the other zero crossing corresponds to the outer turning point of the vibration, i.e., to the maximal displacement of the reed. There is no indication of complete closure of the shallot opening, which would appear as a short horizontal part (zero velocity) in the waveform.

The acceleration of the reed is determined by calculating the time derivative of the velocity waveform. The acceleration (see Fig. 4), which is proportional to the force acting on the tongue, shows a sharp positive peak, corresponding to a strong restoring force. The smaller but wider negative acceleration (force) corresponds to a slowly accelerating movement of the reed towards the shallot.

The velocity waveform approximates a sine wave as the frequency of the vibration increases. The magnitude of the vibration is estimated by calculating the root-mean-square (rms) velocity. The rms values as functions of the reed length can be seen in Fig. 5. The curve has a broad maximum around 25-mm reed length. For shorter reeds the rms value decreases quickly with decreasing reed length.

The blown frequency increases with increasing wind pressure. The frequency of the reed with $L = 29.5$ mm length is determined at different pressures as measured at the foot of the pipe. The frequency versus wind pressure plot is shown in Fig. 6.

The data reported in this paper are from a single G_4 trumpet pipe, but similar effects have been observed recently in other reed pipes as well.[17–19]

IV. THEORETICAL CONSIDERATIONS AND COMPARISON WITH RESULTS

A. Plucked reed

Although the investigated reed has a slightly trapezoidal shape, it will be regarded as a cantilever with clamped and free ends in the following discussion. The differential equation of a lossless cantilever is given as[20]

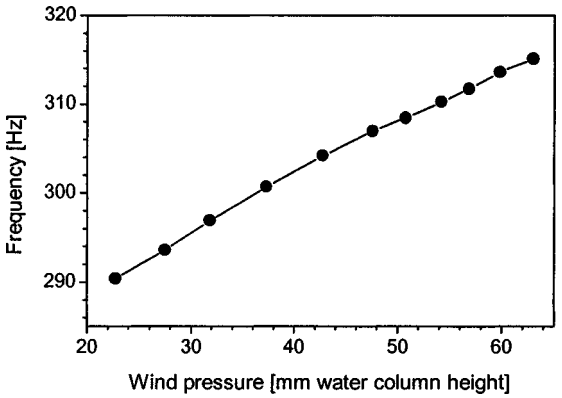


FIG. 6. Frequency versus pressure (water column height) for a blown reed ($l=29.5$ mm) without a resonator.

$$\rho blh\frac{\partial^2\xi(z,t)}{\partial t^2}+\frac{Eblh^3}{12}\frac{\partial^4\xi(z,t)}{\partial z^4}=F(z,t), \tag{1}$$

where ρ, E, b, l, h, ξ, z, and F are the density, Young's modulus, width, length, thickness, and displacement of the reed, the spatial variable, and the force acting on the reed, respectively. The resonance frequency of the plucked reed can be determined by solving Eq. (1) with $F=0$. Assuming $\exp(\pm k_n z+i\omega t)$ or $\exp(\pm ik_n z+i\omega t)$ dependence, the following equation can be derived for the angular frequency ω_n:

$$\omega_n^2=\frac{Eh^2}{12\rho}k_n^4=\frac{\alpha_n^4h^2}{l^4}\frac{E}{12\rho}, \tag{2}$$

where α_n is a numerical constant. Its value can be determined from the boundary conditions.[20] For the fundamental vibration $\alpha_1=1.8748$.

The frequency of the plucked reed vs reed length is shown in Fig. 2. Since the fundamental resonance frequency of the reed is inversely proportional to its length squared [see Eq. (2)], a function of the form of

$$f=\frac{A}{(l-\kappa)^2}, \tag{3}$$

was fitted to the measured frequencies. The fit with $\kappa=3$ mm is excellent (see the solid line in Fig. 2), showing that the effective vibrating length of the reed is about 3 mm shorter than the distance between the tuning wire and the free end of the reed. That is, a 3-mm length of the reed near the wire is pressed so hard against the shallot that it does not vibrate. However, this ''end point'' is not fastened mechanically. Therefore, it may be expected that the vibrating length decreases when the reed moves towards the shallot and increases when the reed moves outwards. This effect may be the reason for the observed shift of the frequency at large-amplitude plucking. This assumption can be investigated as follows.

Assume that the displacement can be written as a product of a time-dependent and a z-dependent function

$$\xi(z,t)=\psi(t)\varphi(z). \tag{4}$$

The function φ(z) must fulfill the boundary conditions of a clamped-free cantilever,[20] i.e., $\varphi(0)=0$, $\varphi^{\mathrm{I}}(0)=0$, $\varphi^{\mathrm{II}}(l)=0$, and $\varphi^{\mathrm{III}}(l)=0$. Moreover, it will be normalized in such a way that its value at the free end equals unity: $\varphi(l)=1$. This function for the fundamental vibration can be written as

$$\varphi(z)=0.5\left[\cosh\left(\alpha_1\frac{z}{l}\right)-\cos\left(\alpha_1\frac{z}{l}\right)\right]-0.37\left[\sinh\left(\alpha_1\frac{z}{l}\right)-\sin\left(\alpha_1\frac{z}{l}\right)\right].$$

The function $\psi(t)$ can be regarded, then, as the time-dependent displacement of the free end of the reed. By substituting Eq. (4) into Eq. (1), a second-order ordinary differential equation can be derived for $\psi(t)$

$$\frac{d^2\psi(t)}{dt^2}+\frac{Eh^2}{12\rho}\frac{\alpha_1^4}{l^4}\psi(t)=0. \tag{5}$$

Assume that the length of the reed depends on the displacement in the following way:

$$l=l_0+\varepsilon\psi(t)=l_0\left(1+\frac{\varepsilon}{l_0}\psi(t)\right), \tag{6}$$

where the second term in the bracket is much smaller than unity. Then, $1/l^4$ can be approximated to the second order as

$$\frac{1}{l^4}=\frac{1}{l_0^4}\left(1-\frac{4\varepsilon}{l_0}\psi(t)-\frac{6\varepsilon^2}{l_0^2}\psi^2(t)\right). \tag{7}$$

By substituting Eq. (7) to Eq. (5), a nonlinear differential equation can be derived

$$\frac{d^2\psi(t)}{dt^2}+\omega_0^2\left(\psi(t)-4\frac{\varepsilon}{l_0}\psi^2(t)-6\frac{\varepsilon^2}{l_0^2}\psi^3(t)\right)=0. \tag{8}$$

Equation (8) can be solved by the method of successive approximations.[21] The first approximation is the solution of the linear equation $\psi_1(t)=a\cos(\omega_0 t)$. The solution of Eq. (8) to second order in the small variable $\varepsilon a/l_0$ can be written as[21]

$$\psi(t)=a\left[2\frac{\varepsilon a}{l_0}+\cos\omega t-\frac{2}{3}\frac{\varepsilon a}{l_0}\cos 2\omega t+\frac{1}{6}\left(\frac{\varepsilon a}{l_0}\right)^2\cos 3\omega t\right], \tag{9}$$

where

$$\omega=\omega_0\left[1-\frac{107}{12}\left(\frac{\varepsilon a}{l_0}\right)^2\right]. \tag{10}$$

The small variable $\varepsilon a/l_0$ is essentially the relative change of the length $\Delta l/l_0$. The nonlinear terms in Eq. (8) cause a decrease of the frequency and the appearance of higher harmonics. The frequency shift is proportional to the square of the amplitude.

Equation (10) can explain the observed behavior of the plucked frequency at large-amplitude plucking. Since the amplitude a decreases during the slow decay of the reed vibration, the frequency, in accordance with Eq. (10), shifts upwards. From the observed frequency shift of 25 Hz at 272-Hz fundamental frequency, the value of 0.1 can be calculated for $\varepsilon a/l_0$ from Eq. (10). This value corresponds to ~10% length change during the decay process.

B. Blown pipe without a resonator

At a first glance it is quite surprising that the plucked reed frequency is much lower than the blown frequency. Moreover, the measurements in Fig. 6 show that the vibration frequency of the reed increases with increasing pressure. Both phenomena can be explained by assuming that the pressure in the boot reduces the free vibrating length of the reed. It was shown in the previous section that the free vibrating length of the plucked reed is already smaller than the distance between the tuning wire and the free end of the reed. The equilibrium position of the plucked reed without the boot pressure is determined by the curvature given by the curving process mentioned in the Introduction. The boot pressure pushes the reed towards the shallot to a new equilibrium position. In this position the part of the reed which cannot vibrate will be somewhat bigger than the value of the fit parameter κ for a plucked reed. Therefore, the vibrating length will be smaller, and the blown frequency higher than that of the plucked reed. Since the frequency is inversely proportional to the square of the vibrating length, quite small length changes could be expected. The relative change of the vibrating length calculated from the measured frequencies between pressures of 63- and 22-mm water column would be less than 5%.

This would be the simplest possible explanation of the observed increase of the blown frequency compared to the plucked one. However, another physical effect, the stiffening of the reed through aerodynamic forces, could also raise the blown frequency. This effect will be discussed later.

The reed generator in the lingual organ pipes is composed of the reed and the shallot. The basic operation of the system can be described by the model given by Fletcher and Rossing (Ref. 2, Chap. 13). The pressure in the boot tends to blow the reed closed,[11] but, due to the special shape of the reed, the pressure in the boot cannot close the opening between the reed and the shallot completely, and a thin gap always remains. Since the boot pressure is usually quite small, the displacement of the reed is also small between the equilibrium positions without and with boot pressure.

The displacement of the free end of a clamped-free cantilever due to a distributed force can be written as[20]

$$\frac{\xi_0}{h} = \frac{3p_0}{2E}\left(\frac{l}{h}\right)^4, \tag{11}$$

where p_0 is the force per unit area of the cantilever. In the case of a blown reed p_0 is the difference between wind pressures in the boot of the pipe and in the shallot. Since $p_0 \approx 800$ Pa, $E \approx 10^{11}$ Pa, $h \approx 0.3$ mm, and $l \approx 30$ mm in our experiments the shift of the equilibrium position is about $\xi_0 \approx 1.2\ \mathrm{h} \approx 0.36$ mm, smaller than the gap between the reed and shallot without boot pressure, i.e., the boot pressure cannot close the reed. The gap between the new equilibrium position and the shallot is very small, less than 1 mm.

The vibration of the reed around its fundamental frequency can be modeled by a mass–spring system. The differential equation of motion can be derived from Eq. (5) by adding a loss term to the equation. The resulting differential equation can be written as

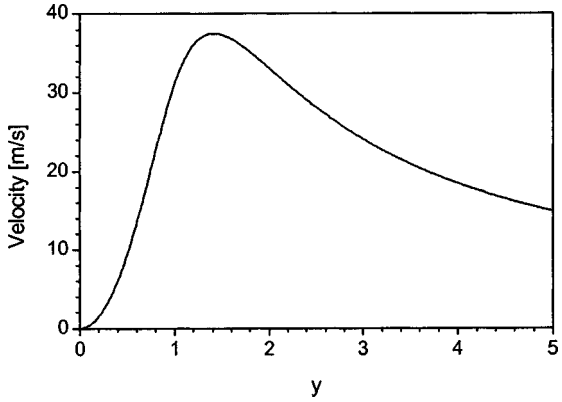


FIG. 7. A typical calculated flow velocity versus gap-width curve. Gap width y is dimensionless.

$$m\frac{d^2\psi(t)}{dt^2} + 2\beta m\frac{d\psi(t)}{dt} + K\psi(t) = F, \tag{12}$$

where ψ, m, β, K, and F are the displacement of the free end of the reed around the equilibrium position, reed mass, attenuation constant, spring constant (stiffness), and driving force, respectively. The mass m and spring constant K can be given as

$$m = \rho bhl \quad \text{and} \quad K = 1.03Eb\left(\frac{h}{l}\right)^3, \tag{13}$$

where ρ, b, h, l, and E are the density, width, thickness, and length of the reed and the Young's modulus, respectively.

The driving force F on the right side of Eq. (12) has two components: the aerodynamic force due to the air flow through the slit between the reed and shallot, and the acoustic force produced by the sound fields in the boot and shallot. It will be shown in the following discussion that the aerodynamic force alone is sufficient for producing a self-sustained oscillation of the reed.

C. Reed oscillation due to the aerodynamic force

The slit between the reed and the shallot is very narrow. For such a slit the flow could be laminar up to about 30–40-m/s flow velocity. Since the flow velocity through the slit would probably be smaller, it can be assumed that the flow between the reed and shallot is laminar. However, the flow through the foothole and the outflow from the shallot should be described by the Bernoulli equation. For stationary flow the volume velocity at the foothole, slit, and shallot opening is equal. Under this condition the flow velocity w in the slit can be determined [see the Appendix, Eq. (A13)]. The details of the calculation can be found in the Appendix.

The dependence of the flow velocity on the dimensionless gap variable y, defined in Eq. (A14), can be seen in Fig. 7. As mentioned in the Appendix, the velocity curve has a maximum at $y = \sqrt{2}$. For small y values the velocity scales approximately with y^2, as expected for a laminar flow. As y increases, the second term under the square root of Eq. (A13) will dominate; therefore, the velocity curve should approach a $1/y$ dependence. The physical reason for the decrease of the velocity is the decreasing pressure difference between the boot and shallot at larger gaps.

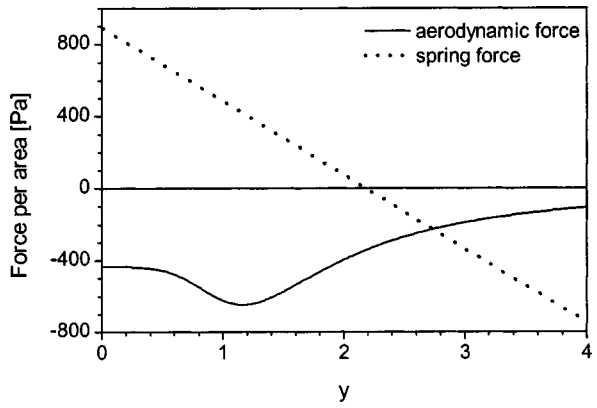


FIG. 8. Calculated aerodynamic and spring forces acting on the reed. Gap width y is dimensionless.

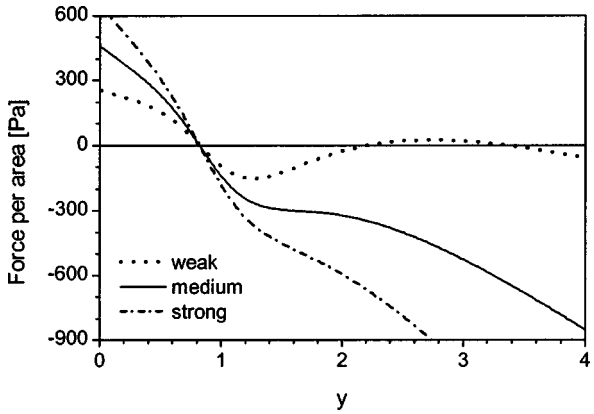


FIG. 9. The resultants of the aerodynamic and spring forces for weak, medium, and strong reeds. Gap width y is dimensionless.

The velocity shown in Fig. 7 was calculated for a constant gap. In the case of an oscillating reed the gap changes during the travel time of a fluid particle through the slit. This effect modifies the velocity [see Eq. (A16)]. The modified velocity was used for calculating the $f(y)$ and $g(y)$ components of the aerodynamic force [see Eqs. (A19)–(A21)]. Both components are strongly nonlinear.

The differential equation of the blown reed can be derived by substituting the displacement ψ for the change of the gap width $x-x_P$ (x_P is the equilibrium position of the plucked reed) and the driving force F from Eq. (A19)

$$m\frac{d^2x}{dt^2}+2\beta m\frac{dx}{dt}+K(x-x_P)=f(y)-g(y)\frac{dy}{dt}. \qquad (14)$$

Equation (12) can be written in a more convenient form by substituting the dimensionless gap variable y in the left side of the equation and by rearranging the loss and force terms

$$\frac{ma}{\sqrt{2}}\frac{d^2y}{dt^2}+\left[\frac{2\beta ma}{\sqrt{2}}+g(y)\right]\frac{dy}{dt}$$
$$=\left[-\frac{Ka}{\sqrt{2}}(y-y_P)+f(y)\right], \qquad (15)$$

where a is the gap at maximum flow velocity. It can be seen from Eq. (15) that $f(y)$ corresponds to a force, while $g(y)$ corresponds to a loss or gain. The expression on the right side of the equation is the sum of the elastic restoring force and the aerodynamic force.

The calculated aerodynamic force $f(y)$ per unit area of the reed vs gap size is shown in Fig. 8. The sign of this force is negative, i.e., it would drive the reed towards the shallot. The spring force of a reed of medium strength is also shown in the figure, assuming an original equilibrium position of the plucked reed at $y_P=2.65$. This force is positive for $y<2.65$ and proportional to the dimensionless displacement $y-y_P$. Depending on the elasticity of the reed material, the reed geometry, and the wind pressure, a broad variation of force curve profiles is possible. The resultants of the aerodynamic and spring forces are shown in Fig. 9 for reeds having different stiffness ($K_W/K_M=0.39$, $K_S/K_M=1.56$, $y_P=5.15$, $y_P=2.04$ for weak and strong reeds, respectively). It can be seen that the shape of the resultant force is very different for weak, medium, and stiff reeds. In all cases a strong restoring force appears as the gap approaches zero. This force impedes closing the gap entirely. The force curve crosses the y axis at least in one point, but three zero crossings (or one zero crossing and one tangential point) are also possible for weak reeds. Positions where the resultant force is zero are the new possible equilibrium positions of the reed when wind pressure is applied.

In the cases of medium and stiff reeds the zero crossings can be regarded as the new equilibrium positions of the reed. As the force is negative for bigger y values, the applied pressure will drive the reed from the original equilibrium position (y_P) to the new position. Since this position is stable, the reed may oscillate around the new equilibrium. However, the force curve is nonlinear, and quite asymmetric to this position. For example, the reed with medium stiffness in Fig. 9 would provide a small negative force in the entire domain between the old and new equilibrium positions; therefore, the reed would approach the new equilibrium quite slowly. That is, the onset of the oscillation would be delayed after applying the pressure. Such behavior is very common in organ reed pipes.

The curve of the weak reed with three zero crossings in Fig. 9 can describe the response of certain reed pipes for slowly increasing pressure. In this case the reed will move from the original equilibrium position to the first zero from the right in Fig. 9. This position is stable; therefore, the reed remains in this position. It may respond with small oscillations for small disturbances. However, these oscillations may drive the reed over the local maximum of the curve, bringing it into the domain of negative force, which will drive the reed to the equilibrium position closest to the shallot. The investigated reed pipe must have this type of force curve, because the observed behavior corresponds to that described above. In the case of slow valve opening the pipe did not sound, but after quite a long time (~1–2 min) a slow and quite uncertain onset of the sound could be observed. The pipe responded normally, however, for a sudden onset of the pressure, because the reed was then driven over the first and second zero crossings by the first pressure stroke.

Stiff reeds cannot be easily influenced by the aerodynamic forces. It can be observed in Fig. 9 that the shape of the resultant force is quite close to that of a linear force. The

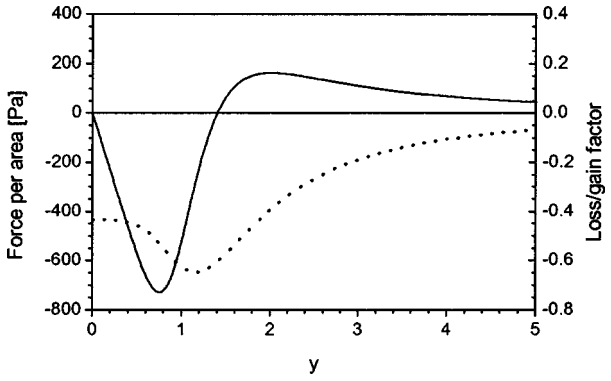


FIG. 10. Aerodynamic force (dotted line) and loss/gain (solid line) versus gap width. Gap width y is dimensionless.

blown frequency will then be close to the plucked one.

For a given pipe geometry and wind pressure the gap y_P in equilibrium position without pressure in the foot should be bigger for weaker reeds than for stronger ones. That is, weaker tongues have to be curved up more than stiffer ones. Since y_P is adjusted by the polishing process, this voicing step is extremely important.

Investigate first the expression in the square bracket on the left side of Eq. (15). Since $g(y)$ corresponds to a mechanical resistance, this component provides a negative resistance for small y values. For larger y values the resistance becomes positive; thus, the aerodynamic effect increases the loss of the reed in this domain. Functions $g(y)$ and $f(y)$ are shown in Fig. 10.

The expressions in the left and right square brackets, i.e., the total loss/gain and the total force of Eq. (15), are shown in Fig. 11. Assume a symmetric small-amplitude oscillation around the equilibrium position given by the zero crossing of the force curve. In a small domain around this point the resistance is negative; thus, the amplitude of an initial small oscillation will increase. The loss/gain during one single period of oscillation can be represented by the integral of the loss curve over the shaded domain of oscillation (see Fig. 11). Therefore, the amplitude will increase until the value of the integral equals zero. At this amplitude the oscillation becomes self-sustained. It is clear from the figure that self-sustained oscillation can be achieved without closing the gap entirely.

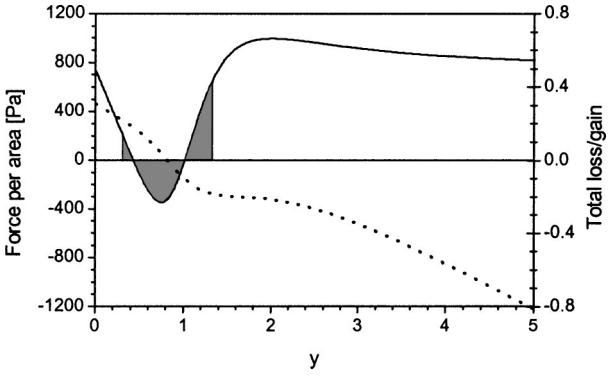


FIG. 11. Total loss/gain curve (solid line) and total force curve (dotted line) versus gap width. Shaded area: domain of self-sustained oscillation. Gap width y is dimensionless.

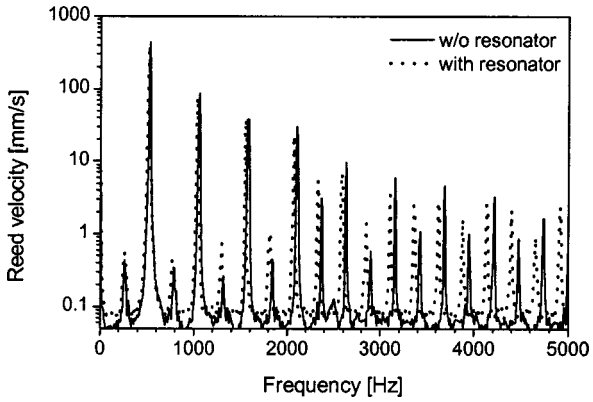


FIG. 12. Subharmonics in reed vibration spectra without and with resonator attached.

The frequency of the self-sustained oscillation can be estimated in the following way: The derivative of the spring force with respect to the displacement is proportional to the negative of the squared angular frequency for a simple harmonic oscillator. Since a small part of the force curve around the equilibrium position (zero crossing) can always be regarded as linear, the slope of the force curve at the zero crossing will be proportional to $-\omega^2$. The slope, as shown in Fig. 9, is always negative, and it is always steeper than the slope of the pure elastic restoring force (see Fig. 8). That is, the blown frequency of the reed is always higher than the plucked frequency. The relative frequency shift, however, will be the biggest for weak reeds, somewhat smaller for medium reeds, and the smallest for strong reeds. That is, the presented model of reed oscillation can explain the increase of the blown frequency by taking into account the stiffening of the reed due to aerodynamic effects.

D. Effects of the acoustic field on reed vibrations

Two observed phenomena can be explained by the effect of the acoustic field on the vibration of the reed. It is obvious that the air mass load on the reed should decrease the blowing frequency compared to the plucked one. Therefore, the observed upshift of the blown frequency was very surprising. However, the decrease of the blown frequency can be seen also in the presented experiments. The difference between the curves of blown and plucked frequency in Fig. 3 will be smaller for shorter reeds (higher frequencies). For very short reeds the blown frequency is smaller than the plucked one. Since the acoustic load (inertance) due to the sound field in the shallot increases linearly with the frequency in the range of the presented measurements, the observed shape of the blown frequency curve can be explained by the combination of the frequency-increasing effects discussed before and of the frequency-lowering effect of the acoustic load.

Another interesting property of reed pipes is the generation of subharmonics. Measured spectra of reed vibration of the blown pipe with and without a resonator attached are shown in Fig. 12. Sharp peaks can be observed halfway between the harmonics. Since subharmonic peaks appear without a resonator, too, they are not produced by the interaction between the reed generator and the acoustic resonator. It was also found that the subharmonic is not reproducible; the

subhar- monic content in the spectrum may be significantly different in subsequent soundings of the pipe. It seems that some kind of instability may be responsible for this effect.

A physical explanation of such instability can result from taking into account the effect of the curving (polishing) process. The reed will be curved, because a mechanical stress is produced in the reed material by the curving process. Since the shape of the curved reed is very complicated, it cannot be described unambiguously by a mathematical function. For a qualitative description, however, it can be assumed that the deformation of the reed corresponds to the deformation of a clamped-free cantilever under a uniform stress σ[20]

$$\xi(z)=\frac{\sigma}{2Eh^3}z^2(z^2-4lz+6l^2). \tag{16}$$

The displacement of the free end is given by

$$\xi(t)=\frac{3\sigma l^4}{2Eh^3}. \tag{17}$$

In turn, the stress produced by the curving process can be estimated from the displacement of the free end

$$\sigma=\frac{2Eh^3}{3l^4}\xi(l). \tag{18}$$

The radius of curvature at $z=0$ can be given as

$$R(0)=\frac{l^2}{4\xi(l)}, \tag{19}$$

and the mean radius of curvature ($R(l)=0$)

$$R=\frac{l^2}{8\xi(l)}. \tag{20}$$

Since the reed is not stress-free, the differential equation of a stressed plate[22] has to be applied in the form of

$$\rho\frac{\partial^2\xi(z,t)}{\partial t^2}+\frac{Eh^2}{12}\frac{\partial^4\xi(z,t)}{\partial z^4}-T\frac{\partial^2\xi(z,t)}{\partial z^2}=0, \tag{21}$$

where T is the stress due to the curving of the reed. This stress will change during the vibration of the reed, because the curvature of the reed changes. Therefore, T can be written as the sum of a constant part and a variable part, i.e., $T=T_0+T_1$. The variable part T_1 of the stress due to a pressure difference p between the upper and lower surface of the reed can be given as[22]

$$T_1=\frac{Rp}{h}=\frac{l^2}{8h\xi(l)}p. \tag{22}$$

The solution of Eq. (22) can be separated into a time-dependent part $\psi(t)$ and a z-dependent part $\varphi(z)$, as described above. The constant part T_0 contributes only to the displacement distribution function $\varphi(z)$; therefore, the following equation can be derived for $\psi(t)$:

$$\frac{d^2\psi(t)}{dt^2}+\frac{Eh^2}{12\rho}\frac{\alpha_1^4}{l^4}\psi(t)-\frac{R\alpha_1^2}{\rho hl^2}p\psi(t)=0. \tag{23}$$

Assuming a periodic pressure in the form of $p=P\cos(\omega t)$, Eq. (24) can be written in the following form:

$$\frac{d^2\psi(t)}{dt^2}+\omega_0^2(1+\eta\cos(\omega t))\psi(t)=0, \tag{24}$$

where

$$\omega_0^2=\frac{E\alpha_1^4}{12\rho}\frac{h^2}{l^4}\quad\text{and}\quad\eta=\frac{3l^4}{2\alpha_1^2h^3\xi(l)}\frac{P}{E}. \tag{25}$$

Equation (24) is the well-known Mathieu equation of parametric instability.[21,22] The amplitude of the pump function can be estimated as $\eta\approx0.09$. Because of the parametric instability an oscillation around half of the pump frequency ω may be developed.[22] That is, Eq. (25) can explain the appearance of subharmonic peaks observed in the measured sound spectra of blown pipes. The physical reason of this effect is the periodic modulation of the mechanical stress in curved reeds due to the sound pressure in the boot and shallot.

V. DISCUSSION

Three rather surprising phenomena (the frequency shift by large-amplitude plucking, higher blown frequency than plucked one, and the appearance of subharmonics in the vibration spectrum) can be explained by qualitative physical models of reed vibration.

The vibration of the plucked reed corresponds to the expected exponentially decaying sine function, but the effective vibrating length is somewhat shorter than the distance between the tuning wire and the free end of the tongue.

The observed frequency difference between small- and large-amplitude plucking can be explained by the change of the vibrating length with the amplitude.

Two different effects contribute to the increase of the blown frequency over the plucked one: the shortening of the vibrating length due to the boot pressure and the increase of the stiffness of the reed due to aerodynamic forces. These effects overcompensate the frequency-lowering effect of air mass loading on the reed; thus, the frequency of the blown reed will be higher than that of the plucked reed. However, air mass loading scales with the frequency; therefore, the blown frequency becomes lower than the plucked one for short reeds, as can be seen in Fig. 3.

The vibrating reed did not strike the surface of the shallot in our experiments. Due to the curved shape of the reed, it can vibrate freely around the equilibrium position. The gap between reed and shallot is very small at the lower turning point of the oscillation, but no mechanical contact of the metal surfaces occurred.

Air flow through the thin slit between shallot and reed has been regarded as laminar in this paper. This is the main difference between the present and earlier treatments[2,11,12,14] of the reed generator of lingual organ pipes. While the velocity of the laminar flow has a shape shown in Fig. 7 as a function of the gap, the velocity of a Bernoulli flow would be maximal at the limit $y=0$ and it would decrease monotonically with increasing gap [$w=a/(1+by^2)$ dependence]. The blown frequency then would always be lower than the plucked one. Thus, the observed behavior of the blown frequency contradicts the assumption of Bernoulli flow. Moreover, the Reynolds number calculated from the velocity

shown in Fig. 7 approximates a constant value (~1600) as the gap increases, and this limit value is smaller than the laminar limit (Re<2320) for the usual lingual pipe geometry.

The laminar theory presented can explain several properties of lingual organ pipes known from the tradition of organ building, and supports the opinion of organ builders concerning the ultimate importance of reed curving for the proper speech and sound quality of reed pipes.

The theory given in this paper creates only a qualitative model, because several effects cannot be quantified properly (shape of the curved tongue, flow resistance of shallot opening and of the bore in the foot of the pipe, etc.). The main problem is the modeling of the flow in the slit. The length of the slit (essentially the thickness of the shallot wall) is not long enough to allow the full development of a laminar flow. Thus, Eq. (A1) in the Appendix probably overestimates the laminar flow velocity. Therefore, the maximal flow velocity U and the gap of the maximal velocity a calculated by Eq. (A14) may deviate significantly from the real values. The calculated value of $a=0.23$ mm may be only about half of the real value, because the amplitude of vibration determined by integrating the velocity profile shown in Fig. 4 is ~0.45 mm.

The real shape of the side opening between the shallot and the reed is far from a triangle, as is assumed in the Appendix. Since each tongue has slightly different shape, it is impossible to take into account the real shape of the side opening at the model calculations. However, this shape can have quite a large influence on the average flow velocity and on the aerodynamical forces.

The flow resistances of the bore and the open end of the shallot are also not known. For both resistance (drag) coefficients, a value of 1.5 was used in the calculation. For better results, the flow resistances should be determined by measurements.

Only one dynamic effect, the change of the gap due to reed vibration, was taken into account. Flow in the boot and the shallot, however, should be described by the time-dependent Bernoulli equation, and compressible (acoustic) flow should be taken into account. In this case the volume of the boot and the acoustic properties of the shallot would enter into the theory. According to the opinion of the authors, however, these effects play only a secondary role in the mechanism of reed oscillation.

The main features can be described by the simple model presented in this paper. This treatment also shows that the reed oscillation is essentially an aerodynamic phenomenon. Although the vibrating reed produces sound, feedback from the acoustic field is not necessary for sustaining the oscillation. However, the acoustic field may modify the vibration of the reed. Such an effect can be observed, in the case of weak reeds, in the vibration spectrum of the reed.[17,18] Much more influence can be expected in reed pipes with resonators. Sound generation by reed pipes with resonator, and the effect of acoustic feedback on the vibration of the reed will be discussed in a second paper.

ACKNOWLEDGMENTS

The authors express their thanks to Neville Fletcher for his helpful comments and to the organ builder B. Welde for providing the lingual pipes for the measurements. T.D.R. expresses thanks to Karl Gertis for supporting his research visit to the Fraunhofer-Institut für Bauphysik, Stuttgart.

APPENDIX: AIR FLOW BETWEEN THE REED AND THE SHALLOT

It is assumed that the side opening between the shallot and the reed has a triangular form so that the height of the opening equals zero at $z=0$ and the value x at $z=l$. Thus, $x(z)=xz/l$. The front opening is a rectangle with height x and width b. The length of the gap in the direction of the flow equals to the thickness δ of the wall of the shallot.

The flow through the slit is regarded as laminar. The velocity through the side opening can be calculated as

$$w(z)=\frac{\Delta p x^2}{12\rho\nu\delta}\frac{z^2}{l^2}, \tag{A1}$$

where Δp, ρ, and ν are the pressure difference between boot and shallot, the density, and kinematic viscosity of the air, respectively. The velocity of the laminar flow through the side opening depends on coordinate z, also. Therefore, the velocity has to be averaged over the length l of the side opening. The average flow velocity can be expressed as the ratio of the volume flow and the area of the opening

$$\overline{w}=\frac{2}{xl}\int_0^l w(z)\frac{xz}{l}dz=\frac{2\Delta p x^2}{12\rho\nu\delta l^4}\int_0^l z^3 dz=\frac{1}{2}\frac{\Delta p x^2}{12\rho\nu\delta}. \tag{A2}$$

The total flux q through the slit can be given as the sum of the fluxes through the front slit and the side ones

$$q=bx\frac{\Delta p x^2}{12\rho\nu\delta}+2\frac{xl}{2}\frac{1}{2}\frac{\Delta p x^2}{12\rho\nu\delta}=x\left(b+\frac{l}{2}\right)\frac{\Delta p x^2}{12\rho\nu\delta}. \tag{A3}$$

The average velocity of the laminar flow through the slit is defined as the ratio of the total flux q and the total open area of $(b+l)x$

$$w=\frac{\Delta p x^2}{12\rho\nu\delta}\frac{b+l/2}{b+l}=\frac{\Delta p x^2}{12\rho\nu\tilde{\delta}}, \tag{A4}$$

where an effective gap length was introduced by the definition

$$\tilde{\delta}=\delta\frac{2l+2b}{l+2b}. \tag{A5}$$

The stationary flow through the pipe can be determined from the following four equations:

$$(1+\varsigma_F)\tfrac{1}{2}\rho w_F^2=p_0-p_B, \tag{A6}$$

$$\frac{12\rho\nu\tilde{\delta}}{x^2}w=p_B-p_S, \tag{A7}$$

$$\varsigma_S\tfrac{1}{2}\rho w_S^2=p_S, \tag{A8}$$

and

$$A_F w_F = (b+l)xw = A_S w_S, \tag{A9}$$

where w_F, w_S, ζ_F, ζ_S, A_F, A_S, p_0, p_B, p_S are flow velocities in the foot hole and at the open end of the shallot, the flow resistance (drag) coefficients of the foothole and shallot opening, the cross-sectional areas of the foothole and the shallot, and the pressures in the windchest, pipe boot, and shallot, respectively.

Equations (A6) and (A8) describe the free flow through the foot hole and the open end of the shallot. Equation (A9) is the mass conservation equation for incompressible flow. Equation (A7) is the same as Eq. (A4), and it describes the laminar flow through the slit between reed and shallot.

The velocities w_F and w_S can be substituted by w using Eq. (A9). By adding Eqs. (A6)–(A8) the pressures p_B and p_S fall out and an equation quadratic in the variable w remains

$$\left(\frac{1+\varsigma_F}{A_F^2}+\frac{\varsigma_S}{A_S}\right)(l+b)^2x^2w^2+\frac{24\nu\tilde{\delta}}{x^2}w-\frac{2p_0}{\rho}=0. \tag{A10}$$

This equation can be written in a simpler form

$$w^2+\frac{2B}{x^4}w-\frac{C}{x^2}=0, \tag{A11}$$

where B and C are the following constants:

$$B=\frac{12\nu\tilde{\delta}}{\left(\frac{1+\varsigma_F}{A_F^2}+\frac{\varsigma_S}{A_S^2}\right)(l+b)^2}$$

and (A12)

$$C=\frac{2p_0/\rho}{\left(\frac{1+\varsigma_F}{A_F^2}+\frac{\varsigma_S}{A_S^2}\right)(l+b)^2}.$$

The solution of Eq. (A11) can be written as

$$w=\frac{B}{x^4}\left(\sqrt{1+\frac{C}{B^2}x^6}-1\right)=2U\frac{\sqrt{1+y^6}-1}{y^4}, \tag{A13}$$

where U and the dimensionless variable y were introduced by the definitions

$$U=\sqrt[3]{\frac{C^2}{8B}},\quad y=\sqrt{2}\frac{x}{a},\ \text{and}\quad a=\sqrt[6]{\frac{8B^2}{C}}. \tag{A14}$$

The quantities U and a depend only on the geometry of the pipe foot and the wind pressure p_0. The flow velocity w has a maximum at $x=a$ where $w=U$. Therefore, parameters U and a can be regarded as the maximum flow velocity through the slit between reed and shallot and the value of the gap width at maximum velocity. For the pipe studied $U=37.5$ m/s and $a=0.23$ mm can be calculated from the geometrical data and assuming 800-Pa wind pressure. The calculated velocity w vs y curve can be seen in Fig. 7.

In case of vibration, the gap changes during the time needed for a fluid particle to cross the slit. Therefore, x should be replaced by the following expression:

$$x+\frac{\nu\delta}{2w}, \tag{A15}$$

where $\nu=\mathrm{d}x/\mathrm{d}t$ is the velocity of the reed. The flow velocity w can be calculated then from Eq. (A11). The first-order approximation can be written as

$$w=2U\frac{\sqrt{1+y^6}-1}{y^4}+\left(\frac{3}{\sqrt{1+y^6}}-1\right)\frac{\delta}{2y}\frac{dy}{dt}. \tag{A16}$$

The force acting on the reed can be calculated by assuming that a pressure difference of p_B-p_S pushes the middle part of the reed towards the shallot, while a Bernoulli force tries to close the gap. Both components must have a negative sign, because they are directed to the shallot. The force can be given as the sum of the two components

$$F\left(y,\frac{dy}{dt}\right)=-(p_B-p_S)bl(l-\gamma)-\frac{1}{2}\rho w^2 bl\gamma, \tag{A17}$$

where γ denotes the ratio of the area over the gap to the total area of the vibrating reed

$$\gamma=\frac{(2l+b)\delta}{lb}. \tag{A18}$$

The force can be calculated by substituting Eq. (A16) into Eq. (A17). After some algebra, the force can be written in the following form:

$$F\left(y,\frac{dy}{dt}\right)=f(y)-g(y)\frac{dy}{dt}, \tag{A19}$$

where

$$f(y)=-2p_0(1-\gamma)\frac{\sqrt{1+y^6}-1}{y^6}-2\rho U^2\gamma\left(\frac{\sqrt{1+y^6}-1}{y^4}\right)^2, \tag{A20}$$

$$g(y)=\left[-\frac{3p_0(1-\gamma)\delta}{2U}+\rho U\delta\gamma\left(\frac{3-\sqrt{1+y^6}}{y^2}\right)\right]\times\left(\frac{\sqrt{1+y^6}-1}{y^3\sqrt{1+y^6}}\right). \tag{A21}$$

Calculated curves of $f(y)$ and $g(y)$ vs y are shown in Fig. 10.

[1] N. H. Fletcher, "Sound production by organ flue pipes," J. Acoust. Soc. Am. **60**, 926–936 (1976).
[2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), Chap. 16.
[3] M. P. Verge, B. Fabre, W. E. Mahu, and A. Hirschberg, "Feedback excitation mechanism in organ pipes," J. Acoust. Soc. Am. **95**, 1119–1132 (1994).
[4] M. Castellengo, "The role of mouth tones in the constitution of attack transients of mouth pipes," Acust. Acta Acust. **85**, 387–400 (1999).
[5] A. Miklós and J. Angster, "Properties of the sound of labial organ pipes," Acust. Acta Acust. **86**, 611–622 (2000).
[6] J. W. Coltman, "Jet drive mechanisms in edge tones and organ pipes," J. Acoust. Soc. Am. **60**, 725–733 (1976).
[7] S. A. Elder, "On the mechanism of sound production in organ pipes," J. Acoust. Soc. Am. **54**, 1554–1564 (1973).
[8] T. L. Finch and A. W. Nolle, "Pressure wave reflections in an organ note channel," J. Acoust. Soc. Am. **79**, 1584–1590 (1986).
[9] S. Yoshikawa and J. Saneyoshi, "Feedback excitation mechanism in organ pipes," J. Acoust. Soc. Jpn. (E) **1**, 175–191 (1980).

[10] W. Lottermoser and J. Meyer, *Orgelakustik in Einzeldarstellungen* (Verlag Das Musikinstrument, Frankfurt am Main, 1966).
[11] N. H. Fletcher, ''Autonomous vibration of simple pressure-controlled valves in gas flows,'' J. Acoust. Soc. Am. **93**, 2172–2180 (1993).
[12] A. Hirschberg, R. W. A. van de Laar, J. P. Marrou-Mauriéres, A. P. J. Wijnands, H. J. Dane, S. G. Kruijswijk, A. J. M. Houtsma, ''A quasi-stationary model of air flow in the reed channel of single-reed woodwind instruments,'' Acust. Acta Acust. **70**, 146–153 (1990).
[13] G. R. Plitnik, ''Vibration characteristics of pipe organ reed tongues and the effect of the shallot, resonator, and reed curvature,'' J. Acoust. Soc. Am. **107**, 3460–3473 (2000).
[14] A. Z. Tarnopolsky, N. H. Fletcher, and J. C. S. Lai, ''Oscillating reed valves—An experimental study,'' J. Acoust. Soc. Am. **108**, 400–406 (2000).
[15] P. Williams and B. Owen, *The Organ* (Norton, New York, 1988).
[16] R. Janke (organ builder), http://www.orgel-info.de
[17] T. D. Rossing, J. Angster, and A. Miklós, ''Reed vibration and sound generation in lingual organ pipes,'' J. Acoust. Soc. Am. **104**, 1767–1768 (1998).
[18] T. D. Rossing, J. Angster, and A. Miklós, ''Reed vibration and sound generation in lingual organ pipes,'' in Proc. of Int. Symp. on Mus. Acoust., Perugia, Italy, Vol. **1**, 313–316 (2001).
[19] J. Braasch, J. Angster, and A. Miklós, ''The influence of the shallot leather facing on the sound of lingual organ pipes,'' in Proc. of Int. Symp. on Mus. Acoust., Perugia, Italy, Vol. **1**, 325–328 (2001).
[20] L. D. Landau and E. M. Lifshitz, *Theoretical Physics, VII. Elasticity* (Pergamon, Oxford, 1960), Chap. 3.
[21] L. D. Landau and E. M. Lifshitz, *Theoretical Physics, I. Mechanics* (Pergamon, Oxford, 1960), Chap. 5.
[22] M. A. Mironov, ''Parametric instability of a circular cylindrical shell propagating a Korteweg wave,'' Acoust. Phys. **41**, 707–711 (1995).

Time-domain simulation of sound production of the sho

Takafumi Hikichi
NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan

Naotoshi Osaka
NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan

Fumitada Itakura
Graduate School of Engineering, Nagoya University, 1 Furo-cho, Chikusa-ku, Nagoya, Aichi 468-8603, Japan

(Received 10 April 2002; revised 31 October 2002; accepted 11 November 2002)

A physical model based on the sound production mechanism of the sho is proposed with intention of applying it to sound synthesis. Time-domain simulation was done using this model, and effects of the tube length and blowing pressure on the sounding frequency and sounds spectra were investigated. The reed vibration, pressure variation inside the tube, and threshold blowing pressure for oscillation were measured by artificially blowing air into the sho. The experimental results are in acceptable agreement with simulation results in terms of the relationships between tube length and threshold pressure and between tube length and the sounding frequency. In addition, recorded sound waveforms and simulated ones have a common feature in the sense that high-frequency components of their spectra increase with increasing blowing pressure. Further, it is concluded that a sho reed acts as an ''outward-striking valve.'' © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1534605]

PACS numbers: 43.75.Pq [NHF]

I. INTRODUCTION

Physical modeling is attracting much attention in attempts to synthesize the sounds of musical instruments. Physical models excel in naturally controlling musical instrumental sounds compared with other conventional methods, such as additive synthesis and sampling synthesis. Dynamic features, such as attack, transience, and frequency variation, can be realized because it has inherent mechanism that changes sound features like real instruments. In addition, artificial instruments can have the same parameters as real ones, which allows users to control the instruments intuitively. Many artificial instruments have recently been developed and offered to computer music composers.[1–3]

However, our understanding is not sufficient to make full use of physical models' abilities. In particular, little attention has been paid to instruments other than Western orchestral instruments. Departing from this trend, this paper treats the sho, a free reed mouth organ used in traditional Japanese court music called ''Gagaku.'' The purposes of this paper are to model the sound production mechanism of the sho, and, through simulations based on that model, show the possibility of physical modeling that synthesizes realistic sounds.

Figures 1(a) and (b) show a sho, the sho disassembled. The sho is mainly composed of a mouthpiece, cavity, and seventeen bamboo pipes. Metal reeds are glued with resin to the lower side of the bamboo pipes. Figure 1(c) shows a close-up view of a reed. The sho is played by holding it in front of the face upright. When a player blows through the mouthpiece and closes the small finger holes on the tubes, oscillation commences and the reeds start sounding. The tubes whose finger holes remain open do not sound even if air is supplied, so the player can control sounding by fingerings. Some tubes have one slot besides the finger hole, which determine the effective lengths. Each pipe has a different length and a different reed which gives a different pitch. The pitch range of a typical sho covers A4 (430 Hz) to F♯6 (1451 Hz). Note that the standard pitch for playing gagaku music is A4 = 430 Hz, which is different from one normally used in Western music. Table I shows position, name and pitch of each pipe of a sho, and Fig. 2 shows top view of the cavity (wind chest) part of the sho. One characteristic of the sho is that it can be played not only by blowing but also by drawing. In typical playing style, the fingers close five or six finger holes simultaneously, and multiple pipes sound in chords.

The instrument is said to originate in the 3rd or 4th century in China. There are other musical instruments, mostly found in Eastern and Southern Asia, that work on a similar mechanism. These include Chinese sheng and Laotian khaen. It is said that the instrument was introduced to Japan around the 8th century, and that its structure has remained virtually unchanged to the present day. In spite of such a long history, its mechanism is not yet fully understood and has not been applied to synthesis.

Acoustically, it can be described as free reeds coupled to pipe resonators. Let us call this structure as ''sho-type.'' The sho-type instrument is categorized as a free reed instrument like the harmonica, accordion, and reed-pipe organ. However, unlike these Western free reed instruments, the reeds of the sho-type instrument are approximately symmetrical, so that the same reed vibrates on both blowing and drawing. Some measurements of sho-type instrument have been re-

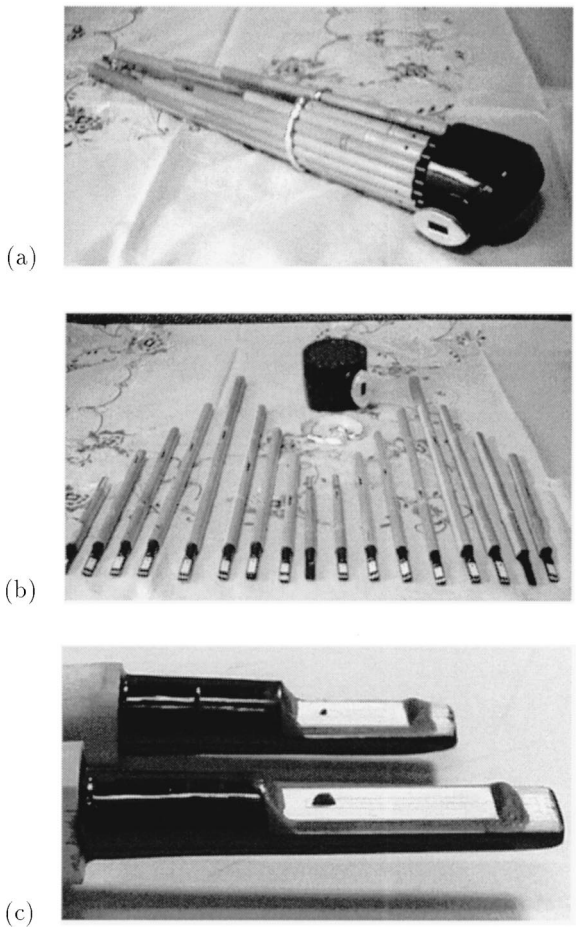


FIG. 1. (a) The sho. (b) Disassembled. (c) Close-up view of a reed.

ported by Cottingham.[4–6] For khaen, the relationship between the pipe length and sounding frequency has been reported, and it was shown that the vibrating frequency of the reed could be pulled to match the pipe resonance. The input impedance of the khaen has also been measured, and it was concluded that the reed acted as ''outward striking.'' On the relationship between blowing pressure versus sounding frequency, it was concluded that sounding frequency decreased with increasing blowing pressure.

TABLE I. Position, name, and pitch of each pipe of a sho. Position is shown in Fig. 2. ''Ya'' and ''mou'' do not have the reeds so they do not contribute to sounding.

Position	Name	Pitch
1	sen	F♯6
2	jyu	G5
3	ge	F♯5
4	otsu	E5
5	ku	C♯5
6	bi	G♯5
7	ichi	B4
8	hachi	E6
9	ya	
10	gon	C♯6
11	shichi	B5
12	gyo	A5
13	jyo	D6
14	bou	D5
15	kotsu	A4
16	mou	
17	hi	C6

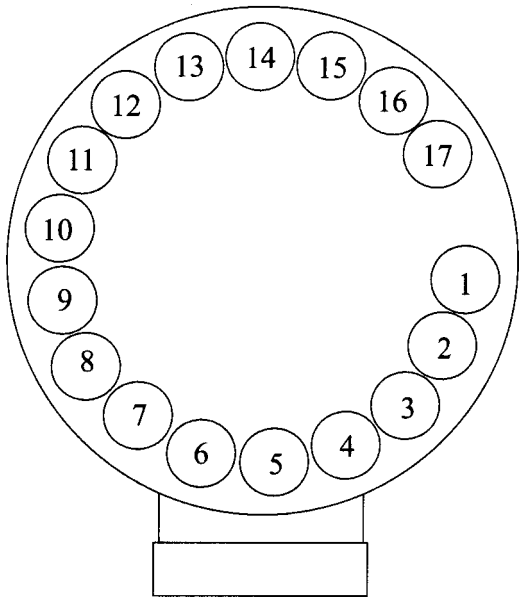


FIG. 2. Top view of a wind chest of the sho. Small circles with numbers show holes where pipes are mounted.

Although sound production mechanism of the sho-type instrument has not been proposed, it is expected that research for other wind instruments may give us hints to model this instrument. According to a standard theory for wind instruments, the sound production mechanism comprises a generator and a resonator. For instruments of this type, the generator corresponds to reeds and airflow, and the resonator to an air column.

As for reeds, the behavior of a free reed has been investigated by Tarnopolsky *et al.* both experimentally and theoretically.[7] A reed was treated as a pressure-controlled valve in a simplified manner, and good correspondence between theoretical and experimental results was found. So, we adopt their formulation to describe the motion of the sho reed. For modeling of the pipe resonators, we can utilize Schumacher's well-known work on woodwind instruments to carry out time domain simulations.[8]

Section II describes the experiment carried out to ascertain the basic physical mechanism of the sho-type instrument. In Sec. III, our sho sound production system is formulated and the oscillation condition is analyzed based on linear theory. The simulation results obtained using our physical model are discussed and compared with experimental results in Sec. IV. Section V summarizes the results and concludes the paper.

II. EXPERIMENTS

Measurements were made on variations in threshold pressure and sounding frequency with tube length to get sufficient understanding on sound production mechanism of the sho-type instrument.

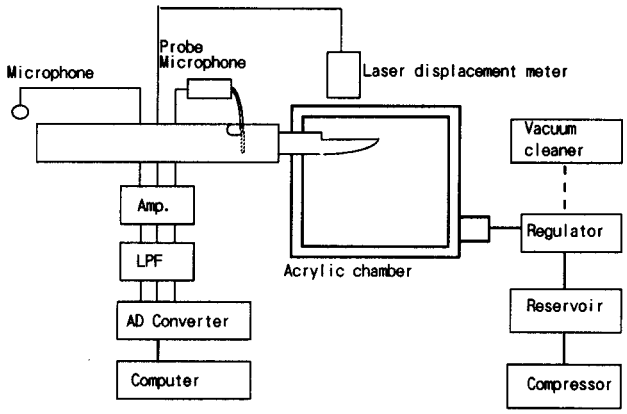


FIG. 3. The experimental setup.

A. Experimental condition

Figure 3 shows details of the experimental arrangement used. We focus on the ''ichi'' tube (B4, 483.7 Hz) because it does not have any slot on the side wall and its configuration is the simplest among the tubes. The ''ichi'' tube was taken from a disassembled sho and mounted in an acrylic chamber made for this experiment. A pressure sensor was connected to the chamber to measure the difference between the pressure inside the chamber and the atmospheric pressure. We will refer to this as blowing pressure. A probe microphone was inserted into the tube via the finger hole, and the hole was closed with rubber tape. A condenser microphone was mounted 5 cm from the open end of the pipe. The instrument was artificially played by a compressor (blow) or by a vacuum cleaner (draw), and blowing pressure was adjusted manually using a regulator. Blowing pressure was in the range of 0 to 1 kPa, which was static with time. The reed vibration was measured by a laser displacement meter, and pressure variations were measured by microphones. All the recorded signals were transferred to computer via an A/D converter. The sampling frequency was set to 24 kHz. The cutoff frequency of the probe microphone was set to 3.6 kHz, and that of the displacement sensor and the condenser microphone were set to 10 kHz.

B. Tube resonance frequency and threshold behavior

First, the tube length dependence of threshold pressure was investigated. The length of the tube was varied and the threshold pressure was measured. The tube length was made longer by simply adding a piece of tube to one end, and blowing pressure was gradually increased until oscillation began or it reached 1 kPa. Figure 4 shows the relationship between tube resonance frequency and threshold pressure for positive (blow) and negative (draw) pressures. Here, the resonance frequency of the tube was calculated from its length under the assumption of one-fourth wavelength resonance, since the pipes can be regarded as having one end closed. There is a strong dependency on pipe length, i.e., there is a range in which the threshold is low, and also a range in which the reed cannot oscillate (200–250 Hz). Further, the threshold pressure for negative pressure was lower than for positive pressure. One of the reasons for this is

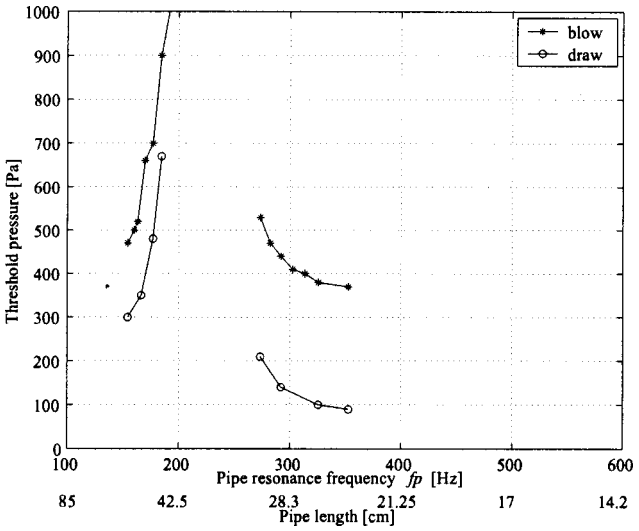


FIG. 4. Tube resonance frequency f_p versus threshold pressure obtained by experiment. Blow means positive, and draw means negative pressure.

probably the small reed asymmetry. Another may be the difference in the configuration of upstream/downstream sides of the reed. In addition, hysteresis was found in the experiment, i.e., threshold pressures obtained by gradually increasing the blowing pressure were slightly different from those obtained by decreasing the pressure.

C. Tube resonance frequency and sounding frequency

It was expected that sounding frequency would vary with tube length. To clarify this, the relationship between the tube resonance frequency and the sounding frequency under constant pressure (positive: 0.8 kPa, negative: 0.5 kPa) was examined. Figure 5 shows experimental results for positive (blow) and negative (draw) pressures. Integer multiples of the pipe resonance frequency are shown by dashed–dotted lines. The dashed line shows the reed frequency and is in-

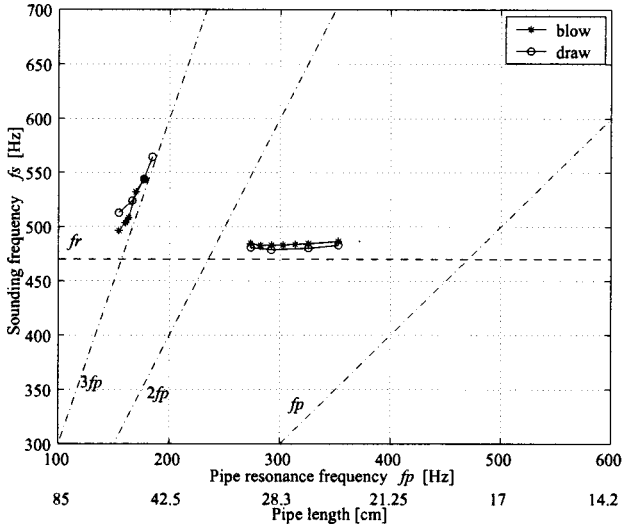


FIG. 5. Tube resonance frequency f_p versus sounding frequency f_s obtained by experiment. Blow means positive, and draw means negative pressure. Dashed–dotted lines show integer multiples of the pipe resonance frequency and dashed line shows the reed frequency.

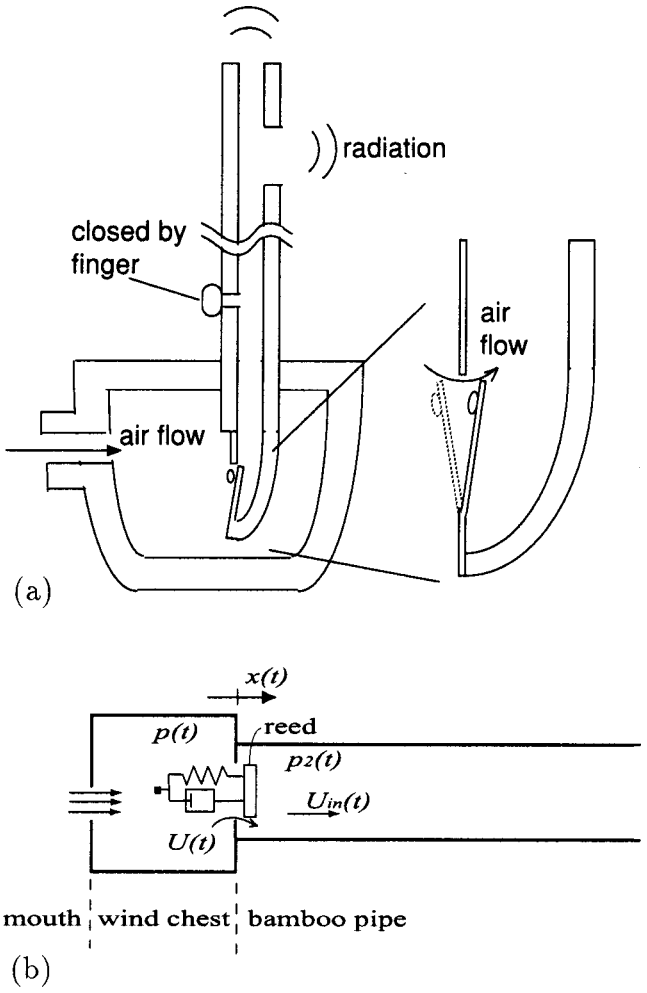


FIG. 6. (a) Configuration of an actual sho. (b) Simplified physical model.

cluded to make the data easy to understand. Here we will divide whole frequency range into three parts for explanation. First, when tube resonance frequency is low, the sounding frequency increases quickly with increasing pipe resonance until it reaches about 200 Hz. Around this frequency, the pitch reached 570 Hz. Then, there is a no-oscillation range of 200–250 Hz. And then again, there is a range where oscillation occurs around 480 Hz.

From these results, we found that (1) oscillation commenced above the reed frequency (dashed line in the figure), and (2) oscillation commenced above the first or third multiple of the pipe resonance frequency (shown by the dashed–dotted lines). These frequency relationships suggest that the reeds in a sho operate in an ''outward-striking'' manner.

III. FORMULATION

From the experimental results mentioned above, it was shown that the reed behaves as an outward-striking valve. Based on this finding, sound production system of the sho is formulated.

Figure 6 shows the actual configuration of the sho and its simplified model. A player blows the air into the wind chest as shown in Fig. 6(b), and as a result, the pressure inside the cavity $p(t)$ increases. We consider the pressure

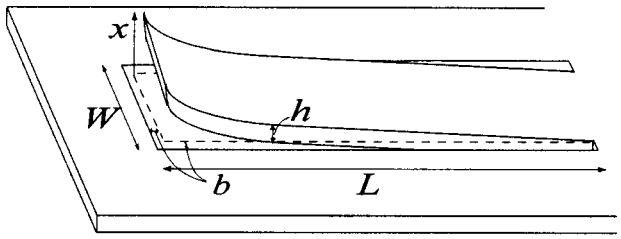


FIG. 7. A close-up schematic view of a reed.

$p(t)$ as an input to our system, and neglect effects of the flow, cavity vibration, etc., on the pressure.

As for reeds, the behavior of an outward-striking valve has been investigated by Tarnopolsky *et al.*[7] and Fletcher.[9] The formulation in Secs. III A and III B below are based on their work.

A. Modeling the reed vibration

The vibration of the reed is that of a simple cantilever. Let pressure at the upstream of the reed (i.e., pressure inside the cavity) at time t be $p(t)$, and pressure at the downstream (pressure inside the pipe) be $p_2(t)$. As shown in the Appendix, the equation of motion of the reed has the form

$$\frac{d^2x}{dt^2}+\frac{\omega_r}{Q}\frac{dx}{dt}+\omega_r^2(x-x_0)=\frac{1.5WL}{m}(p(t)-p_2(t)), \tag{1}$$

where x is the displacement at the tip of the reed, Q the resonance Q value, ω_r the angular resonance frequency, and x_0 the initial displacement. W, L, and m are the width, length, and mass of the reed, respectively.

B. Airflow through the reed

From Bernoulli's equation, the relationships among $p(t)$, $p_2(t)$, and volume velocity through the slit $U(t)$ are as follows:

$$p(t)=p_2(t)+\frac{\rho}{2}\left[\frac{U(t)}{CF(x)}\right]^2+\frac{\partial}{\partial t}\left[\frac{\rho U(t)\delta}{CF(x)}\right], \tag{2}$$

where C is the flow contraction coefficient, ρ the air density, δ the inertia parameter. The flow contraction coefficient represents the effect of the slit configuration. The area of the slit $F(x)$ is described as

$$F(x)=W[x^2+b^2]^{1/2}+2L[a(x)^2+b^2]^{1/2},$$

where $a(x)$ is the average displacement of the sides of the reed, b the clearance gap around the reed. Figure 7 shows the reed configuration in detail. For a sho reed, we assume $x_0=0$. Considering it displaces both ways, and from the form of the mode function, it follows that (see the Appendix),

$$a(x)\approx|x_0+0.4(x-x_0)|=|0.6x_0+0.4x|=0.4|x|.$$

C. Reflection function of the tube

The acoustical characteristics of the tube are specified by the input impedance $Z_{\text{in}}(f)$. Input impedance is defined by the ratio of sound pressure and volume velocity at the end

of the tube with frequency f. Reflection coefficient $R(f)$ is defined by the input impedance $Z_{\text{in}}(f)$ and characteristic impedance of the tube Z_0 as

$$R(f)=\frac{Z_{\text{in}}(f)-Z_0}{Z_{\text{in}}(f)+Z_0}.$$

Reflection function $r(t)$ is defined by the inverse Fourier transform of $R(f)$. It represents the pressure waveform, which is reflected back to the end of the tube when impulse pressure is injected at $t=0$. Using this reflection function, we calculate the pressure inside the tube $p_2(t)$ as

$$\begin{aligned} p_2(t)&=Z_0U_{\text{in}}(t)+r(t)*(p_2(t)+Z_0U_{\text{in}}(t)),\\ U_{\text{in}}(t)&=U(t)+0.4WL\frac{dx}{dt}, \end{aligned} \tag{3}$$

where the asterisk denotes convolution and $U_{\text{in}}(t)$ is the net volume velocity input to the tube. The quantity $0.4WL(dx/dt)$ is approximately the volume velocity displaced by a cantilever of width W and length L when its tip is moved by (dx/dt).

The pipe shape is relatively simple. We have measured the resonance frequency of the ''ichi'' pipe, and have concluded that the pipe could be approximated as a cylinder.[10] So, we use a simple reflection function of Gaussian type and adapt Schumacher's method to calculate the pressure at the entrance of the pipe.

D. Linear theory of oscillation

In the preceding sections, the sound production system was formulated with two linear elements and their nonlinear interaction. To gain insight into this system before simulation, this section investigates the condition under which self-oscillation continues within linear theory when the amplitude of oscillation is small.

For brass instruments, self-oscillation condition based on linear theory has been discussed by Adachi and Sato.[11] The derivation shown here is based on their work.

First, let the dc component and ac component of the variables be described as $\overline{U}$ and $\tilde{U}$, $\overline{x}$ and $\tilde{x}$, etc. Assuming the input pressure p is static, the dc components satisfy the following stationary conditions:

$$\omega_r^2\overline{x}=\frac{1.5WLp}{m}, \tag{4}$$

$$p=\frac{\rho}{2}\left[\frac{\overline{U}}{C\overline{F}}\right]^2, \tag{5}$$

which are derived from Eqs. (1) and (2).

The dc components $\overline{U}$, $\overline{F}$, $\overline{x}$ are described as follows:

$$\overline{x}=\frac{1.5WLp}{m\omega_r^2}, \tag{6}$$

$$\overline{F}=W[\overline{x}^2+b^2]^{1/2}+2L[0.16\overline{x}^2+b^2]^{1/2}, \tag{7}$$

$$\overline{U}=\sqrt{\frac{2p}{\rho}}C\overline{F}. \tag{8}$$

Equation (2), which governs the nonlinear airflow dynamics, is linearized as

$$\left(2+\sqrt{\frac{2\rho}{p}}\delta\frac{d}{dt}\right)\frac{\tilde{U}}{\overline{U}}=-\frac{p_2}{p}+2\frac{\tilde{F}}{\overline{F}}, \tag{9}$$

where airflow velocity is much greater than that of the reed. In the ordinary range of parameter values, the inertia of airflow can be neglected. Therefore, we omit the second term on the left-hand side of Eq. (9).

Substituting $x=\overline{x}+\tilde{x}$ into Eq. (1), where the angular frequency of the reed $\omega(=2\pi f)$, we have

$$\frac{\tilde{x}}{p_2}=-\frac{1.5WL}{m\omega_r^2}\Lambda\left(\frac{\omega}{\omega_r}\right), \tag{10}$$

where $\Lambda(\Omega)$ is defined as

$$\Lambda(\Omega)=\frac{1}{1-\Omega^2+j(\Omega/Q)}.$$

Reed compliance G is defined as $G(f)=\tilde{F}(f)/p_2(f)$, similar to the previous work on brass instruments,[11] where $\tilde{F}(f)$ and $p_2(f)$ denote Fourier components of $\tilde{F}$ and p_2 at frequency f, respectively. From Eq. (10), G becomes

$$G(f)=\frac{\tilde{F}(f)}{p_2(f)}=-\frac{1.5WL(W+0.8L)}{m\omega_r^2}\Lambda\left(\frac{\omega}{\omega_r}\right). \tag{11}$$

Equation (11) shows that $|G(f)|$ takes its maximum near the reed resonance frequency f_r, and that $\angle G(f)$ shows π decrease from the lower to the higher side of f_r. Substituting the reed compliance $G(f)$ and input impedance Z_{in} into the linearized equation (9), we obtain the self-oscillation condition as follows:

$$K(f)\equiv\sqrt{\frac{2p}{\rho}}C\overline{G}(f)Z_{\text{in}}(f)=1, \tag{12}$$

where $\overline{G}(f)$ is defined as

$$\overline{G}(f)=G(f)-\frac{\overline{F}}{2p}.$$

Note that the magnitude of $\overline{G}$ is also maximized near the reed resonance frequency and that the angle of $\overline{G}$ has the same sign as the angle of G. Oscillation happens under the condition that there exists a frequency f such that $K(f)$ is real and larger than 1.

The magnitude condition $K(f)>1$ requires that the magnitudes of both $\overline{G}$ and Z_{in} be large. This indicates that the oscillation has a frequency f that is near one of the resonance frequencies of the pipe as well as near the reed resonance frequency.

The phase condition is written as $\angle K=\angle\overline{G}+\angle Z_{\text{in}}=0$. Because $\angle\overline{G}$ is positive, the phase condition is satisfied only if $\angle Z_{\text{in}}$ is negative for oscillation. This indicates that the oscillation happens on the higher frequency side of the input impedance peak. According to Fletcher's classification based on consideration of a simple pressure-controlled valve,[12] the sho reed is classified as ''outward-striking'' valve.

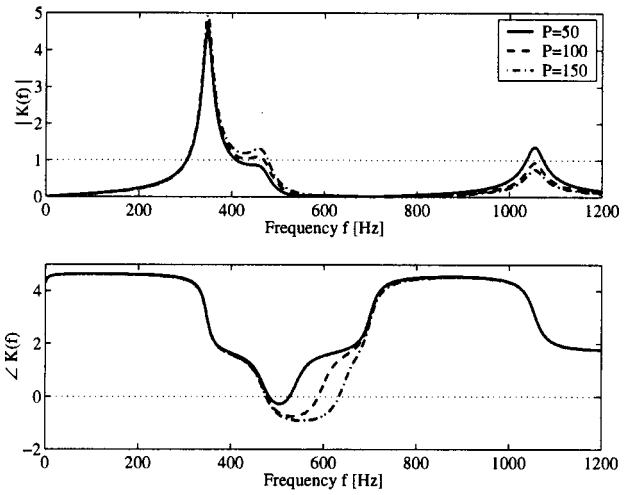


FIG. 8. The magnitude and phase of $K(f)$ defined by Eq. (12) with blowing pressures of 50, 100, and 150 Pa.

Next, a quantitative analysis is done for threshold pressure and sounding frequency using Eq. (12). Threshold pressure is given as a pressure when $|K(f)|=1$ holds at a frequency f that satisfies $\angle K(f)=0$. The magnitude and phase of $K(f)$ defined by Eq. (12) are plotted in Fig. 8 with blowing pressures of 50, 100, and 150 Pa ($Q=10$). From Fig. 8, it is clear that, as the pressure increased, (1) the phase of $K(f)$ changed so as to have a crossing point with zero axis, (2) the crossing frequency shifted to the lower side, and (3) the magnitude of $K(f)$ became larger near the reed frequency. So, threshold pressure was determined by gradually increasing blowing pressure until the magnitude of $K(f)$ exceeded 1 at a frequency f that satisfies $\angle K(f)=0$, which was determined to be the sound frequency.

Figure 9 shows (a) threshold pressure and (b) sound frequency with some values of Q and additional mass rate α_m determined by the above procedure. Here, the additional mass corresponds to a piece of lead attached to the tip of the reed. The amount of this additional mass was specified by a percentage of the original reed mass.

In experiment, threshold pressures were 370 Pa for positive pressure and 90 Pa for negative pressure for normal pipe length ($L_p=24.1$ cm), as shown in Fig. 4. In Fig. 9(a), threshold pressure increases with increasing additional mass rate, and increases with decreasing Q value. Considering the experimental results, the ranges of $Q=8-10$ and $\alpha_m=0-30\%$ seem appropriate. By using these values, threshold pressure in the range of 140 to 350 Pa was obtained, which was considered to be reasonable.

In Fig. 9(b), when $Q=8$ or 10 was used, the sounding frequency obtained by linear analysis was smaller than the 483 Hz obtained from experiment. It is expected that nonlinearity may play a role in shifting the sounding frequency to the higher value. Further, the effect of additional mass on the sounding frequency was found to be small when the natural frequency of the reed was kept constant: increasing the added mass slightly lowers sounding frequency.

IV. SIMULATION

Equations (1)–(3) were discretized, and a simulation was done in 48 kHz sampling. Given $p(t)$, the following

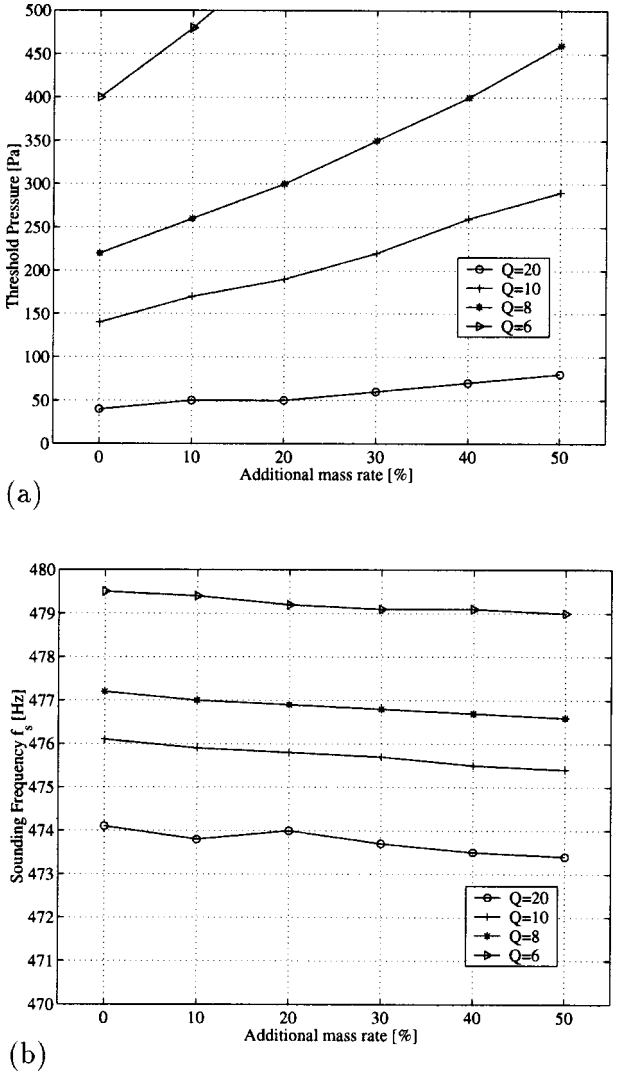


FIG. 9. (a) Threshold pressure versus additional mass rate α_m obtained based on the linear theory analysis ($Q=20,10,8,6$). (b) Sound frequency versus additional mass rate α_m obtained based on the linear theory analysis ($Q=20,10,8,6$).

three variables were calculated: displacement of the reed $x(t)$, pressure inside the tube $p_2(t)$, and volume velocity $U(t)$. Parameter values used in the simulation are shown in Table II. Simulation results will be compared with both experimental and theoretical results.

A. Threshold behavior

The length of the tube was varied, and the tube length dependence of threshold pressure was investigated. Delay time in a reflection function was adjusted in inverse proportion to the tube length, and the blowing pressure was gradually increased until oscillation began or it reached 1 kPa. Figure 10(a) shows simulation results for some Q and additional mass values that were suggested by linear theory analysis. In the simulation, the same results were obtained for positive and negative pressure cases. A strong dependency on pipe length was observed, which is consistent with experimental results. There is a range in which the threshold is low, and also a range in which the reed cannot oscillate (210–350 Hz). These characteristics can be interpreted from

TABLE II. Parameters used in the time-domain simulation.

Variable	Symbol	Value
Reed		
Length	L	10 mm
Width	W	2 mm
Thickness	h	0.3 mm
Displacement at the tip	x	
Initial displacement	x_0	0
Gap between the reed and plate	b	0.01 mm
Resonance Q value	Q	8–35
Natural angular frequency	ω_r	$2\pi\times 470$ Hz
Material density	ρ_r	8×10^3 kg/m^3
Additional mass rate	α_m	0–0.3
Mass	m	$\rho_r WLh(1+\alpha_m)$
Tube		
Pressure inside the tube	p_2	
First resonance frequency	f_p	150–600 Hz
Radius	r_p	3.5 mm
Characteristic impedance	Z_0	$\rho c/\pi r_p^2$
Others		
Blowing pressure	p	0–1.0 kPa
Volume velocity through the slit	U	
Channel length	δ	1 mm
Flow contraction coefficient	C	0.61
Air density	ρ	1.2 kg/m^3
Sound velocity	c	340 m/s

the frequency relation between the pipe resonance frequency and natural frequency of the reed. It was also found that threshold pressure increased with Q and with additional mass.

However, there is a considerable discrepancy between Fig. 10(a) and 4: there is a wider nonoscillation range, or simulation results are shifted to the higher frequency. It is suspected that this is partially due to incorrectness of parameter values, and partially due to the nonlinearity of the model. To get more reasonable results, the simulation was redone using different Q and reed frequency f_r values, as shown in Fig. 10(b). By using more larger Q value, the agreement between experiment and simulation results was improved. Using the reed frequency $f_r=400$ also produced an acceptable result.

Next, threshold pressure was compared in the original pipe length $L_p=24.1$ cm ($f_p=353$ Hz) between simulation and theory. The threshold pressure by theory did not exactly coincide with the simulation. A Q value of 10 was suggested from linear theory analysis, but a larger value was shown to be appropriate in the simulation. It is inferred that this discrepancy is due to the nonlinear terms, which were neglected in the linear theory analysis. When $Q=35$, simulation and experiment showed a better correspondence, and the oscillation condition was also satisfied, i.e., $|K(f)|>1$ held.

B. Sounding frequency

Then, the relationship between the tube resonance frequency and the sounding frequency under constant pressure was examined. Figure 11 shows simulation results using the same parameters used in the preceding section. In contrast to Fig. 10, the Q value and mass affect sounding frequency only slightly. When $f_r=400$, the result was different from the others and from the experimental result, and was found to be inappropriate.

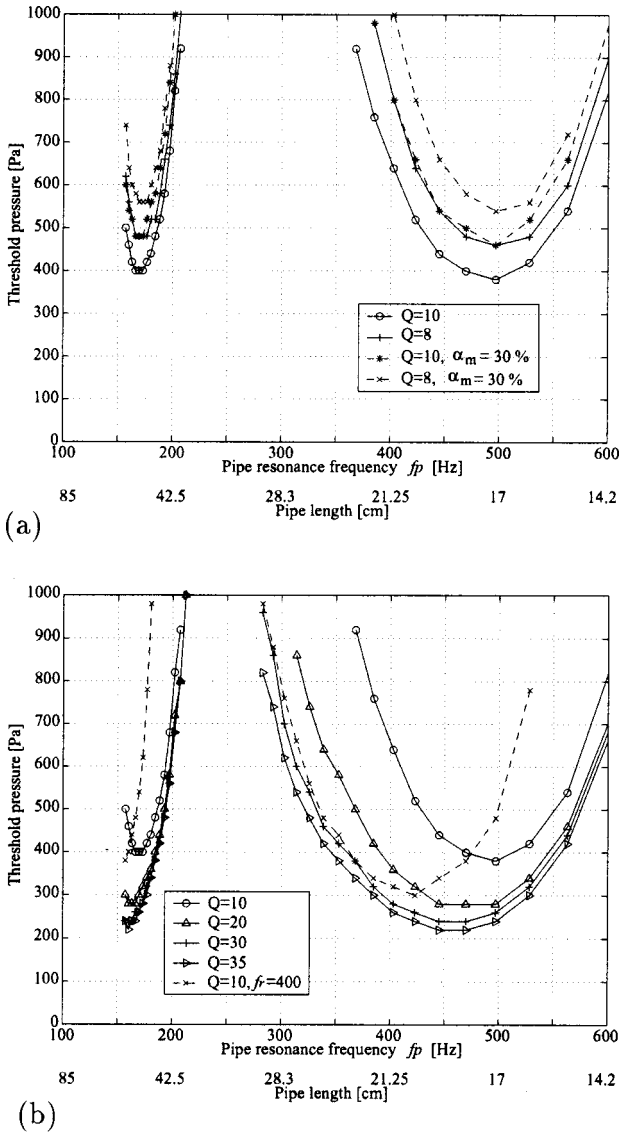


FIG. 10. (a) Tube resonance frequency f_p versus threshold pressure obtained by simulation with two Q values and additional mass rate α_m used as parameters. (b) Q values and reed frequency f_r are changed to fit to the experimental result.

The data in Fig. 11 are in fair agreement with those in Fig. 5. In the simulation, oscillation commenced above the reed frequency (dashed line in the figure), and also above the first or third multiple of the pipe resonance frequency (dashed–dotted lines), the same as for the experimental results. Pitch reached 570 Hz in the experiment, and 630 Hz in the simulation. Although the experimental result was limited to below 360 Hz, the sounding frequency rises with increasing tube resonance frequency in the simulation. In terms of the range of the pipe resonance frequency that can oscillate, $Q=35$ seems appropriate.

In the original pipe length, $L_p=24.1$ cm, the sound frequency obtained by simulation was 477 Hz, and the difference between the experiment and simulation was about 1.2%. The sound frequency obtained from linear theory was

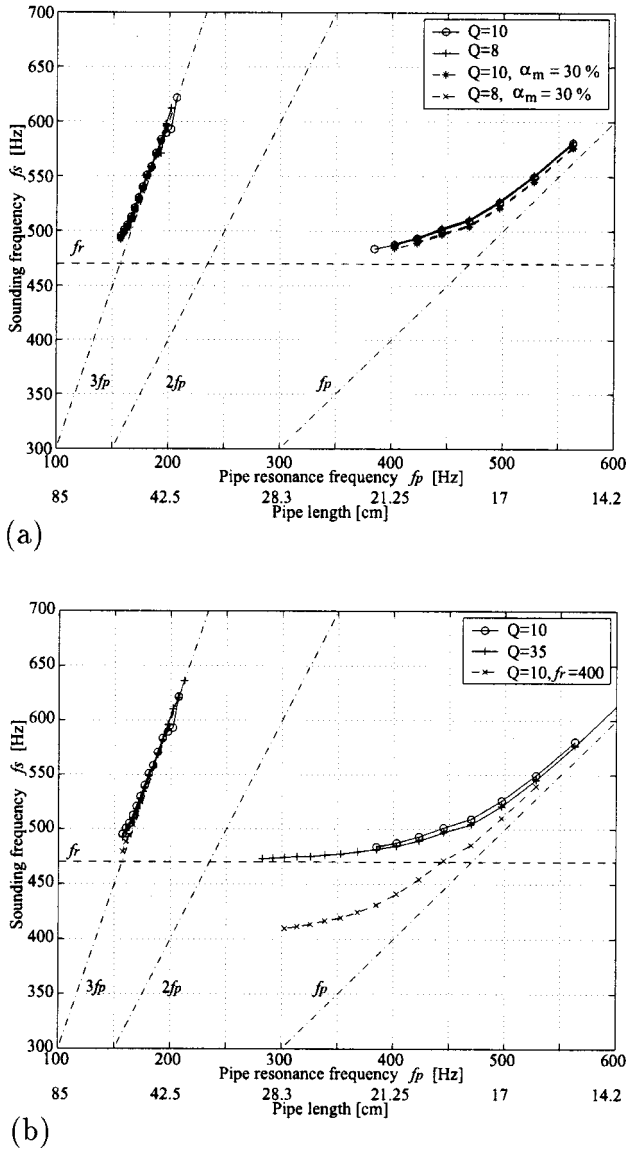


FIG. 11. (a) Tube resonance frequency f_p versus sounding frequency f_s obtained by simulation with two Q values and additional mass rate α_m used as parameters. Dashed–dotted lines show integer multiples of the pipe resonance frequency and dashed line shows the reed frequency. (b) Q values and reed frequency f_r are changed to fit to the experimental result.

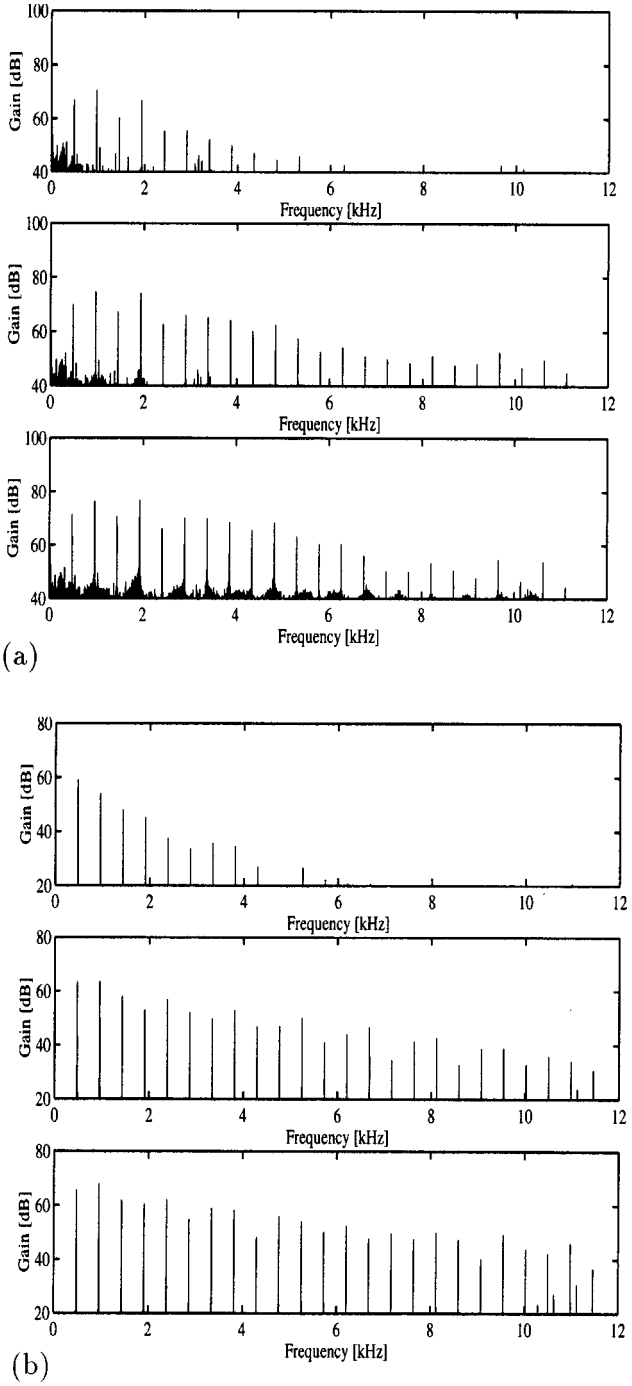


FIG. 12. Variation in spectra with different blowing pressure. (a) Spectra of recorded sound (positive, 0.4, 0.6, 0.8 kPa from the top). (b) Spectra of simulated sound (positive, 0.4, 0.6, 0.8 kPa from the top). Simulated sound was calculated by multiplying the simulated volume velocity by the transfer function of the pipe. The high frequency components increased with increasing blowing pressure, the same as for the recorded sounds.

417 Hz, so it shifted about 1.3% higher in the simulation based on the model that includes nonlinear terms.

C. Sound spectrum

Next, simulated sound spectra were compared with those obtained in the experiment. Figure 12(a) shows spectra of recorded sounds with different blowing pressures, i.e., 0.4, 0.6, 0.8 kPa from the top. High-frequency components in the spectra increased with increasing blowing pressure. The same tendency was found in the negative case.

Radiated sound pressure cannot be simulated using our model, but volume velocity signal can. According to Caussé's method,[13] given the pipe shape, the transfer function between the volume velocity at the entrance of a pipe and sound pressure at the exit of a pipe can be calculated. In the calculation of the transfer function, as a termination condition, spherical radiation is assumed. Sound pressure at the exit of the pipe was then calculated by multiplying the simulated volume velocity with this transfer function. Figure 12(b) shows spectra of simulated sounds with blowing pressures of 0.4, 0.6, 0.8 kPa from the top. The high frequency components increased with increasing blowing pressure, the same as for the recorded sounds.

The second and fourth harmonics of the spectra of the recorded sounds are enhanced, as shown in Fig. 12(a). It

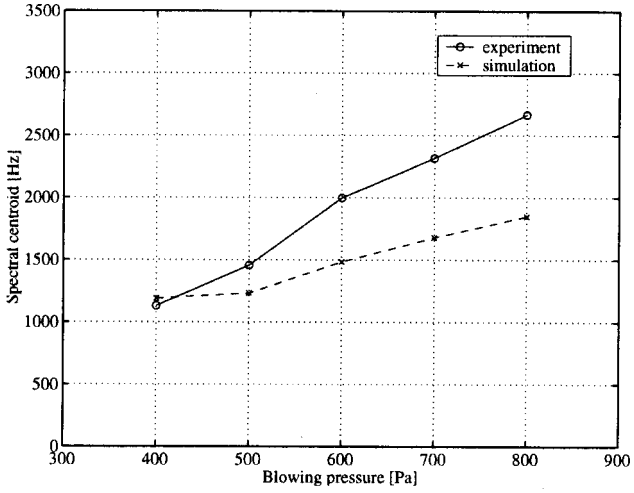


FIG. 13. Spectral centroid versus blowing pressure. The plots ○ and × show experimental and simulation results, respectively.

seems that this is due to the 3/4 lambda and 5/4 lambda resonances of the pipe. In order to find out whether this inference is correct or not, the transfer function of the pipe was calculated, and the frequencies of the 3/4 lambda and 5/4 lambda resonances were 1054 Hz and 1764 Hz, respectively. On the other hand, the frequencies of the second and fourth harmonics are about 966 Hz and 1932 Hz. So, it is concluded that predominance of the second and fourth harmonics is not due to the peaks of the transfer function of the pipe.

To evaluate changes in spectral components with blowing pressure quantitatively, spectral centroid defined by the following equation was calculated and compared:

$$C = \frac{F_s}{N} \frac{\Sigma_{k=0}^{N/2-1} |Y(k)| \cdot k}{\Sigma_{k=0}^{N/2-1} |Y(k)|},$$

where $Y(k)$ is the discrete FFT spectra, F_s the sampling frequency and N the FFT length. Figure 13 shows the spectral centroids of the recorded and simulated sounds. Although there is some discrepancy, both plots show similar trend: the spectral centroid gradually increases as blowing pressure increases.

V. CONCLUSIONS

We investigated the effects of tube length on the threshold pressure and on sounding frequency by simulation, and the result was proved to coincide with that obtained by experiment. Recorded and simulated sounds have a common feature in the sense that high-frequency components of their spectra increase with increasing blowing pressure. Further, from the frequency relationships, it was concluded that the reeds of the sho act as ''outward-striking valves.'' Especially, noteworthy is that the frequency relationship also holds even for negative pressure.

In terms of the relationship between the reed resonance frequency and one of the pipe resonance frequencies, the sound production mechanism can be explained as follows. First, the oscillation commences with a frequency *f*, which is near the reed frequency f_r and above one of the pipe resonance frequencies. When the finger hole is opened, the effective pipe length is shortened and the first pipe resonance frequency f_p becomes so high that the condition for oscillation is not satisfied. When the finger hole is closed, f_p becomes lower than f_r and the oscillation condition is satisfied.

Our model does not yet predict the finer details of the experimental results. Here, the difference in threshold pressure is considered. In the experiment, the threshold pressure was different for positive and negative pressures, whereas in simulation it perfectly coincided. In our model, the reed is assumed to be symmetrical and its thickness is neglected. But, the back of the reed is actually slightly chiseled. As a result, the reed goes out of the slot more easily when drawing than when blowing. By incorporating this asymmetry, it is expected that the difference between positive and negative pressures can be represented.

Measurements of the blowing pressure dependency of sounding frequency were also done, although we have not shown here. As stated in the introduction, previous studies on khaen reeds have shown a linear decrease of playing frequency with increasing blowing pressure. Cottingham *et al.* also reported same trends on harmonium-type reeds without a pipe resonator, and they showed theoretical calculations which agree well with the experimental data.[14] However, our measurement results obtained from all tubes showed variations among tubes. Pipe configuration and particles that were spread on the reeds might affect reed oscillation, so more work should be done before drawing conclusions.

We believe this dependency of sounding frequency on blowing pressure may play an important role in producing characteristic timbre of the sho. The sho is usually played in chords with gradual crescendo and decrescendo, which causes slight pitch variations on each note. In addition, traditional chords have many dissonant tones, from the Western tonal musical theory point of view, so sounds have inherent complex beats and keep changing with time.

From the sound-quality point of view, sounds produced by our model have basic characteristics of real tones, although they are somewhat monotonous. The reason is due to the fact that we used only simple control parameters with time. It is expected that use of realistic control parameters may make the sounds more realistic. Other future work which was not mentioned above is to improve the simulation by including the effects of radiation, vortices, and wall vibration of the instrument in the model.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Ken'ichiro Ishii, Director of the NTT Communication Science Laboratories, and Dr. Hiroshi Murase, Manager of the Media Information Laboratory for their support. They wish to thank reviewers for their helpful suggestions. They also thank CIAIR research members at Nagoya University for fruitful discussions. Part of this work is supported by the Center of Excellence (COE) formation program of the Ministry of Education, Culture, Sports, Science and Technology of Japan (No. 11CE2005).

APPENDIX

The complete derivation of Eq. (1) can be found in the Appendix in Ref. 7. Here, it is briefly summarized. First, the equation of motion of the cantilever is described as

$$\rho_r Wh \frac{\partial^2 \xi}{dt^2} + R \frac{\partial \xi}{\partial t} + K \frac{\partial^4 \xi}{\partial s^4} = W(p(t) - p_2(t)), \quad \text{(A1)}$$

where $\xi(s,t)$ is the displacement of the reed, s the distance from its clamped root, ρ_r the material density, W the width, h the thickness, K the bending stiffness of the reed, and R its damping coefficient.

When the reed is assumed to vibrate in the normal mode, we can write $\xi(s,t) = [x(t) - x_0]\Psi(s)$, where $\Psi(s)$ is the form of the mode function, normalized so that $\Psi(L) = 1$, where L is the reed length. Multiplying both sides of (A1) by $\Psi(s)$ and integrating over the reed length L then gives

$$\frac{d^2x}{dt^2} + \frac{\omega_r}{Q}\frac{dx}{dt} + \omega_r^2(x - x_0) = \frac{\gamma WL}{m}(p(t) - p_2(t)), \quad \text{(A2)}$$

where Q is the resonance Q value and m is the effective mass of the reed as given by

$$m = \rho_r WLh \quad \text{(A3)}$$

and

$$\gamma = \frac{\int_0^L \Psi(s)ds}{\int_0^L \Psi(s)^2 ds} \approx 1.5. \quad \text{(A4)}$$

From a knowledge of the form of $\Psi(s)$, we can also evaluate the vertical component of the side opening $a(x)$ when its tip opening is x.

$$a(x) = x_0 + (x - x_0)\int_0^L \Psi(s)ds \approx 0.6x_0 + 0.4x.$$

[1] C. Roads, "Physical modeling and formant synthesis," *The Computer Music Tutorial* (MIT Press, London, 1996), Chap. 7, pp. 263–315.
[2] J. O. Smith, "Physical modeling synthesis update," Comput. Music J. **20**, 44–56 (1996).
[3] G. P. Scavone and P. R. Cook, "Real-time computer modeling of woodwind instruments," Proceedings of the International Symposium on Musical Acoustics (Acoustical Society of America, New York, 1998), pp. 197–202.
[4] J. P. Cottingham and C. A. Fetzer, "Acoustics of the khaen," Proceedings of the International Symposium on Musical Acoustics (Acoustical Society of America, New York, 1998), pp. 261–266.
[5] J. P. Cottingham, "The acoustics of a symmetrical free reed coupled to a pipe resonator," Proceedings of the 7th International Congress on Sound and Vibration, 2000, pp. 1825–1832.
[6] J. P. Cottingham, "The Asian free reed mouth organs," in Proceedings of the International Symposium on Musical Acoustics (Fondazione Scuola di San Giorgio, Venezia, 2001), pp. 61–64.
[7] A. Z. Tarnopolsky, N. H. Fletcher, and J. C. S. Lai, "Oscillating reed valves-An experimental study," J. Acoust. Soc. Am. **108**, 400–406 (2000).
[8] R. T. Schumacher, "*Ab initio* calculations of the oscillations of a clarinet," Acustica **48**, 71–85 (1981).
[9] N. H. Fletcher, "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180 (1993).
[10] T. Hikichi and N. Osaka, "Measurements of the resonance frequencies and the reed vibration of the sho," Acoust. Sci. Technol. **23**, 25–27 (2002).
[11] S. Adachi and M. Sato, "Time-domain simulation of sound production in the brass instrument," J. Acoust. Soc. Am. **97**, 3850–3861 (1995).
[12] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York, 1998), Chap. 13, pp. 401–428.
[13] R. Caussé, J. Kergomard, and X. Lurton, "Input impedance of brass musical instruments—Comparison between experiment and numerical models," J. Acoust. Soc. Am. **75**, 241–254 (1984).
[14] J. P. Cottingham, C. H. Reed, and M. Busha, "Variation of frequency with blowing pressure for an air-driven free reed," J. Acoust. Soc. Am. **105**, Pt. 2, 1001 (1999).

The effect of superior auditory skills on vocal accuracy

Ofer Amir,[a] Noam Amir, and Liat Kishon-Rabin
Department of Communication Disorders, Sackler Faculty of Medicine, Tel-Aviv University, Israel

(Received 19 July 2002; revised 17 November 2002; accepted 17 November 2002)

The relationship between auditory perception and vocal production has been typically investigated by evaluating the effect of either *altered* or *degraded* auditory feedback on speech production in either normal hearing or hearing-impaired individuals. Our goal in the present study was to examine this relationship in individuals with *superior* auditory abilities. Thirteen professional musicians and thirteen nonmusicians, with no vocal or singing training, participated in this study. For vocal production accuracy, subjects were presented with three tones. They were asked to reproduce the pitch using the vowel /a/. This procedure was repeated three times. The fundamental frequency of each production was measured using an autocorrelation pitch detection algorithm designed for this study. The musicians' superior auditory abilities (compared to the nonmusicians) were established in a frequency discrimination task reported elsewhere. Results indicate that (a) musicians had better vocal production accuracy than nonmusicians (production errors of 1/2 a semitone compared to 1.3 semitones, respectively); (b) frequency discrimination thresholds explain 43% of the variance of the production data, and (c) all subjects with superior frequency discrimination thresholds showed accurate vocal production; the reverse relationship, however, does not hold true. In this study we provide empirical evidence to the importance of auditory feedback on vocal production in listeners with superior auditory skills. [DOI: 10.1121/1.1536632]

PACS numbers: 43.75.St, 43.70.Bk, 43.66.Hg [AL]

I. INTRODUCTION

Auditory feedback influences speech and vocal production in a complex manner. Typically, this relation has been studied extensively by examining the effect of either *altered* or *degraded* auditory feedback on speech production in either normal hearing or hearing-impaired population. Few studies, however, have examined this relation in populations with *superior* auditory abilities such as musicians. Our goal in the present study was to evaluate whether musicians, who demonstrate superior auditory skills, would also have higher vocal production accuracy.

Studies with normal-hearing individuals showed immediate voice changes when auditory feedback was altered: vocal intensity increased when individuals were subjected to background noise (also known as the Lombard effect),[1] the speech rate decreased when auditory feedback was artificially delayed,[2] and fundamental frequency changed when auditory feedback frequencies have been altered.[3] A more recent study reported changes in vowel production to compensate for feedback alterations in the first three formants of the vowel; changes that were large enough to influence the vowel's perceived phonetic identity.[4] These data support the hypothesis that auditory information is used in a closed-loop system, which provides moment-to-moment feedback for the control of vocal production.

Studies with the hearing impaired showed differences in the role of auditory feedback on speech production between those deafened after speech and language acquisition had been completed (postlingual) and those deafened before the age of two (prelingual). In *post-lingually* deafened adults, hearing loss had a minimal effect on speech *intelligibility* but a slow and gradual effect on certain speech and *vocal parameters*.[5–14] The data support the hypothesis of a predominantly open-loop speech motor control system once the speaker establishes the relationship between motor commands and resulting sound output (as occurs in individuals with the late onset of deafness). It is in those cases that the speaker uses their knowledge to compute the motor sequence for desired speech/vocal production in the absence of auditory feedback.[15]

In *prelingual* hearing-impaired children, the absence or partial auditory information prior to and during speech acquisition has a deleterious effect on speech production and its intelligibility.[16–18] These children develop abnormal phonemic-motor patterns because of their need to rely on visual, tactile and proprioceptive feedback.[19–21] The fact that the partial restoration of hearing after many years of auditory deprivation does not result in good speech production skills supports the nonlinear relationship between perception and production and the involvement of additional factors such as the plasticity of the speech production mechanism to accept changes.

While the hearing impaired represent one end of the auditory abilities spectrum, musicians are typically viewed as representing the other end of this spectrum. As discussed above, the deleterious effect of absent or degraded auditory abilities on speech and vocal production have been widely demonstrated. Yet, it is not clear whether individuals with exceptional auditory abilities (e.g., musicians) would also demonstrate better-than-normal vocal abilities.

The superior auditory performance of musicians has

[a]Electronic mail: oferamir@post.tau.ac.il

been established on tests that reflect specific facets of music and on basic psychoacoustic tasks. Musicians demonstrated superior processing of timbre and rhythm,[22] the identification of mistuned harmonics,[23] the labeling of musical intervals (frequency ratio),[24–26] musical memory,[27] and a smaller difference limen for frequency (DLF).[28–30]

Physiological data suggest that the differences in behavioral tests between musicians and nonmusicians stem from neurological and/or functional differences in the auditory system. Micheyl,[31] for example, found that musicians demonstrated a significant reduction in cochlear emission in response to contralateral stimuli, suggesting different auditory-nerve efferent activity in musicians compared to nonmusicians. Functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET) showed a pronounced hemispheral asymmetry in the planum temporal among musicians, which is assumed to be related to their superior auditory abilities.[32] Studies of Evoked Related Potentials (ERPs) reported musicians to exhibit a larger P_3 in response to music stimuli compared to nonmusicians.[33] Musicians also showed increased neural activity (using magneto encephalography) in response to musical tones compared to pure tones.[34]

The question of whether individuals with exceptional auditory abilities, such as musicians also demonstrate better-than-normal vocal production has been investigated directly in only two published studies. The first was conducted by Seashore in 1919.[35] In this pioneer study, Seashore asked a group of singing teachers to evaluate their students' singing accuracy. He then tested these students' DLF and concluded that there is ''a slight tendency toward relationship'' (p. 58). Nevertheless, this study should be examined with caution due to several methodological issues. The validity of the variables used in this study is difficult to evaluate. Singing accuracy was not evaluated directly. Instead, the participants' vocal ''brightness'' was rated, subjectively, by the teachers, with no reported reliability. Pitch discrimination, on the other hand, was evaluated as accurately as possible for that time (using a series of tuning forks). Moreover, Seashore himself raised doubts regarding the young participants' ability to comprehend the task requirements and present their actual musical capacity.

The second study to have addressed this question was conducted by Ternstrom, Sundberg, and Collden.[36] They asked a group of trained singers to sustain their pitch while producing different vowels. This task was performed both with normal auditory feedback and with masked auditory feedback. No control group was included in the study. They reported that the singers were less accurate in maintaining their pitch in the presence of background noise than with normal feedback.

Given the methodological concerns in these two studies: the absence of control groups and the fact that both studies examined the performances of trained singers and not musicians with no vocal training, it appears that the question of whether better *auditory abilities* result in improved vocal production has yet to be addressed. One can only speculate why this issue has not been investigated in depth. One possible explanation is that studies that focused on the relation between auditory perception and vocal production were interested primarily in pathological speech. This led to testing the theories in clinical populations, such the hearing impaired. Another possibility is that speech/voice production was viewed as inherently limited by the constraints of the articulatory system. Furthermore, any mispronunciations or inaccuracies can be resolved by the speaker's knowledge of the language. Thus, it might seem logical to assume that musicians would not produce voice more accurately that nonmusicians due to the objective mechanical constraints of the vocal production system. Finally, it is possible that the methodological challenges of measuring minute changes in vocal production and compare them with subtle perceptual parameters posed technological obstacles that made such a study more difficult to perform.

It is our belief, however, that investigating the relationship between exceptional auditory abilities and vocal production is of interest and may complement the existing data on the role of auditory feedback on vocal production. It will also shed light on the question of whether the importance of auditory feedback is unique to speech or can be extended to nonverbal stimuli.

Our purpose in this study, therefore, is to test whether musicians who have significantly better auditory frequency discrimination than nonmusicians, will exhibit better-than-normal performance on vocal production accuracy task. Prior to the present study, the DLF of 16 musicians and 14 nonmusicians were examined for reference tones 250, 1000, and 1500 Hz in a three-interval, three-alternative forced-choice adaptive procedure.[28] The musicians showed significantly better DLF than nonmusicians for all frequencies. Once the superior auditory performance of musicians has been established, we proceeded to test 26 of these subjects (13 musicians and 13 nonmusicians) in an accuracy imitative vocal production task. It is our purpose in this paper to report on the results of the production task and on the comparison between perception and production performance in musicians and nonmusicians.

II. METHOD

A. Subjects

Twenty-six male subjects participated in the study: 13 were professional musicians and 13 nonmusicians, approximately matched in age and education. The musicians were 20–33 years of age (average 25 years old), playing at least one musical instrument for 7–24 years (an average of 13 years). All of them were members of a formal musical group (an orchestra or a band).

The nonmusicians were 23–34 years of age (average 27 years old). These subjects had no previous musical training (less than 1 year) or experience in psychoacoustic testing. All subjects had no previous vocal and singing training or experience. All subjects had pure-tone air-conduction thresholds less than 15 dB HL bilaterally at octave frequencies from 250–4000 Hz.[37] Thresholds for relative DLF were established for each participant prior to the collection of the production data, as reported extensively in Kishon-Rabin *et al.*[28] These data are summarized in Table I for each subject and

TABLE I. Individual participants' relative DLF (*rel*DLF%) for the three tones tested, based on the data presented in Kishon *et al.* (2001).

		*rel*DLF%		
Group	Subject	250 Hz	1000 Hz	1500 Hz
Musicians	1	0.95	0.26	0.34
	2	0.37	0.29	0.26
	3	0.60	0.36	0.27
	4	1.40	0.44	0.80
	5	1.65	0.56	0.48
	6	1.70	1.14	1.05
	7	1.80	0.70	0.73
	8	0.85	0.45	0.59
	9	1.30	0.61	0.62
	10	0.47	0.10	0.33
	11	0.92	0.26	0.67
	12	0.87	0.23	0.31
	13	0.57	0.45	0.34
Nonmusicians	1	2.30	1.09	1.19
	2	2.97	1.68	1.12
	3	1.05	0.61	0.31
	4	3.77	1.96	1.32
	5	2.05	1.00	1.02
	6	2.20	0.34	0.45
	7	1.77	1.05	1.11
	8	3.42	1.58	1.23
	9	2.02	0.69	1.04
	10	2.27	0.90	1.18
	11	2.30	0.63	1.30
	12	1.67	0.58	1.03
	13	3.32	1.83	1.25

frequency. Note that the values are expressed in percentage relative DLF (*rel*DLF% $= \Delta f / f^{*}100$).

B. Stimuli

Three reference tones at frequencies 131, 165, 196 Hz (C3, E3, G3, respectively) were selected as representing the mid-range frequencies of the average untrained male voice register.[8] The sine waves were generated digitally using the Sound Forge 4.5 computer program (version 4.5 g, Sonic Foundry, Inc.) at a sampling rate of 22 050 Hz, 16 bits/sample, with a duration of 2 s, and were stored on a hard disk of a personal computer.

C. Procedure

The subjects stood in a quiet room 15 cm from a dynamic Sony microphone (F-170). Signals were presented to the subjects binaurally, through headphones (MDR-CD270) directly from the computer at 80–85 dB SPL.[37]

Each tone was presented three times, totaling nine target stimuli. These were then presented in random order. The subjects were instructed to listen to each stimulus until it ended and then reproduce it, using the vowel /a/ at the same pitch as accurately as possible. The subjects' productions were recorded directly into a computer using a sampling rate of 22 050 Hz. Each production lasted approximately 2 s. The subjects were also asked to produce a vocal sweep of frequencies in order to ensure that the stimuli were within their dynamic vocal range.

D. Vocal analysis

1. Pitch detection algorithm

Pitch detection was performed by computing the autocorrelation over successive windows of 30 ms, with an overlap of 20 ms. The location of the largest local maximum in the autocorrelation curve was taken to be the fundamental period at that window. We remark that when this method is applied to the pitch detection of normal speech, it is prone to false detection under certain circumstances, such as the presence of strong high harmonics and a weak fundamental frequency. Nevertheless, in the present study such conditions did not occur.

The resolution of this method is limited by the sampling rate, giving a different *relative* error for each detected frequency. Specifically to this study, the fundamental periods for frequencies 131, 165, and 196 Hz are 168.32, 133.64, and 112.5 samples. Since the maximum error in detecting the peak of the autocorrelation function can be half a sample, adding 0.5 to each of these periods and translating back to frequencies gives 131.25, 164.55, and 195.13 Hz. The maximum relative errors are thus 0.3%, 0.37%, and 0.44%, respectively. These percentages give an upper bound on errors due to the limited frequency resolution in the vicinities of frequencies used here. In order to improve the resolution, the autocorrelation curve was interpolated by a factor of 4, using FIR interpolation. This reduced the upper bounds on relative resolution errors to 0.07%, 0.09%, and 0.11%, respectively. Thus, the resolution errors are far below the production errors themselves, as shown in the next section, and are further reduced by averaging over the utterances.

2. Applying the pitch detection routine

An analysis was performed by presenting the experimenter with a graphic window containing the recorded production. The experimenter selected the middle 50% of each file. The fundamental frequency was computed over this segment, and averaged. If the chosen section presented exceptional instability (>4%) in frequency or intensity, a similar section from another part of the recording was analyzed. In recordings with no stable section at the initially required length, a shorter section was used, subject to the condition that it would not be shorter than 0.5 s. In addition, a randomly chosen set of 20% of the responses was remeasured, by the same judge and by a second judge, to evaluate interjudge and intrajudge reliability of the fundamental frequency measurements. Correlations between original and repeated measurements were $\mathbf{r}=0.99$, $p<0.001$ for interjudge reliability and $\mathbf{r}=1$, $\mathbf{p}<0.001$ for intrajudge reliability.

III. RESULTS

As described above, each participant produced the target tones (131, 165, and 196 Hz) three times. The three fundamental frequency measurements were averaged for each target frequency and participant. The mean individual production data for the three frequencies are presented in Table II.

The distribution of the production values for each target tone within the two groups are illustrated in Fig. 1. In this box plot graph, the box represents the interquartile range,

TABLE II. Fundamental frequencies (in Hz) of the productions performed by each musician and nonmusician for each of the three target tones (values reported represent means of three repetitions of each production).

Group	Subject	Target tone		
		131 Hz	165 Hz	196 Hz
Musicians	1	132.86	172.88	195.44
	2	129.02	165.45	196.49
	3	125.73	159.92	188.68
	4	123.25	158.12	190.58
	5	124.18	153.44	157.57
	6	131.45	157.15	183.02
	7	133.54	165.43	195.88
	8	127.07	167.84	195.22
	9	127.93	163.56	196.79
	10	127.43	157.31	197.71
	11	131.68	163.59	192.84
	12	123.20	162.59	187.93
	13	129.02	166.75	195.75
Nonmusicians	1	106.61	184.40	232.75
	2	118.29	148.23	173.09
	3	134.96	167.82	192.95
	4	129.52	162.40	180.78
	5	128.05	163.38	194.33
	6	131.03	158.61	196.95
	7	126.91	157.07	185.03
	8	182.60	279.84	327.75
	9	128.12	142.22	165.32
	10	126.41	142.73	148.29
	11	113.26	168.38	197.49
	12	128.80	161.63	186.02
	13	88.92	88.18	161.90

which contains 50% of the values. The line within the box marks the median, the whiskers above and below the box extend to the 90th and 10th percentiles, and the outlying data are graphed as filled circles. Clearly, the nonmusicians group had a wider range of values than the musicians group. Standard deviations for the nonmusicians group were markedly larger than for the musicians group (21.03 vs 3.51, 41.71 vs 5.30 and 44.80 vs 10.75 for frequencies 131, 165, and 196 Hz, respectively). Tests for the equality of variance revealed

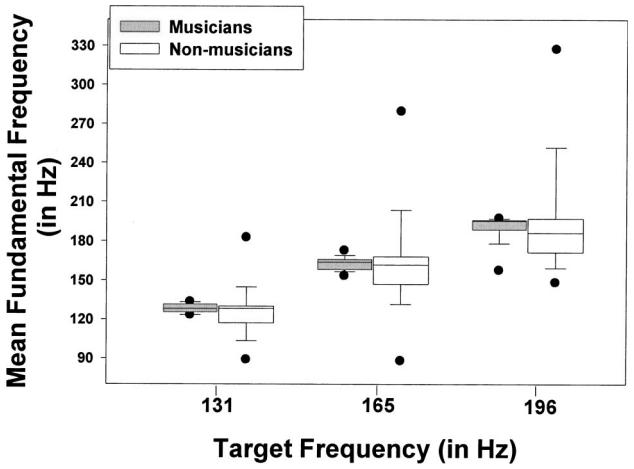


FIG. 1. Distribution of the fundamental frequencies produced by musicians and nonmusicians for each target tone. The box represents the interquartile range, which contains 50% of values. The line within the box marks the median, the whiskers above and below the box extend to the 90th and 10th percentiles, and the outlying data are graphed as filled circles.

that these group differences were statistically significant for all frequencies ($\mathbf{p}<0.0005$). In addition, to evaluate the intrasubject reproducibility between the three tones produced by each subject for each frequency, an intraclass correlation was employed, yielding a Cronbach's alpha of 0.92 for the musicians group and 0.63 for the nonmusicians group. Note that the majority of the vocal productions (approximately 72%) were produced at frequencies *lower* than the expected frequencies.

The accuracy of the vocal production was calculated as the absolute difference between the observed fundamental frequency and the reference frequency relative to the reference frequency in percent. This measure, which we termed *relative accuracy* (*rel*Accuracy%), is assumed to reflect the accuracy of production. This value decreases as the difference between the observed frequency of vocalization and the target frequency decreases. For example, for a reference tone of 131 Hz and a measured production of 144 Hz, the *rel*Accuracy% is 9.92% ($100*|131-144|/131$). Means of the *rel*Accuracy% for both groups are presented in Table III. Data are presented separately for the three tones as well as a calculated mean value for each participant. In addition, the mean frequency discrimination threshold (in *rel*DLF%), adopted from Kishon-Rabin *et al.*[28] is reported for each participant. Note that in approximately 3% of the measurements shown in Table II, production was closer to one octave above or below the target frequency. In these cases, the reference frequency was adjusted accordingly and presented in Table III. For example, subject 13 of the nonmusicians produced 88.92 Hz when the target was 131 Hz. In this case, the reference frequency was considered 65.5 Hz (131/2) and the *rel*Accuracy% computed as 35.57% (Table III).

The *rel*Accuracy% grand mean (combining all three tones) was 2.88% (SD=2.67) for the musicians group, and 8.94% (SD=7.53) for the nonmusicians group. Thus, the musicians group produced the tones approximately three times more accurately than the nonmusician group. Using an analysis of variance with repeated measures (MANOVA) with Group as a fixed factor and Frequency as the repeated factor, these group differences were found to be statistically significant [$\mathbf{F}(1.24)=4.48$, $\mathbf{p}<0.05$]. However, no significant differences were found among the three frequencies ($\mathbf{p}=0.95$), as well as no Frequency X Group interaction ($\mathbf{p}=0.80$). Also, an Equality-of-Variance Two-Sample T-Test revealed a significantly larger distribution of the *rel*Accuracy% values in the nonmusicians group, in comparison to the musicians group ($\mathbf{p}<0.0005$).

A. Relation between frequency discrimination and accuracy of production

A Pearson correlation was performed between frequency discrimination and production using the *rel*DLF% and the *rel*Accuracy% averaged each across the tested frequencies for each subject. This correlation, for the two groups combined, is illustrated in Fig. 2. A significant correlation was found between the two measures ($\mathbf{r}=0.67$, $\mathbf{p}<0.001$). This analysis suggests that approximately 43% of the variance of the production data can be explained by auditory perception. Figure 2 also demonstrates the relatively small between-

TABLE III. Individual production data (in *rel*Accuracy%) and perceptual data (in *rel*DLF%) (Ref. 28) of the participants in the musicians (M) and the nonmusicians (NM) groups.

Group	Participant	*rel*Accuracy 131 Hz	165 Hz	196 Hz	Mean value	*Rel*DLF% Mean value
M	1	1.42	4.77	0.29	2.16	0.52
	2	1.51	0.27	0.25	0.68	0.31
	3	4.03	3.08	3.74	3.61	0.41
	4	5.92	4.17	2.77	4.28	0.88
	5	5.21	7.01	19.61	10.61	0.90
	6	0.34	4.76	6.62	3.91	1.30
	7	1.94	0.26	0.06	0.75	1.07
	8	3.00	1.72	1.69	1.71	0.63
	9	2.34	0.87	0.40	1.21	0.84
	10	2.73	4.66	0.40	2.75	0.30
	11	0.52	0.85	0.87	1.00	0.62
	12	5.95	1.46	4.12	3.84	0.47
	13	1.51	1.06	0.52	0.90	0.46
Mean		2.80	2.69	3.14	2.88	
(SD)		(1.92)	(2.18)	(5.34)	(2.67)	
NM	1	18.62	11.76	18.75	16.38	1.53
	2	9.70	10.16	11.69	10.52	1.92
	3	3.02	1.71	1.56	2.10	0.66
	4	1.13	1.58	7.77	3.49	2.35
	5	2.25	0.98	0.85	1.36	1.36
	6	0.02	3.87	0.48	1.46	1.00
	7	3.12	4.81	5.60	4.51	1.31
	8	39.39	15.20	16.40	23.66	2.08
	9	2.20	13.81	15.65	10.55	1.25
	10	3.50	13.50	24.34	13.78	1.45
	11	13.54	2.05	0.76	5.45	1.41
	12	1.68	2.04	5.09	2.94	1.09
	13	35.75	6.88	17.40	20.01	2.13
Mean		10.30	6.80	9.72	8.94	
(SD)		(13.29)	(5.36)	(8.12)	(7.53)	

subject variability for both perception and production in the musicians group compared to the nonmusicians group.

The data in Fig. 2 shows that *rel*DLF% of 12 of the 13 musicians is under 1.1 and the same proportion of the production accuracy of musicians is less than 4.3%. Furthermore, 85% (11/13) of the musicians have auditory perception *and* vocal production accuracy of less 1.1 and 4.3, respectively. These musicians are within the performance range indicated by the horizontal and vertical arrows in Fig. 2. In contrast, only three nonmusicians fall within this range of performance. It can also be seen that all subjects but one (regardless of musical experience) showed good production accuracy for *rel*DLF% smaller than 1.1. For perception thresholds greater than 1.1, the production data demonstrate greater variability: four of the nonmusicians have *rel*Accuracy% of less than 6, whereas the other six remaining subjects in this group have values of 10 to 24.

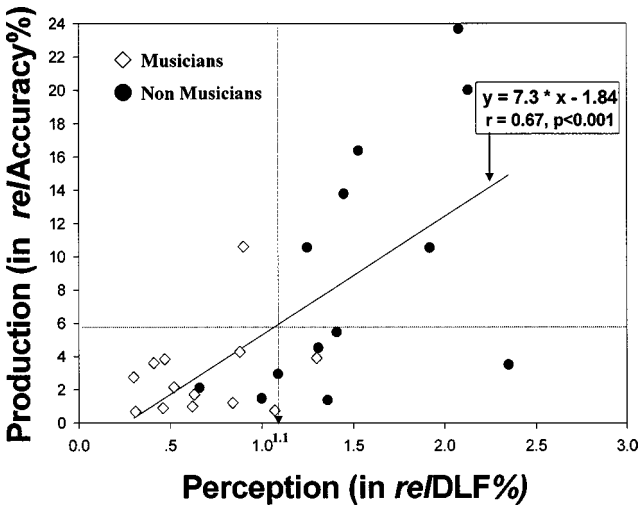


FIG. 2. Individual production data (*rel*Accuracy%) as a function of individual perception data (*rel*DLF%) (Ref. 28) for musicians (open symbols) and nonmusicians (filled symbols). The solid line represents the best fitting linear function for *all* data. The arrows represent the boundary range of performance of 12 of the 13 musicians for perception (vertical arrow) and production (horizontal arrow).

IV. DISCUSSION

In this paper we investigated the role of auditory perception on vocal production in a population with exceptional auditory abilities. If such individuals, who had no previous experience in voice training, show better-than-normal vocal production accuracy, it could have important implications on the importance of auditory feedback for vocal production that may be not specific to speech. Such information complements existing investigations on the perception–production relationship, which used primarily degraded or altered auditory feedback and verbal stimuli.

The current results indicate that, as a group, musicians who showed exceptional frequency discrimination ability also showed greater vocal production accuracy. This finding

is highlighted by the fact that these musicians had no formal vocal experience. Thus, it is possible that listeners use immediate auditory feedback for vocal production. The interesting question remains *how* do musicians use the auditory information to vocalize accurately. It is possible that they are tuned to acoustic parameters in vocal production that are otherwise ignored by nonmusicians. Another hypothesis is that musicians are able to transfer the underlying assumptions of the ''motor theory'' for speech[38] to the perception of auditory stimuli produced by musical instruments. The motor theory suggests that the relationship between perception and production of speech stems from the listeners ability to translate acoustic patterns to articulatory gestures and *vice versa*. It is possible that musicians develop mental representations of sounds as they are produced by musical instruments and then translate it, when producing sounds via the human vocal system. Furthermore, musicians may have had many years of fine auditory perception to motoric-production training. This hypothesis is supported by the finding of reduced intersubject variability (in both perception and production) in the musicians' group, which is commonly observed in studies where learning has occurred. Clearly, many of these issues need to be substantiated empirically in future studies.

An additional interpretation of our results is derived when converting the data to semitones. The musicians group had average production errors that were no greater than half of a semitone for each frequency. In contrast, the nonmusicians had mean errors of approximately 1.3 semitones. Keeping in mind the fact that the musical scale is based on notes that are defined in semitones units, inaccuracies that are greater than one semitone are perceived as a melody change. Thus, plus or minus one-half semitone may be viewed by musicians as a musical boundary (analogous to categorical boundary), where ''crossing'' this boundary creates a musical meaningful difference. Alternatively, inaccuracies less than one semitone could create the subjective feeling of a ''mistune,'' but would not create a meaningful difference. It should be noted that the nonmusicians in the present study demonstrated larger vocal inaccuracies compared to those reported in Weiner *et al.*[39] It is difficult, however, to discuss these differences due to the lack of background information regarding the musical training of the Weiner *et al.* subjects.

When looking at the individual relationship between perception and production data we found that perception explained approximately 43% of the variance of the production data. All listeners but one, regardless of musical experience, that had superior frequency discrimination (*rel*DLF% less than 1.1) demonstrated accurate pitch vocalization (*rel*Accuracy% between 0.27 and 4.5). However, listeners with poor frequency discrimination (*rel*DLF% greater than 1.1) were divided in terms of their vocal ability: six of them showed poor vocal pitch accuracy (*rel*Accuracy% greater than 10), whereas five subjects showed accurate production (*rel*Accuracy% less than 6). Although these findings emphasize the importance of auditory frequency discrimination for the accurate pitch production of non-verbal sounds, they also suggest that subjects may be able to use other musical skills and/or different mechanisms to vocalize accurately.

It should be noted that although the correlation analysis was based on the average of the tested frequencies for both perception and production, it might be more reasonable to correlate one frequency at a time. The underlying assumption would be that accuracy in production frequency is related to frequency discrimination at that frequency range. A reanalysis of the data correlating auditory frequency discrimination only at 250 Hz, to vocal accuracy at 131–196 Hz resulted in r^2 of 0.31, a value smaller than that observed for the mean frequencies. Thus, our data did not support the assumption that perception and production accuracy should be tested with the same frequency. We assumed that averaging the tested frequencies for a single measure for both auditory frequency discrimination and vocal accuracy is valid because no statistical differences were found between the tested frequencies and by doing so, the intrasubject variability is reduced. Nonetheless, we recommend that future studies explore the importance of using the same tested frequency in both auditory perception and production accuracy and the carryover to other frequencies.

In summary, in the present study we provide empirical evidence of the important role that auditory feedback has in vocal production when superior though nonvocal musical skills are involved. Specifically, individuals with superior frequency discrimination abilities were able to vocally imitate pure tones with great accuracy. Frequency discrimination thresholds, however, could not be predicted from production accuracy. It appears that while all individuals with small *rel*DLF% exhibited accurate vocalization, some subjects exhibited accurate pitch vocalization, despite poor frequency discrimination. These individuals may be using other auditory abilities that were not included in the present study but may be linked to the vocal task evaluated here. Future studies examining the relationship between perception and production should include several auditory perceptual and production tasks. The present data also shed light on the importance of auditory experience on improved vocalizations. This may have implications on vocal training of singers. It would be of interest to investigate whether auditory training improves vocalization.

ACKNOWLEDGMENTS

The authors would like to thank Ms. Y. Vexler and Ms. Y. Zaltz for their contribution in the collection of the data. We would also like to thank Ms. Eti Shabtai for the statistical analyses.

[1] H. L. Lane and B. Tranel, ''The Lombard sign and the role of hearing in speech,'' J. Speech Hear. Res. **14**, 677–709 (1971).
[2] I. Davidson, ''Sidetone delay, reading rate, articulation and pitch,'' J. Speech Hear. Res. **2**, 266–270 (1959).
[3] J. Elman, ''Effects of frequency-shifted feedback on the pitch of vocal productions,'' J. Acoust. Soc. Am. **70**, 45–50 (1981).
[4] J. F. Houde and M. I. Jordan, ''Sensorimotor adaptation in speech production,'' Science **279**, 1213–1216 (1998).
[5] G. Plant, ''The speech of adults with acquired profound hearing losses: I. A perceptual evaluation,'' Eur. J. Disord. Commun. **28**, 273–288 (1993).
[6] G. Plant, ''The effects of an acquired profound hearing loss on speech production,'' Br. J. Audiol. **18**, 39–48 (1984).
[7] T. Read, ''Improvement in speech production following use of the UCH/RNID cochlear implant,'' J. Laryngol. Otol. Suppl. **18**, 45–49 (1989).
[8] R. S. Waldstein, ''Effects of postlingual deafness on speech production:

Implications for the role of auditory feedback,'' J. Acoust. Soc. Am. **88**, 2099–2114 (1990).
[9] C. Binnie, R. Daniloff, and H. Buckingham, ''Phonetic disintegration in a five-year old following sudden hearing loss,'' J. Speech Hear Disord. **47**, 181–189 (1982).
[10] S. B. Leder and J. B. Spitzer, ''Longitudinal effects of single-channel cochlear implantation on voice quality,'' Laryngoscope **100**, 395–398 (1990).
[11] S. B. Leder and J. B. Spitzer, ''Speaking fundamental frequency, intensity, and rate of adventitiously profoundly hearing-impaired adult women,'' J. Acoust. Soc. Am. **93**, 2146–2151 (1993).
[12] H. L. Lane and J. Webster, ''Speech deterioration in postlingually deafened adults,'' J. Acoust. Soc. Am. **89**, 859–866 (1991).
[13] S. B. Leder, J. B. Spitzer, J. Kirchner, C. Phillips, P. Milner, and F. Richardson, ''Voice intensity of prospective cochlear implant candidates and normal hearing males,'' Laryngoscope **97**, 224–227 (1987a).
[14] S. B. Leder, J. B. Spitzer, C. Phillips, P. Milner, J. Kirchner, and F. Richardson, ''Speaking rate of adventitiously deaf male cochlear implant candidates,'' J. Acoust. Soc. Am. **82**, 843–846 (1987b).
[15] M. L. Matthies, M. A. Svirsky, H. L. Lane, and J. S. Perkell, ''A preliminary study of the effects of cochlear implants on the production of sibilants,'' J. Acoust. Soc. Am. **96**, 1367–1373 (1994).
[16] H. Levitt and H. Stromberg, ''Segmental characteristics of the speech of hearing impaired children: Factors affecting intelligibility,'' in *Speech of the Hearing Impaired; Research, Training, and Personnel Preparation*, edited by I. Hochberg, H. Levitt, and M. J. Osberger (University Park Press, Baltimore, 1983).
[17] C. R. Smith, ''Residual hearing and speech production in deaf children,'' J. Speech Hear. Res. **18**, 795–811 (1975).
[18] D. Ling, *Speech and the Hearing Impaired Child* (A. G. Bell Association for the Deaf, Washington, DC, 1976).
[19] M. A. Svirsky and S. B. Chin, ''Speech production,'' in *Cochlear Implants*, edited by S. B. Waltzman and N. L. Cohen (Thieme Medical Publishers, New York, 2000), pp. 293–309.
[20] M. J. Osberger, M. Maso, and L. K. Sam, ''Speech intelligibility of children with cochlear implants, tactile aids or hearing aids,'' J. Speech Hear. Res. **36**, 186–203 (1993).
[21] E. A. Tobey, S. Angelette, C. Murchison, J. Micosia, S. Sprague, S. J. Staller, J. A. Brumacombe, and A. L. Beiter, ''Speech production performance in children with multichannel cochlear implants,'' Am. J. Otol. **12**, 165–173 (1991a).
[22] M. Prior and G. A. Troup, ''Processing of timbre and rhythm in musicians and non-musicians,'' Cortex **24**, 451–456 (1988).
[23] S. Koelsch, E. Schroger, and M. Tervaniemi, ''Superior pre-attentive auditory processing in musicians,'' NeuroReport **10**, 1309–1313 (1999).
[24] E. M. Burns and A. J. M. Houtsma, ''The influence of musical training on the perception of sequentially presented mistuned harmonics,'' J. Acoust. Soc. Am. **106**, 3564–3570 (1999).
[25] W. D. Ward, ''Absolute Pitch, Part II,'' Sound **2**, 33–39 (1963).
[26] E. M. Burns and D. W. Ward, ''Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals,'' J. Acoust. Soc. Am. **63**, 456–468 (1978).
[27] R. W. Lundin, *An Objective Psychology of Music* (The Ronald Press Company, New York, 1967), pp. 21–29.
[28] L. Kishon-Rabin, O. Amir, Y. Vexler, and Y. Zaltz, ''Pitch discrimination: Are professional musicians better than non-musicians,'' J. Basic Clin. Physiol. Pharmacol. **12**, 125–144 (2001).
[29] B. C. J. Moore and R. W. Peters, ''Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity,'' J. Acoust. Soc. Am. **91**, 2881–2893 (1992).
[30] M. F. Spiegel and C. S. Watson, ''Performance on frequency discrimination tasks by musicians and non-musicians,'' J. Acoust. Soc. Am. **76**, 1690–1696 (1984).
[31] K. Micheyl, ''Difference in choclear efferent activity between musicians and non-musicians,'' NeuroReport **8**, 1047–1050 (1997).
[32] R. J. Zatorre, D. W. Perry, C. A. Beckett, C. F. Westbury, and A. C. Evans, ''Functional anatomy of musical processing in listeners with absolute pitch and relative pitch,'' Proc. Natl. Acad. Sci. U.S.A. **95**, 3172–3177 (1998).
[33] Y. Barnea, ''Absolute pitch: Electrophysiological evidence,'' Int. J. Psychophysiol **16**, 29–38 (1994).
[34] C. Pantev, R. Ostenveld, A. Engelien, B. Ross, L. E. Roberts, and M. Hoke, ''Increased auditory cortical representation in musicians,'' Nature (London) **392**, 811–813 (1998).
[35] C. E. Seashore, *The Psychology of Musical Talent* (Silver, Burdett and Company, Boston, 1919), pp. 59–74.
[36] S. Ternstrom, J. Sundberg, and A. Collden, ''Articulatory f0 perturbations and auditory feedback,'' J. Speech Hear. Res. **31**, 187–192 (1988).
[37] ANSI S3.6-1989, ''Specifications for Audiometers'' (ANSI, New York, 1989).
[38] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, ''The discrimination of speech sounds with and across phoneme boundaries,'' J. Exp. Psychol. **54**, 358–368 (1957).
[39] J. B. Weiner, L. Lee, J. Cataland, and J. C. Stemple, ''An assessment of pitch-matching abilities among speech-language pathology graduate students,'' Am. J. Speech Lang. Pathol. **5**, 91–95 (1996).

Surface response of a viscoelastic medium to subsurface acoustic sources with application to medical diagnosis

Thomas J. Royston,[a] Yigit Yazicioglu, and Francis Loth
University of Illinois at Chicago, Chicago, Illinois 60607

(Received 2 August 2002; revised 8 November 2002; accepted 18 November 2002)

The response at the surface of an isotropic viscoelastic medium to buried fundamental acoustic sources is studied theoretically, computationally and experimentally. Finite and infinitesimal monopole and dipole sources within the low audible frequency range (40–400 Hz) are considered. Analytical and numerical integral solutions that account for compression, shear and surface wave response to the buried sources are formulated and compared with numerical finite element simulations and experimental studies on finite dimension phantom models. It is found that at low audible frequencies, compression and shear wave propagation from point sources can both be significant, with shear wave effects becoming less significant as frequency increases. Additionally, it is shown that simple closed-form analytical approximations based on an infinite medium model agree well with numerically obtained ''exact'' half-space solutions for the frequency range and material of interest in this study. The focus here is on developing a better understanding of how biological soft tissue affects the transmission of vibro-acoustic energy from biological acoustic sources below the skin surface, whose typical spectral content is in the low audible frequency range. Examples include sound radiated from pulmonary, gastro-intestinal and cardiovascular system functions, such as breath sounds, bowel sounds and vascular bruits, respectively. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1536153]

PACS numbers: 43.80.Cs, 43.80.Ev, 43.80.Qf, 43.80.Vj [FD]

I. INTRODUCTION

Passive listening (auscultation) has been used *qualitatively* by physicians to aid in the diagnosis of a wide range of medical conditions, involving for example the pulmonary system (e.g., breath sounds), the gastro-intestinal system (e.g., bowel sounds) and the cardiovascular system (e.g., bruits caused by partially occluded arteries). Recently, researchers have tried to apply more quantitative measurement and analysis techniques to increase the diagnostic utility of the ''passive listening'' approach in all of these areas (e.g., Refs. 1–8).

A common characteristic in all these applications is that the sounds produced by biological sources are dominantly in the low audible frequency range, nominally less than 1000 Hz. Unlike in the case of medical ultrasonic imaging, at these lower frequencies one cannot necessarily assume that dilatational (compression) waves dominate the response relative to other wave types which may be present and may be dispersive, including shear and surface waves. Additionally, multiple reflections and standing wave patterns may also be more prevalent, rendering a ray acoustics approach to the problem intractable. Nonetheless, the unique information about structure and function that may be obtainable from the passive monitoring of sounds, as opposed to medical imaging methodologies, warrants further investigation of sound propagation in the body at low audible frequencies.

As a fundamental step in addressing the complexities described above, in the present article we focus on the behavior of elementary acoustic sources, monopoles and dipoles, in a viscoelastic medium comparable to soft biological tissue. In all cases, the interest is on the resulting surface motion of the medium due to the buried elementary source. While only a handful of studies can be found in the medical diagnostic literature that mathematically treat this scenario, there are a larger number of studies in the seismology literature treating buried sources in elastic and viscoelastic half-spaces that can be adapted to the biological problem. In addition to the surface problem (theoretically treated as a half-space here), an infinite space response is also of interest as some skin surface-based sensors may have an impedance comparable to soft tissue and approximate an anechoic boundary condition.

Closed-form analytical or approximate solutions for the surface response to buried dipoles and monopoles could be particularly useful in the analysis and design of novel sensor configurations, as any distributed complex source can be approximated in terms of a finite number of elementary sources.

While ultimately interest may lie more in infinitesimal (point) sources, finite monopole and dipole solutions are also considered in order to understand differences and to help in comparisons with experimental and computational finite element studies.

Consequently, the objectives of the present work are as follows.

(1) Review and consolidate relevant theoretical solutions for finite and infinitesimal monopole and dipole sources in viscoelastic media of infinite and semi-infinite (half-space) bounds;

[a]Author to whom correspondence should be addressed. Electronic mail: troyston@uic.edu

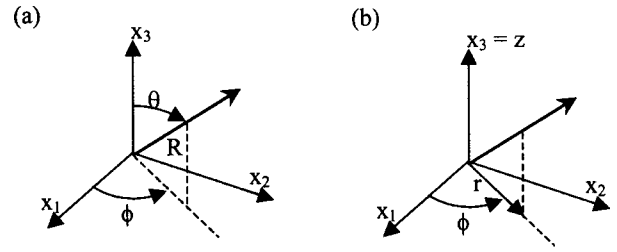


FIG. 1. Spherical (a) and cylindrical (b) reference frames.

(2) Approximate these same basic problems in a finite element analysis framework, validating the framework when possible (and investigating the effect of finite boundaries), which can then be used to study more complex and anatomically accurate geometries;
(3) Experimentally evaluate selected cases; and
(4) Evaluate source behavior as measured at the surface theoretically, computationally and experimentally and identify, if possible, simple analytical solutions that may accurately approximate more accurate but cumbersome numerical or computational solutions.

II. THEORY

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:[9–12]

$$(\lambda+\mu)\nabla\nabla\cdot\mathbf{u}+\mu\nabla^2\mathbf{u}=\rho\ddot{\mathbf{u}}, \tag{1a}$$

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. The density of the medium is denoted by ρ and displacements in the medium are denoted by $u=[u_r,u_\phi,u_z]^T$ in terms of a cylindrical coordinate system or by $u=[u_R,u_\phi,u_\theta]^T$ in terms of a spherical coordinate system. These coordinate systems are shown in Fig. 1. Normal stresses and strains in each direction are denoted by σ_{ii} and ε_{ii}, respectively, while a shear stress in the pth direction on the ith face is denoted by σ_{ip} where $i,p=r$, R, z, ϕ, or θ.

Wave motion in the *infinite* 3-dimensional viscoelastic medium consists of a superposition of dilatational and shear wave types, $\mathbf{u}=\mathbf{u}_\alpha+\mathbf{u}_\beta$, respectively. For the *semi-infinite* half-space problem an additional surface (Rayleigh) wave $\mathbf{u_R}$ will exist. The dilatational wave component can be represented by a scalar wave potential Π such that $\mathbf{u}_\alpha=\nabla\Pi$. Shear wave motion can be represented using a vector wave potential $\mathbf{H}$ such that $\mathbf{u}_\beta=\nabla\times\mathbf{H}$ where, in spherical coordinates,

$$\mathbf{H}=\hat{\mathbf{i}}_R R\psi+l\nabla\times(\hat{\mathbf{i}}_R R\chi), \tag{2}$$

where ψ and χ are two scalar potentials such that

$$\mathbf{u}_\beta=\nabla\times(\hat{\mathbf{i}}_r R\psi)+l\nabla\times\nabla\times(\hat{\mathbf{i}}_r R\chi). \tag{3}$$

Here, $\hat{\mathbf{i}}_R$ is the unit vector in the radial direction and l is a scalar factor having a dimension of length, whose role is to fit units. When defined in terms of potentials of this sort the displacements in spherical coordinates can be written as

$$u_R=\frac{\partial\Pi}{\partial R}+l\left[\frac{\partial^2(R\chi)}{\partial R^2}-R\nabla^2\chi\right], \tag{4a}$$

$$u_\phi=\frac{1}{R\sin\theta}\frac{\partial\Pi}{\partial\phi}-\frac{1}{R}\frac{\partial(R\psi)}{\partial\theta}+l\frac{1}{R\sin\theta}\frac{\partial^2(R\chi)}{\partial\phi\,\partial R}, \tag{4b}$$

$$u_\theta=\frac{1}{R}\frac{\partial\Pi}{\partial\theta}+\frac{1}{R\sin\theta}\frac{\partial R(r\psi)}{\partial\phi}+l\frac{1}{R}\frac{\partial^2(R\chi)}{\partial\theta\,\partial R}. \tag{4c}$$

Also, for the infinite medium problem the governing equation of motion can be separated into the wave equations which must be satisfied by the potential functions

$$\nabla^2 f=\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2}, \tag{5}$$

where $c=c_\alpha=\sqrt{(\lambda+2\mu)/\rho}$ for $f=\Pi$ and $c=c_\beta=\sqrt{\mu/\rho}$ for $f=\psi$ or χ. For the case of harmonic motion the problem reduces to the Helmholtz equation. This can be solved by the separation of variables; in the presence of axisymmetry about the z or x_3 axis (Fig. 1) a typical spherical wave function must be of the following form to satisfy the governing equations:

$$f_n=\sqrt{2/\pi}E_n(kR)P_n[\cos(\theta)]e^{j\omega t}, \tag{6}$$

where $k=\omega/c$. Here, $E_n(kR)$ denotes a spherical Bessel function, with $j_n(kR)$ and $y_n(kR)$ being appropriate for standing waves and $h_n^{(1)}(kR)$ and $h_n^{(2)}(kR)$ being appropriate for traveling disturbances. Also, $P_n[\cos(\theta)]$ denotes a Legendre function.

A. Infinitesimal and finite monopole in an infinite viscoelastic medium (Refs. 10 and 11)

Consider a spherical cavity of radius "a" located at the origin and subjected to internal pressure (normal stress in the radial direction) $\sigma_{RR}(R=a)=-p_0e^{j\omega t}$ in an infinite viscoelastic medium. For this case we have $\mathbf{u}_\alpha=\nabla\Pi=u_R$ where $\Pi=\Pi(R,t)$ in spherical coordinates and $\mathbf{u}_\beta=0$. The following potential describes outgoing waves:

$$\Pi=A_0h_0^{(2)}(k_\alpha R)e^{j\omega t},\quad h_0^{(2)}(k_\alpha R)=\frac{je^{-ik_\alpha R}}{k_\alpha R}. \tag{7}$$

The spherical Hankel function $h_0^{(2)}(k_\alpha R)$ of the second kind is used to denote outgoing waves to be consistent with the use of $e^{+j\omega t}$ denoting harmonic time dependence. For spherical symmetry we have

$$u_R=\frac{\partial\Pi}{\partial R}=A_0k_\alpha\left(1-\frac{j}{k_\alpha R}\right)\frac{e^{-jk_\alpha R}}{k_\alpha R}e^{j\omega t}$$
$$=A_0k_\alpha h_1^{(2)}(k_\alpha R)e^{j\omega t}. \tag{8}$$

Additionally, Hooke's law relating stress σ and strain ε adapted to the case of spherical symmetry provides

$$\sigma_{RR}=\lambda(\varepsilon_{RR}+\varepsilon_{\theta\theta}+\varepsilon_{\phi\phi})+2\mu\varepsilon_{RR}$$
$$=\lambda\left(\frac{\partial^2\Pi}{\partial R^2}+\frac{2}{R}\frac{\partial\Pi}{\partial R}\right)+2\mu\frac{\partial^2\Pi}{\partial R^2}$$
$$=-[(\lambda+2\mu)k_\alpha^2R^2h_0^{(2)}(k_\alpha R)$$
$$-4\mu k_\alpha Rh_1^{(2)}(k_\alpha R)]A_0e^{j\omega t}/R^2. \quad (9)$$

Consequently, we have that

$$A_0=\frac{p_0a^2}{(\lambda+2\mu)k_\alpha^2a^2h_0^{(2)}(k_\alpha a)-4\mu k_\alpha ah_1^{(2)}(k_\alpha a)}. \quad (10)$$

The source strength Q denotes the amplitude of volume displacement across the monopole boundary at $R=a$. This is a useful concept for extending to the infinitesimal monopole with $a\to 0$ but Q finite. For a finite monopole the source strength Q is related to u_R by $Q=4\pi a^2u_a$ where $u_a=u_r(R=a)$. For the infinitesimal monopole A_0 can be replaced by Q using Eq. (8) evaluated at $R=a$.

B. Infinitesimal monopole in a semi-infinite viscoelastic half-space (Ref. 12)

An approximate approach to estimate motion at the free surface of a half-space due to an infinitesimal monopole buried in the half-space is to take the solution based on an infinite medium (the previous section) and double its value. For 1-dimensional planar dilatational wave propagation normal to the pressure release surface this would yield the exact solution. But, for the present case of spherical wave motion incident with the surface over a range of angles from normal to oblique, such an approximation does not account for the conversion of dilatational waves to shear and Rayleigh (surface) waves for most of these angles. This fact becomes evident in the exact analysis provided below.

Ewing *et al.*[12] used the mirror image of a point compressional source in an infinite space to derive an expression for the response to a point source in a half-space. A cylindrical coordinate system is used; the free surface is perpendicular to the z-axis and resides at $z=0$ and the buried monopole source is located at $z=h$, with its mirrored counterpart at $z=-h$. The problem is axisymmetric around the z-axis so there is one dilatational and one shear potential, Π and ψ, respectively, with $\chi=0$. The amplitude of potentials for the two compressional sources acting harmonically can be written as

$$\Pi=A_0[h_0^{(2)}(k_\alpha[R-h])+h_0^{(2)}(k_\alpha[R+h])]e^{j\omega t}$$
$$=A_0\frac{j}{k_\alpha}\left[\int_0^\infty J_0(k_\alpha r)e^{-\nu_\alpha|z-h|}\frac{k\partial k}{\nu_\alpha}\right.$$
$$\left.+\int_0^\infty J_0(k_\alpha r)e^{-\nu_\alpha|z+h|}\frac{k\partial k}{\nu_\alpha}\right]e^{j\omega t}, \quad (11)$$

with $\psi=0$ where $\nu_\alpha=\sqrt{k^2-k_\alpha^2}$ and J_0 represents a zeroth order Bessel function of the first kind. For $0<z<h$ we put $|z-h|=h-z$ and obtain

$$\Pi=A_0\frac{2j}{k_\alpha}\int_0^\infty\cosh[\nu_\alpha z]e^{-\nu_\alpha h}J_0(kr)\frac{k\partial k}{\nu_\alpha}e^{j\omega t}. \quad (12)$$

Consequently, in this axisymmetric problem the cylindrical radial (r-direction) and vertical (z-direction) displacement amplitudes at the half-space surface, q_0 and w_0, respectively, are found to be

$$q_0=A_0\frac{-2j}{k_\alpha}\int_0^\infty e^{-\nu_\alpha h}J_1(kr)\frac{k^2\partial k}{\nu_\alpha}e^{j\omega t}, \quad (13a)$$

$$w_0=0. \quad (13b)$$

The stresses at the free surface determined by the potentials are

$$\sigma_{rz}(z=0)=0, \quad (14a)$$

$$\sigma_{zz}(z=0)=\frac{A_0j}{k_\alpha}2\mu\int_0^\infty\frac{(2k^2-k_\beta^2)}{\nu_\alpha}$$
$$\times e^{-\nu_\alpha h}J_0(kr)k\ dk\,e^{j\omega t}, \quad (14b)$$

violating the free surface condition. Thus, an additional system of surface stresses is required to satisfy the condition $\sigma_{zz}(z=0)=0$. If an opposing stress is applied to the assumed free surface then the surface stress boundary conditions can be set to 0 and the resulting displacements q_0 and w_0 can be obtained by summing the expressions in Eq. (13) with expressions for the displacements due to the applied surface stress. To apply a surface stress equal in amplitude but opposite in sign to Eq. (14) requires that two additional potential functions be introduced:

$$\Pi_s=A_se^{-\nu_\alpha(h-z)}J_0(kr), \quad (15a)$$

$$\psi_s=B_se^{-\nu_\beta(h-z)}J_0(kr), \quad (15b)$$

where $\nu_\beta=\sqrt{k^2-k_\beta^2}$. The horizontal and vertical displacement on the free surface due to these potentials are

$$q_{so}=-(kA_se^{-\nu_\alpha z}+k\nu_\beta B_se^{-\nu_\beta z})J_1(kr), \quad (16a)$$

$$w_{so}=(-\nu_\alpha A_se^{-\nu_\alpha z}+k^2B_se^{-\nu_\beta z})J_0(kr), \quad (16b)$$

and the stresses are,

$$\sigma_{rz}=\mu[2k\nu_\alpha A_s-k(2k^2-k_\beta^2)B_s]J_1(kr), \quad (17a)$$

$$\sigma_{zz}=\mu[(2k^2-k_\beta^2)A_s-2k^2\nu_\beta B_s]J_0(kr). \quad (17b)$$

Implementing the desired surface stress conditions one finds the unknown coefficients as

$$A_s=\frac{2k^2-k_\beta^2}{F(k)}\frac{Z}{\mu}, \quad (18a)$$

$$B_s=\frac{2\nu_\alpha}{F(k)}\frac{Z}{\mu}, \quad (18b)$$

where

$$Z=-\frac{A_0j}{k_\alpha}2\mu\frac{2k^2-k_\beta^2}{\nu_\alpha}e^{-\nu_\alpha h}k\ dk, \quad (18c)$$

$$F(k)=(2k^2-k_\beta^2)^2-4k^2\nu_\alpha\nu_\beta. \quad (18d)$$

It then can be shown that the horizontal (r-direction) and vertical (z-direction) displacement amplitudes for the buried monopole are given by the following expressions:

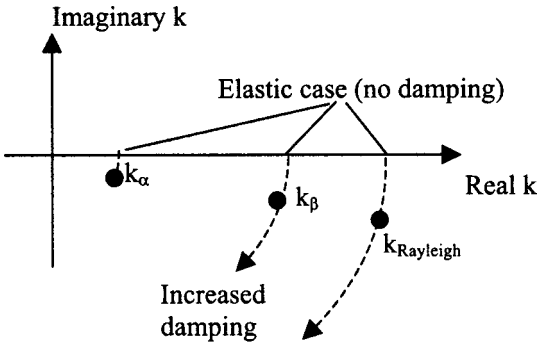


FIG. 2. Location of poles and determination of contour.

$$q_0 = \frac{4A_0 j}{k_\alpha} \int_0^\infty \frac{k_\beta^2 k^2 \nu_\beta}{F(k)} e^{-\upsilon_\alpha h} J_1(kr)\, dk\, e^{j\omega t}, \tag{19a}$$

$$w_0 = -\frac{2A_0 j}{k_\alpha} \int_0^\infty \frac{k_\beta^2 k (2k^2 - k_\beta^2)}{F(k)} e^{-\upsilon_\alpha h} J_0(kr)\, dk\, e^{j\omega t}. \tag{19b}$$

These integrals are very difficult to evaluate analytically for either the elastic or viscoelastic case. Numerical integration for the elastic case is also particularly difficult due to the existence of singularities along the path of integration on the real axis (Fig. 2). It is possible to calculate the effects of the Rayleigh pole using the Cauchy Principle Value Theorem, but this only accounts for surface (Rayleigh) wave propagation. To account for compression and shear wave propagation, which also will affect motion at the surface, it is necessary to use hyperbolic branch cuts on the complex plane, which are not easily evaluated. Alternatively, for the viscoelastic case, which results in the singularities of the integrand moved off the real axis (Fig. 2), numerical integration is possible. In the results reported later in this paper the integral above for vertical surface motion is evaluated using a fifth order Dormand–Price integration algorithm up to a large wave number ($k/k_\alpha \approx 10{,}000$ depending on the case) using a variable wave number step. If the upper boundary of integration is chosen sufficiently large, the solution asymptotically approaches a final value. In this study, convergence of the integral was presumed when the integrated value stabilized to at least 6 significant digits and k was well beyond the location of any poles. The integrals are calculated using the SIMULINK toolbox of MATLAB commercial software. Doing this it is also possible to observe the convergence of the calculation in real time during the operation. Computation time on a 600 MHz Pentium PC typically took less than a minute.

C. Finite monopole in a semi-infinite viscoelastic half-space

While equivalent source strengths for finite and infinitesimal monopoles in infinite media can be used to accurately extrapolate results between cases, this is not the case for the half-space problem since reflections will occur at the surface of the medium which will then transmit back to the finite source and be reflected further. The infinitesimal monopole case presented above in Sec. II B does not account for scattering of the reflected waves by the source. As the finite source becomes smaller and moves farther away from the surface the infinitesimal theory will become more accurate.

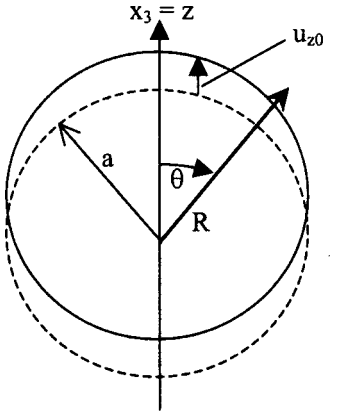


FIG. 3. Rectilinear oscillations of an embedded sphere.

Several authors have analytically addressed the case of a finite monopole buried in an elastic half-space.[13–15] In these studies a set of integral equations are derived. Unfortunately, the mathematics becomes cumbersome and, in addition to not leading to closed form solutions, implementing a numerical integral solution is also cumbersome. Additionally, in some references an asymptotic solution is obtained that is only valid for deep sources, which means when scattering from the monopole is minimal. Our interest in this paper is to review fundamental analytical solutions that are easily found and which may be more useful in design and analysis. When the analytical solution approach reaches a complexity beyond that of other approximate numerical computational solution approaches, like finite element analysis; then its utility may become dubious.

Also, the primary interest here ultimately is to use infinitesimal basic sources to model more complex distributed sources in an approximate sense. So, it is not of interest in the present article to heavily devote effort to the finite monopole problem.

D. Finite dipole in an infinite viscoelastic medium (Ref. 11)

Consider a rigid sphere of radius "a" embedded in an isotropic viscoelastic medium that is executing rectilinear motion along the z-axis given by

$$u_z = u_{z0} e^{j\omega t}. \tag{20}$$

This gives rise to dilatational and shear waves in a surrounding medium. Because of axisymmetry, $u_\phi = 0$ and the displacement vector is

$$\mathbf{u} = [u_R, 0, u_\theta]^T, \tag{21}$$

in spherical coordinates where $u_R = u_R(R,\theta,t)$ and $u_\theta = u_\theta(R,\theta,t)$. Referring to Fig. 3, the boundary conditions for "welded" contact between the sphere and surrounding medium are

$$u_R(R=a,\theta,t) = u_{z0}\cos[\theta], \tag{22a}$$

$$u_\theta(R=a,\theta,t) = -u_{z0}\sin[\theta]. \tag{22b}$$

For the case of ''lossless slip'' contact between the sphere and medium, the second equation is replaced by the following where $\sigma_{R\theta}$ denotes the shear stress in the θ direction on the sphere surface:

$$\sigma_{R\theta}(R=a,\theta,t)=0. \tag{22c}$$

For the case of axisymmetry with respect to the z axis (no dependence on ϕ), then $\psi=0$ (specifically since shear in the $\phi\theta$ plane$=0$) and the displacements u_R and u_θ become

$$u_R=\frac{\partial\Pi}{\partial R}+l\left[\frac{\partial^2(R\chi)}{\partial R^2}-R\nabla^2\chi\right], \tag{23a}$$

$$u_\theta=\frac{1}{R}\frac{\partial\Pi}{\partial\theta}+l\frac{1}{R}\frac{\partial^2(R\chi)}{\partial\theta\,\partial R}. \tag{23b}$$

The following forms of Π and χ satisfy all of the above constraints:

$$\Pi=A_1h_1^{(2)}(k_\alpha R)\cos(\theta)e^{j\omega t}, \tag{24a}$$

$$\chi=B_1h_1^{(2)}(k_\beta R)\cos(\theta)e^{j\omega t}. \tag{24b}$$

The associated displacements u_R and u_θ may then be conveniently expressed in elementary functions as

$$u_R=N_1\cos(\theta)\{[2+2jk_\alpha R-(k_\alpha R)^2]e^{-jk_\alpha R}+2N_2(-jk_\beta R-1)e^{-jk_\beta R}\}e^{j\omega t}/R^3, \tag{25a}$$

$$u_\theta=-N_1\sin(\theta)\{(-jk_\alpha R-1)e^{-jk_\alpha R}+N_2[1+jk_\beta R-(k_\beta R)^2]e^{-jk_\beta R}\}e^{j\omega t}/R^3, \tag{25b}$$

where the arbitrary constants A_1 and B_1 have been replaced with constants N_1 and N_2 with $N_1=-jA_1/k_\alpha^2$ and $N_2=-lB_1k_\alpha^2/A_1k_\beta^2$. The coefficients N_1 and N_2 can be specified based on the boundary conditions for ''welded'' contact or ''lossless slip'' contact. (Note, the above solutions hold for any ratio of the sphere size to the wavelength.) For the welded contact case, we obtain the following coefficient values:

$$N_1=-u_0a^3e^{jk_\alpha a}\frac{3+3jk_\beta a-(k_\beta a)^2}{(1+jk_\beta a)(k_\alpha a)^2+2(1+jk_\alpha a)(k_\beta a)^2-(k_\alpha a)^2(k_\beta a)^2}, \tag{26a}$$

$$N_2=e^{-j(k_\alpha-k_\beta)a}\frac{3+3jk_\alpha a-(k_\alpha a)^2}{3+3jk_\beta a-(k_\beta a)^2}. \tag{26b}$$

The force, $F=F_{z0}e^{j\omega t}$, which must be applied to the sphere to maintain steady-state sinusoidal vibrations at frequency ω may be obtained from the equations of motion and is

$$\begin{aligned}F_{z0}&=-\frac{4}{3}\pi a^3u_{z0}\omega^2-\int_0^\pi 2\pi a(a\sin\theta)\\&\quad\times(\sigma_{RR}\cos\theta-\sigma_{R\theta}\sin\theta)\partial\theta\Big|_{R=a}\\&=\tfrac{4}{3}\pi\omega^2\{-a^3\rho_0u_{z0}+N_1\rho[(1-k_\alpha a)e^{-jk_\alpha a}\\&\quad+2N_2(1-k_\beta a)e^{-jk_\beta a}]\},\end{aligned} \tag{27}$$

where ρ_0 denotes the density of the rigid sphere material. Taking the radius of the sphere ''a'' to be very small ($k_\alpha a$ and $k_\beta a$ also small) for $\rho_0=\rho$ we have

$$\begin{aligned}F_{z0}&\approx\tfrac{4}{3}\pi\omega^2\rho u_{z0}\{-a^3-9a/(k_\alpha^2+2k_\beta^2)\}\\&\approx-\frac{4}{3}\pi\omega^2\rho u_{z0}\frac{9a}{k_\alpha^2+2k_\beta^2},\end{aligned} \tag{28}$$

where $N_1=-3u_{z0}a/(k_\alpha^2+2k_\beta^2)$ and $N_2=1$. Likewise, under the same conditions for u_R and u_θ we have

$$u_R=-u_{z0}a\frac{3}{k_\alpha^2+2k_\beta^2}\cos(\theta)\{[2+2jk_\alpha R-(k_\alpha R)^2]e^{-jk_\alpha R}+2(-jk_\beta R-1)e^{-jk_\beta R}\}e^{j\omega t}/R^3, \tag{29a}$$

$$u_\theta=u_{z0}a\frac{3}{k_\alpha^2+2k_\beta^2}\sin(\theta)\{(-jk_\alpha R-1)e^{-jk_\alpha R}+[1+jk_\beta R-(k_\beta R)^2]e^{-jk_\beta R}\}e^{j\omega t}/R^3. \tag{29b}$$

For the case of $\theta=0$ we have for $u_R=u_z$,

$$u_z=-u_{z0}a\frac{3}{k_\alpha^2+2k_\beta^2}\{[2+2jk_\alpha z-(k_\alpha z)^2]e^{-jk_\alpha z}+2(-jk_\beta z-1)e^{-jk_\beta z}\}e^{j\omega t}/z^3. \tag{30}$$

The following frequency response function can be calculated and compared with results for the infinitesimal dipole given in the next section:

$$\frac{u_z}{F_{z0}}=\frac{1}{4\pi\omega^2\rho}\{[2+2jk_\alpha z-(k_\alpha z)^2]e^{-jk_\alpha z}-2(jk_\beta z+1)e^{-jk_\beta z}\}e^{j\omega t}/z^3. \tag{31}$$

Note that these formulations are in the frequency domain because the theoretical analysis is easier when harmonic excitation and response is assumed. Also this was preferred because it is easier to incorporate damping by the use of complex Lame parameters. If the time response is required, it is possible to use the frequency response functions together with the Fourier transformed form of the forcing function. The inverse Fourier transform of the product of the frequency response function and forcing function in the frequency domain gives the time response of the dipole.

E. Infinitesimal dipole (point force) in an infinite viscoelastic medium (Ref. 11)

The approach in the previous sections makes use of the homogeneous wave equation,

$$c^2\nabla^2 f=\frac{\partial^2 f}{\partial t^2}, \tag{32}$$

which describes free waves, while a physical cause of the phenomenon lies in the boundary and initial conditions. Unlike this, the relation

$$c\nabla^2 f-\frac{\partial^2 f}{\partial t^2}=-f_0, \tag{33}$$

known as a nonhomogeneous wave equation, involves directly the forcing term $f_0\neq 0$, which depends on time and space. For the dynamic equations of displacements for isotropic linear elastic media, we have $L_{il}u_l=-f_i$, where the operator is defined as

$$L_{il}=-\delta_{il}\rho\frac{\partial^2}{\partial t^2}+(\lambda+\mu)_{,il}+\mu\nabla^2\partial_{il}. \tag{34}$$

Here, δ_{il} denotes the Kronecker delta with $\delta_{il}=1$ if $i=l$ and $=0$ otherwise. Forcing at a point $\mathbf{x}$ with a unit impulse can be written more explicitly as

$$f_i=\delta_{il}\delta(\mathbf{x})\,\delta(t). \tag{35}$$

Take the point force to be located at the origin of a reference frame and directed along a coordinate x_p. The displacement along x_l occurring at a point $\mathbf{x}=[x_1,x_2,x_3]^T$ due to the unit impulse directed along the x_p-axis will be labeled as $G_{lp}(\mathbf{x},t)$. After these modifications the dynamic displacement equation becomes

$$L_{il}G_{lp}(\mathbf{x},t)=-\delta_{ip}\delta(\mathbf{x})\,\delta(t). \tag{36}$$

In order to obtain the harmonic response a 4-dimensional Fourier transform is applied. The 4-dimensional Fourier transform $f(\mathbf{k},\omega)=f(k_1,k_2,k_3,\omega)$ of a function $f(\mathbf{x},t)$ is defined by

$$f(\mathbf{k},\omega)=\int_{V_4} f(\mathbf{x},t)e^{j(\mathbf{kx}+\omega t)}\,dV_4, \tag{37}$$

where $dV_4=d\mathbf{x}\,dt$. After this operation and via algebraic manipulation we have

$$G_{ip}(\mathbf{k},\omega)=\frac{1}{\mu k^2-\rho\omega^2}\left[\delta_{ip}-\frac{(\lambda+\mu)k_ik_p}{(\lambda+2\mu)k^2-\rho\omega^2}\right]. \tag{38}$$

The inverse Fourier transform for this equation gives the displacement

$$G_{ip}(\mathbf{x},t)=\frac{1}{16\pi^4}\int_{W_4} G_{ip}(\mathbf{k},\omega)e^{-j(\mathbf{kx}+\omega t)}\,dW_4, \tag{39}$$

where $dW_4=dk_1\,dk_2\,dk_3\,d\omega=d\mathbf{k}\,d\omega$. The following limited inverse transformation can be used to obtain the frequency response only,

$$G_{ip}(\mathbf{x},\omega)=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty} G_{ip}(\mathbf{k},\omega)e^{-j\mathbf{kx}}\,d\mathbf{k}. \tag{40}$$

Using this inverse transformation the frequency response for the displacement along x_i due to a forcing along x_p is given as

$$G_{ip}(\mathbf{x},\omega)=\frac{1}{4\pi\rho\omega^2}\left[k_\beta^2\frac{e^{-jk_\beta R}}{R}\delta_{ip}-\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_p}\left(\frac{e^{-jk_\alpha R}-e^{-jk_\beta R}}{R}\right)\right], \tag{41}$$

where $R=|\mathbf{x}-\mathbf{x}'|$ with $\mathbf{x}'$ indicating the point of application of the source. For the "vertical" response (z-direction) to a vertical force (z-direction) where the force is at the origin and the response is at $R=\sqrt{r^2+z^2}$ where $z=h=x_3$ and $r=\sqrt{x_1^2+x_2^2}$, we have that

$$\frac{z_{rh}}{F_0}=G_{zz}(\omega)=\frac{1}{4\pi\rho\omega^2}\left[k_\beta^2\frac{e^{-jk_\beta R}}{R}-\frac{1}{R^5}\left\{e^{-j[k_\alpha+k_\beta]R}\left[\begin{array}{l}-e^{jk_\beta R}(r^2\{1+k_\alpha^2h^2+jk_\alpha R\}+h^2\{-2+k_\alpha^2h^2-2jk_\alpha R\})\\+e^{jk_\alpha R}(r^2\{1+k_\beta^2h^2+jk_\beta R\}+h^2\{-2+k_\beta^2h^2-2jk_\beta R\})\end{array}\right]\right\}\right]. \tag{42}$$

Note that for $r=0$ and $R=z$ we have

$$\frac{u_z}{F}=G_{zz}(\omega)=\frac{1}{4\pi\omega^2\rho}\{[2+2jk_\alpha z-(k_\alpha z)^2]e^{-jk_\alpha z}-2(jk_\beta z+1)e^{-jk_\beta z}\}e^{j\omega t}/z^3. \tag{43}$$

This is in exact agreement with Eq. (31) of the previous section, which was derived based on a finite dipole formulation using the homogeneous equations with the effect of forcing taken into account in the boundary conditions and taking the limit as the radius of the dipole "a" became very small. The approach here has used the nonhomogeneous approach and arrived at the same result.

F. Infinitesimal and finite dipole in a semi-infinite viscoelastic half-space

Analogous to the case of the monopole, an approximate approach to estimate motion at the free surface of a half-space due to an infinitesimal dipole buried in the half-space would be to take the solution based on an infinite medium (previous section) and double its value. The dipole source produces both dilatational and shear wave motion; however, in addition to conversions between dilatational and shear wave types, which will still occur at the surface, Rayleigh (surface) waves are not accounted for in such an approximation.

Pekeris[16] derived integral expressions for wave motion due to the buried vertical infinitesimal dipole in a half-space.

A cylindrical coordinate system is used; the free surface is perpendicular to the z-axis and resides at $z=0$ and the buried vertical dipole source is located at $z=h$. The problem is axisymmetric around the z-axis so there is one dilatational and one shear potential, Π and ψ, respectively, with $\chi=0$. In order to satisfy the zero stress condition at the free surface and to have continuity within the half-space, horizontal and vertical motion, q_0 and w_0, respectively, at the surface due to a harmonic point force $F(t)=F_0e^{j\omega t}$ are given by the following expressions:

$$\frac{q_0}{F_0}=\frac{-1}{2\pi\mu}\int_0^\infty \frac{k^2}{F(k)}[-2\sqrt{k^2-k_\alpha^2}\sqrt{k^2-k_\beta^2}e^{-h\sqrt{k^2-k_\alpha^2}} + (2k^2-k_\beta^2)e^{-h\sqrt{k^2-k_\beta^2}}]J_1(kr)dk\,e^{j\omega t}, \tag{44a}$$

$$\frac{w_0}{F_0}=\frac{1}{2\pi\mu}\int_0^\infty \frac{k\sqrt{k^2-k_\alpha^2}}{F(k)}[-(2k^2-k_\beta^2)e^{-h\sqrt{k^2-k_\alpha^2}} + 2k^2e^{-h\sqrt{k^2-k_\beta^2}}]J_0(kr)dk\,e^{j\omega t}. \tag{44b}$$

Chao[17] derived similar integral expressions for the response at the surface due to a horizontal buried point force in a half-space. As in the case of the infinitesimal monopole expressions, these integrals are difficult to evaluate analytically for either the elastic or viscoelastic case. Numerical integration for the elastic case is also particularly difficult due to the existence of singularities along the path of integration on the real axis (Fig. 2). Achenbach[18] has derived the closed form solution for the Rayleigh wave component, but this is not expected to be dominant except possibly in the near vicinity of a shallow depth source. To account for compression and shear wave propagation, which also will affect motion at the surface, it is necessary to use hyperbolic branch cuts on the complex plane, which are not easily evaluated. Alternatively, for the viscoelastic case, which results in the singularities of the integrand moved off of the real axis (Fig. 2), numerical integration is possible. In the results reported later in this paper the integral above for vertical surface motion is evaluated using the same technique described for the infinitesimal monopole in a halfspace in a previous section (Sec. II B).

With respect to the problem of a finite dipole in semi-infinite viscoelastic or elastic half-space, the authors have not found an explicit treatment in the literature.

G. Multilayered half-space problems and infinitesimal buried sources

A number of studies in the seismology literature have addressed the case of buried monopole or dipole sources in multilayer half-spaces, where the layering is parallel to the half-space surface and each layer has similar but distinct isotropic and homogeneous material properties. Numerical methods are necessary. As an example, in a recent study, Pak *et al.*[19] has essentially extended the work of Perkeris and others to the case of a buried point force in a multilayered half-space, by enforcing continuity across each layer and the zero stress condition at the surface. (This reference is also an excellent bibliographical source citing 33 studies in this area.) For the multilayered problem, instead of one complex integral expression for vertical or horizontal motion at a point, there are multiple integral expressions with singularities that must be solved. The same issues of difficulty that afflict Perkeris's solution are also present in the multilayer problem. Pak *et al.* have developed an alternative more efficient means of solving these integrals, employing the method of asymptotic decomposition, which still requires numerical contour integration.

An application of these techniques to the layered biological soft tissue problem is left for future work. Of more interest in the present study is to determine whether simple closed form analytical approximations may be sufficiently accurate to be used as initial design tools in place of numerical integral or finite element approaches.

III. COMPUTATIONAL FINITE ELEMENT ANALYSIS

Finite element analysis with ANSYS® software Version 6.1 was used to approximately simulate the cases of a *finite monopole, vertical finite dipole* and *vertical infinitesimal dipole* buried in a viscoelastic medium with finite boundaries. Boundaries of the model were taken to approximately match those of the experiment described in the next section. The axisymmetric nature of all of these problems was utilized in constructing finite element models with, consequently, a greatly reduced number of nodes and elements. Eight node planar elements (plane82) with four corner and four side nodes or four node planar elements without side nodes (plane42) were used. Because of the simple geometry, direct mesh generation was employed with a nominal nodal resolution of 1 mm with minor adjustments in the vicinity of the finite sources. The 8 node elements seemed to provide superior performance only in the case of the point dipole source, given identical node spacing (8 node elements twice the width and height of 4 node elements). The depth of the axisymmetric model was 135.7 mm and its radial width was 172 mm. Material property values used in each of the example case studies are provided at the beginning of Sec. V in terms of the Lame constants, λ_1, λ_2, μ_1, μ_2, and density, ρ. For application in finite element analysis using the indicated solid elements, Young's modulus E, Poisson's ratio ν, a linear viscous damping coefficient ζ, and density ρ are needed. The viscoelastic parameters are related to the Lame constants by the following expressions, where approximations are also provided based on the following soft tissue characteristics in the frequency range of interest:[9,20–21] $\lambda_1 \gg \mu_1$, $\lambda_1 \gg \omega\mu_2$, and $\lambda_2=0$,

$$E(1+j\omega\zeta)=\frac{(\mu_1+j\omega\mu_2)(3\lambda_1+2\mu_1+2j\omega\mu_2)}{\lambda_1+\mu_1+j\omega\mu_2} \approx 3\mu_1+j3\omega\mu_2, \tag{45a}$$

$$\nu=\frac{1}{2}\frac{\lambda_1}{\lambda_1+\mu_1+j\omega\mu_2}\approx\frac{1}{2}\frac{\lambda_1}{\lambda_1+\mu_1}. \tag{45b}$$

Note that results can be very sensitive to Poisson's ratio and a value of $\frac{1}{2}$ is not permissible. A typical model is shown in Fig. 4. Fixed boundary conditions were applied to the bottom and outer surface of the cylindrical region, with the top surface being free and radial motion along the axisymmetric

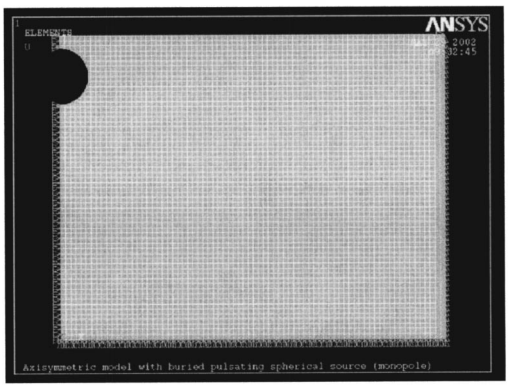


FIG. 4. Typical finite element model of a finite dimension source.

axis being constrained. All results shown in Sec. V are obtained using the harmonic analysis solution routine in ANSYS. The finite monopole is implemented by applying a uniform radial displacement amplitude to the nodes that comprise the monopole surface. For the vertical finite dipole, these same nodes are subjected to a uniform vertical displacement amplitude. For the infinitesimal dipole, a force is applied vertically to one of the nodes (and the cavity shown in Fig. 4 is not present).

IV. EXPERIMENT—FINITE DIPOLE IN A VISCOELASTIC HALF-SPACE

An experimental study was conducted for the case of the *finite dipole* buried in a viscoelastic medium with finite boundaries. Dimensions of the model (depth and radius) approximately matched those of the finite element model described in the previous section. A diagram of the experimental apparatus is shown in Fig. 5. A steel sphere is mounted to a mechanical shaker (Bruel & Kjaer #4808) via a steel stinger that has a smooth, lubricated surface and via an impedance head (PCB Model #288B02). The stinger and sphere are mounted in a container over the shaker via a rubber diaphragm at the base of the container. A phantom gel material in liquid form before it cools is poured into the container until it covers the sphere to the desired depth. The phantom material is 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 this soft tissue phantom is based on prior investigations.[9] The gel container is mounted on a vibration isolated optics bench. The shaker is mounted on a separate support structure below the bench, with the stinger coming up through a clearance hole drilled in the optics bench below the annular rubber diaphragm at the base of the gel-filled container.

A drive signal for the shaker is generated in an Agilent #35670 dynamic signal analyzer and amplified using a Bruel & Kjaer Model #2712 amplifier. A burst chirp input is applied that sweeps from 0 to 400 Hz in 2 seconds (The response of the shaker rolls off at frequencies less than 40 Hz attenuating its output.) The surface response of the phantom material is measured using a laser Doppler vibrometer (LDV; Polytec Model # CLV 800). Very small pieces of 3M retroreflective tape are mounted on the semi-translucent phantom material to aid in LDV measurement. Additionally, the vertical motion of the rigid sphere embedded in the phantom material just below its surface can be measured using the LDV because of the translucent nature of the phantom material.

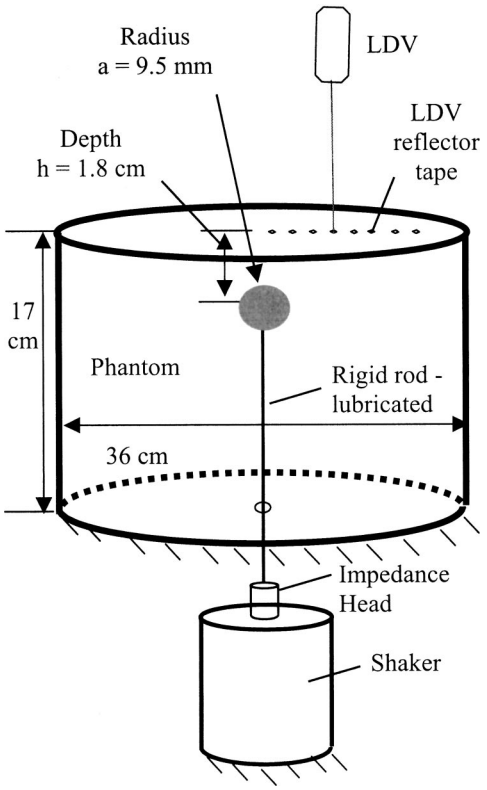


FIG. 5. Schematic of experimental finite dipole study.

The analyzer records the impedance head acceleration on Channel 1 and the LDV measurement on Channel 2. Time averaging of 16 records is used to improve the signal-to-noise. Spectral analysis is performed in Matlab® on a PC. The impedance head acceleration serves as a common normalizing signal for all measurements made with the 2 channel analyzer. In Matlab, the frequency response of the system is calculated, using the rigid sphere velocity as the input and the phantom vertical surface velocity as the output.

V. RESULTS AND DISCUSSION

For all analytical and computational studies, the following material property values were used: $\lambda_1 = 2.6$ GPa, $\lambda_2 = 0$, $\mu_1 = 2000$ Pa, $\mu_2 = 6$ Pa s, and $\rho = 1000$ kg/m^3. The property values nominally match those of the gel phantom material employed in the experimental study for the finite dipole, based on measurements of surface wave speed and attenuation.[9] Additionally, these material properties are comparable to those of biological soft tissue.[9,21–22]

A. Buried finite and infinitesimal monopole studies

Results of these studies can be found in Figs. 6, 7, and 8, which graph the surface vertical velocity response as a function of frequency and position for a smaller monopole at shallow (Fig. 6) and deeper depths (Fig. 8) and for a larger monopole at the more shallow depth (Fig. 7). For these cases, the infinite medium analytical solution for the infini-

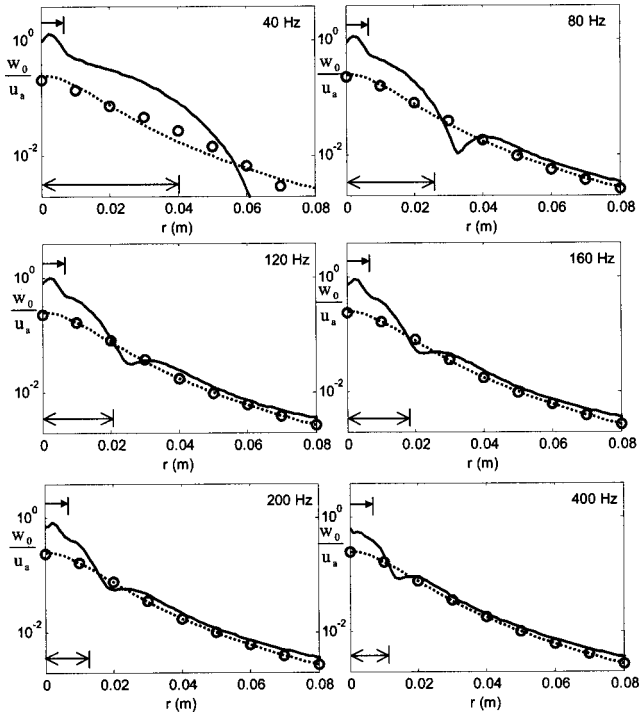


FIG. 6. Monopole results. Small (radius=6.4012 mm) finite monopole at shallow ($h=18.105$ mm) depth. Surface vertical displacement referenced to monopole surface radial displacement. Key: —— FEA, — — —, analytical solution of Sec. II A doubled; ○, numerical integral solution of Sec. II B. An arrow in the upper left corner of each figure indicates extent of source beneath surface. An arrow in the lower left corner indicates the surface (Rayleigh) wave length.

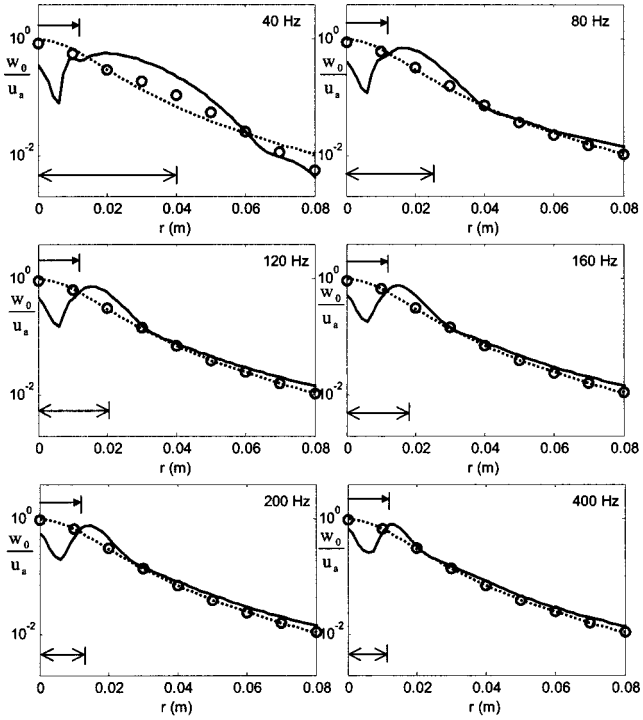


FIG. 7. Monopole results. Large (radius=12.8024 mm) finite monopole at shallow ($h=18.105$ mm) depth. Surface vertical displacement referenced to monopole surface radial displacement. Key: —— FEA, — — —, analytical solution of Sec. II A doubled; ○, numerical integral solution of Sec. II B. An arrow in the upper left corner of each figure indicates the extent of source beneath the surface. An arrow in the lower left corner indicates the surface (Rayleigh) wave length.

tesimal monopole (Sec. II A) is doubled as an approximation that accounts only for outgoing dilatational waves. Numerical integration per section II B is used to solve for the response in a halfspace to the equivalent infinitesimal monopole. Finite element analysis results are also provided for the equivalent finite monopole, but in a bounded medium with a free surface.

First, in comparing theoretical solutions for a point source based on the closed-form analytical ''doubled infinite medium approximation'' and the numerically evaluated ''exact half-space theory,'' it is seen that agreement is good at 80, 120, 160, 200 and 400 Hz in Figs. 6, 7, and 8. At 40 Hz, there is some discrepancy, especially in Figs. 6 and 7. Recall that the approximate approach only accounts for spherically radiating compression waves from the source. The exact half-space solution accounts for conversion of these waves to shear and surface waves. This comparison suggests that such a mode conversion phenomena may only be of significance at very low audible frequencies and for sources close to the surface.

In all three cases, finite element results approach theoretical predictions as the distance from the source and frequency increase. Directly over the source, particularly for the two shallow source cases (Figs. 6 and 7), there are larger discrepancies whose radial extent appears to scale with source size. This is highlighted by projecting the source radius onto each figure in the upper left corner. For all three cases there is also a more extended discrepancy near the source that seems to scale with frequency. This is highlighted by marking the wavelength λ_s of the surface wave at each frequency on the appropriate figure in the lower left corner. Because this discrepancy scales with λ_s, as the frequency increases agreement between FE results and theoretical predictions improves overall.

There are three possible reasons for differences between theory and FE results: (1) scattering from the finite source is not accounted for in the theoretical solution; (2) for compression waves, the finite boundaries do not provide a ''semi-infinite half-space'' condition, and (3) insufficient FE mesh resolution. With regard to reason #3, it was found that further refinement of the mesh had only a minor effect that was most evident for the higher frequencies and could not possibly account for the observed discrepancies, which decreased with increasing frequency. It is suspected that both of the first two reasons may contribute to the differences, with scattering more linked to the source ''footprint'' and finite boundary effects more linked to frequency-scaled discrepancies. Note, compression wavelengths nominally are equal to $\sqrt{\lambda_1/\rho/f}$ where f is the cyclic frequency in Hertz.

At 40 and 400 Hz, the compression wavelength in the phantom is approximately 40 and 4 meters, respectively. In addition, given the shear viscosity of the material, plane compression waves propagating the radial length of the phantom are only attenuated by less than 1%, so multiple reflections are possible. Treating the cylindrical phantom region as an acoustic fluid (neglecting shear elasticity) with fixed radial boundaries, a fixed bottom and a free top surface,

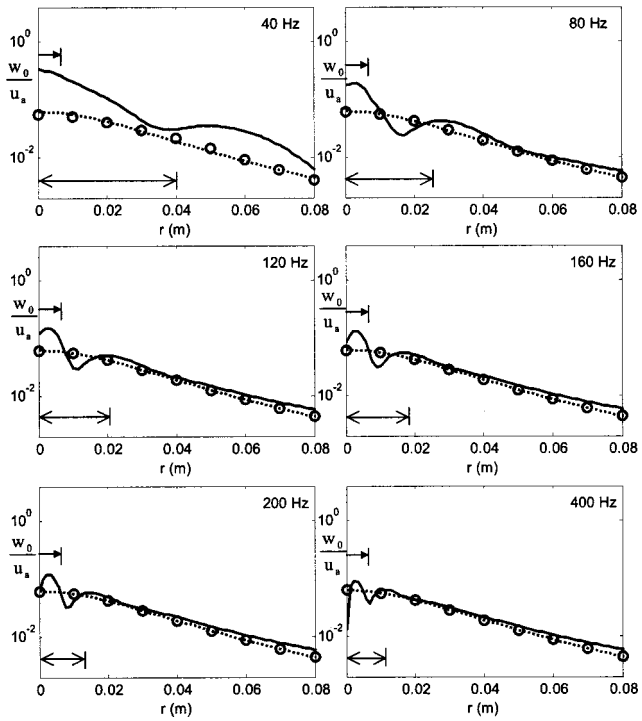


FIG. 8. Monopole results. Small (radius=6.4012 mm) finite monopole at deep ($h=36.211$ mm) depth. Surface vertical displacement referenced to monopole surface radial displacement. Key: ——, FEA; — — —, analytical solution of Sec. II A doubled; ○, numerical integral solution of Sec. II B. An arrow in upper left corner of each figure indicates the extent of source beneath the surface. An arrow in the lower left corner indicates the surface (Rayleigh) wave length.

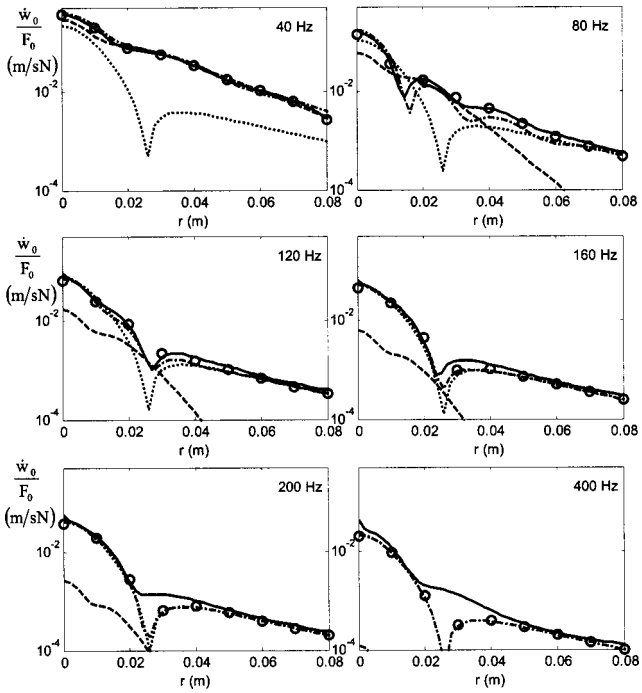


FIG. 9. Infinitesimal dipole results. Shallow ($h=18.105$ mm) depth. Surface vertical velocity referenced to dipole force. Key: ——, FEA; – – –, analytical solution of Sec. II E doubled (- - -, dilatational wave component; – – –, shear wave component); ○, numerical integral solution of Sec. II F.

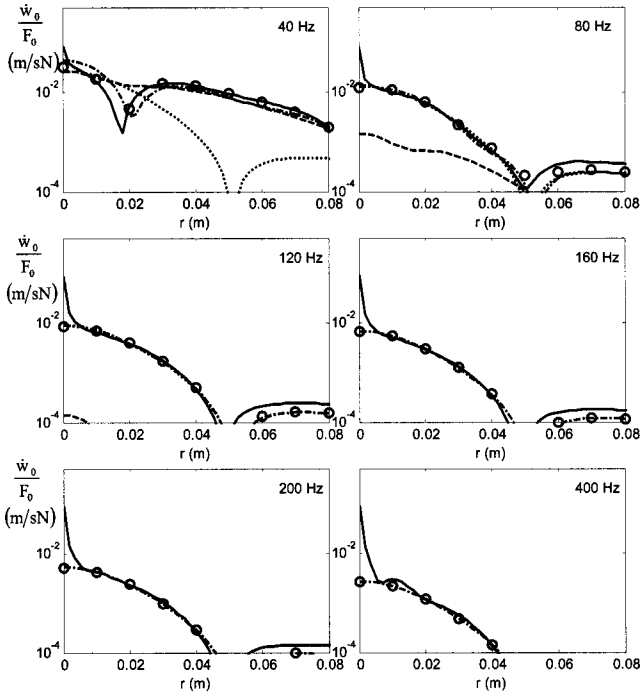


FIG. 10. Infinitesimal dipole results. Deep ($h=36.211$ mm) depth. Surface vertical velocity referenced to dipole force. Key: ——, FEA; – - –, analytical solution of Sec. II E doubled (- - -, dilatational wave component; – – –, shear wave component); ○, numerical integral solution of Sec. II F.

the lowest axisymmetric resonant mode has a natural frequency of about 3 kHz. In summary, it is concluded that an accurate theoretical analysis of surface motion on the finite dimension problem requires consideration of the boundaries,

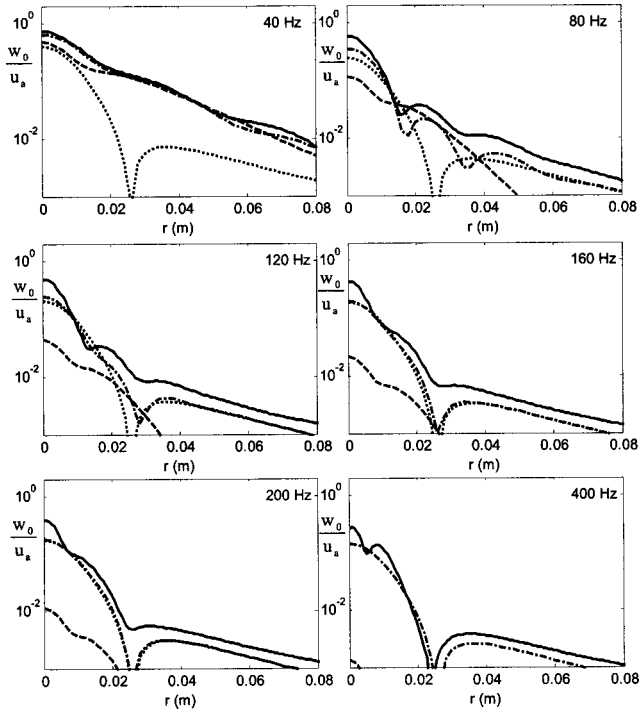


FIG. 11. Finite dipole results. Small (radius=6.4012 mm) finite dipole at shallow ($h=18.105$ mm) depth. Surface vertical displacement referenced to dipole surface vertical displacement. Key: ——, FEA; – - –, analytical solution of Sec. II D doubled (- - -, dilatational wave component; – – –, shear wave component).

especially as one approaches the low end of the audible frequencies.

B. Buried infinitesimal dipole studies

Results of these studies can be found in Figs. 9 and 10, which graph the surface vertical velocity response as a function of frequency and position for vertical point forces at shallow (Fig. 9) and deeper (Fig. 10) depths. For this case, the infinite medium analytical solution for the infinitesimal dipole (Sec. II E) is doubled as an approximation that accounts only for outgoing dilatational and shear waves. Numerical integration is used to solve for the response in a half-space according to Sec. II F; finite element analysis results are also provided.

First, in comparing theoretical solutions for a point source based on the closed-form analytical ''doubled infinite medium approximation'' and the numerically evaluated ''exact half-space theory,'' it is seen that agreement is very good for both cases at most frequencies and radial positions. This suggests that mode conversion and surface wave propagation is not significant in this situation relative to outgoing shear and compression waves. This is an important observation, in that the closed-form analytical solution can be used much more easily in iterative design and analysis studies. The closed-form solution also easily provides a quantitative comparison of the relative strengths of compression and shear wave motion measured at the surface as a function of frequency and distance from the source. Depending on the distance from the source, in the low audible frequency range both wave-types can be significant. Eventually the compression waves dominate as frequency and distance increase due to the higher rate of attenuation per wavelength of the shear waves. Note, there are some fluctuations, particularly in the compression wave strength as one moves on the surface farther away from the source due to the fact that the dipole directivity pattern is not spherically symmetric.

Finite element results do not exactly match theory, but overall are in *better agreement* with theory than was the case for the monopole, particularly at the lower frequencies. In the point dipole (point force) case scattering from the source should not be present theoretically or in FE simulation. However, it appears that there may be some localized anomaly directly over the source in the FE simulation, particularly observable in Fig. 10; this may be associated with trying to represent an infinitesimal source in a finite mesh. Additionally, the finite boundaries differentiate the FE simulation from the theoretical studies. The limitations of the boundaries with respect to compression wave propagation were discussed in the previous section. On the other hand, shear wavelengths nominally are equal to real $\lfloor \sqrt{\mu_1 + j\omega\mu_2/\rho} \rfloor / f$. At 40 Hz and 400 Hz, the shear wavelength in the phantom is approximately 4.17 cm and 1.3 cm, respectively. In addition, given the shear viscosity of the phantom material, plane shear waves propagating the radial length of the phantom are attenuated by more than 99%. It is concluded that, while the finite boundaries of the FE model negate the half-space assumption with regard to compression wave propagation, the assumption may still be accurate for shear and surface wave propagation. Since the dipole source produces both shear and compression waves, where as the monopole only produces compression waves, it is not surprising that better agreement between FE simulation and halfspace theory is achieved for the dipole, particularly at the lower frequencies.

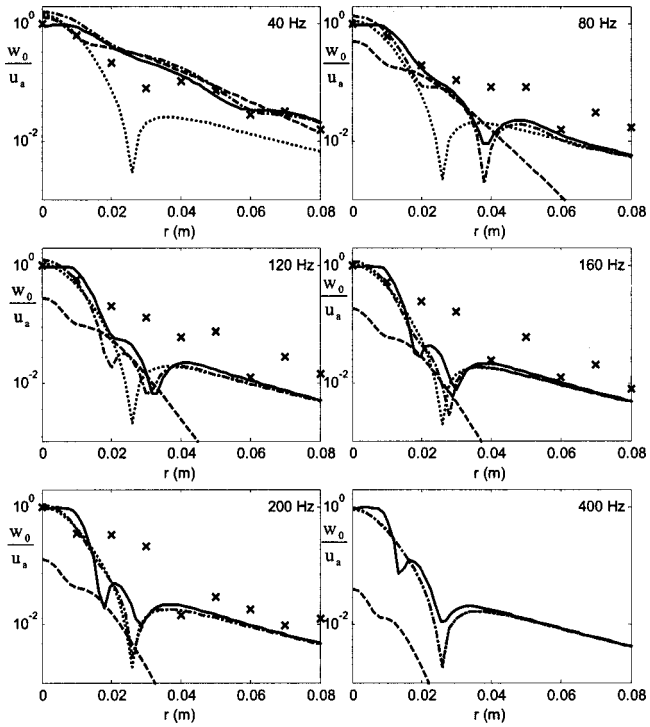


FIG. 12. Finite dipole results. Large (radius=12.8024 mm) finite dipole at shallow (h=18.105 mm) depth. Surface vertical displacement referenced to dipole surface vertical displacement. Key: ——, FEA; – – – analytical solution of Sec. II D doubled (- - -, dilatational wave component; – – –, shear wave component); ×, experimental measurement.

C. Buried finite dipole studies

Results of these studies can be found in Figs. 11, 12, and 13, which graph the surface vertical velocity response as a function of frequency and position for a smaller dipole at shallow (Fig. 11) and deeper (Fig. 13) depths and for a larger dipole (Fig. 12) at the more shallow depth. For this case, the infinite medium analytical finite dipole solution (Sec. II D) is doubled as an approximation that accounts for outgoing dilatational and shear waves. Finite element analysis (FEA) results are provided for all three cases and experimental measurements are also presented for one of the cases.

Especially at lower frequencies and near the source, agreement between the approximate theoretical half-space solution and FE simulations is better than for the monopole studies, but not as good as for the infinitesimal dipole studies. This may be primarily due to scattering from the finite dipole source, now present in the FE simulation but not accounted for in the theoretical solution.

Experimental results qualitatively follow the same trends with frequency and distance from the source as theory and FE simulation. Several experimental limitations have been identified that probably account for large discrepancies between the presented FE simulation and experimental measurements at certain frequencies and distances. It is believed

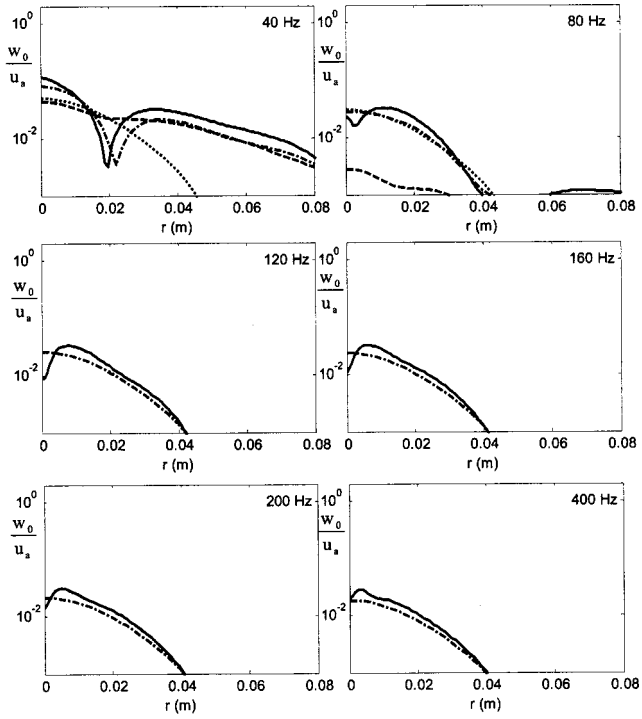


FIG. 13. Finite dipole results. Small (radius=6.4012 mm) finite dipole at deep (h=36.211 mm) depth. Surface vertical displacement referenced to dipole surface vertical displacement. Key: ——, FEA; – - –, analytical solution of Sec. II D doubled (- - -, dilatational wave component; – – –, shear wave component).

that the primary culprit may be vibratory motion of the LDV sensing head, itself, which was mounted to an overhead xy stage supported on the optics bench. Particularly within the range of 60 to 250 Hz, unity coherence was observed between an accelerometer mounted at the LDV sensor head and the impedance head mounted on the shaker during the chirp protocol. Additionally, the measured LDV sensor head vibration was sufficient in amplitude to contaminate measurements at the $r \geqslant 3$ cm positions. The path of vibration would be from the sphere, stinger and diaphragm, in contact with the gel and phantom container base, which rested on the optics table, through the table and into the xy stage support structure to the LDV. The reactive response of the optics table, itself, may have also contributed to the discrepancies. In hindsight the LDV should have been separately isolated from both the phantom and shaker instead of just from the shaker. Also, the phantom should have been placed on a more rigid and/or damped surface. Other FE simulations (not shown) confirm that minor differences between the FE and experimental phantom geometries and the fact that the actual radial phantom boundary was nonrigid (3/16 inch thick plastic) had a negligible effect. As noted earlier, attempts were made to minimize coupling between the stinger, diaphragm and gel via a lubricant. But, even if a ''welded'' contact condition existed, FE simulations suggest the effect of stinger and diaphragm motion on surface response would be negligible relative to the effect of sphere motion. Also, physical inspection of the disassembled experimental setup supported the assumption of a ''welded'' contact between the sphere and gel, as opposed to a ''lossless'' slip condition.

VI. CONCLUSION

The response at the surface of an isotropic viscoelastic medium to buried finite and infinitesimal monopole and dipole sources within the low audible frequency range (40–400 Hz) has been studied. Properties of the viscoelastic medium were chosen to correspond to soft biological tissue so that the reported results have relevance to the medical diagnostic methodology of recording subsurface biological sounds using noninvasive skin surface-based sensors. Analytical and numerical integral solutions that account for compression, shear and surface wave response to the buried sources were formulated and compared with numerical finite element simulations and for one case, an experimental study on a finite dimension phantom model. Some key observations based on these studies are the following.

(1) At low audible frequencies, compression and shear wave propagation from point sources can both be significant. In general, as frequency increases beyond a few hundred Hz, the shear wave component becomes less significant relative to the compression wave component.
(2) Simple closed-form analytical approximations based on an infinite medium model agree well with numerically obtained ''exact'' half-space solutions for the frequency range and material of interest in this study.
(3) At the frequencies of interest here, a half-space approximation for a finite viscoelastic medium with bounds comparable to human subjects is reasonable for shear wave propagation but is inaccurate for compression waves. For compression waves, finite boundaries may result in significant reflection.

ACKNOWLEDGMENTS

The financial support of the National Institutes of Health (NCRR Grant No. 14250) and the Whitaker Foundation (BME Grant No. RG 01-0198) is acknowledged.

[1] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, ''Respiratory sounds—Advances beyond the stethoscope,'' Am. J. Respir. Crit. Care Med. **156**, 974–978 (1997).
[2] H. A. Mansy, R. Balk, T. J. Royston, and R. H. Sandler, ''Pneumothorax detection using computerized analysis of breath sounds,'' Med. Biol. Eng. Comput. **40**, 526–532 (2002).
[3] J. J. Fredberg and S. K. Holford, ''Discrete lung sounds: crackles (rales) as stress relaxation quadrupoles,'' J. Acoust. Soc. Am. **73**, 1036–1046 (1983).
[4] H. A. Mansy and R. H. Sandler, ''Enhancement of bowel sound signals in sedated rats by heart sound removal using adaptive filtering,'' IEEE Eng. Med. Biol. Mag. **16**, 105–117 (1997).
[5] X. Kong, H. A. Mansy, and R. H. Sandler, ''Multi resolution analysis of gastrointestinal sounds for small bowel obstruction identification,'' Int. J. Comput. Appl. **8**, 7–12 (2001).
[6] N. L. Owsley and A. J. Hull, ''Beamformed nearfield imaging of a simulated coronary artery containing a stenosis,'' IEEE Trans. Med. Imaging **17**, 900–908 (1998).
[7] C. E. Chassaing, S. D. Stearns, M. H. van Horn, and C. A. Ryden, ''Noninvasive turbulent blood flow imaging system,'' United States Patent No. 6,278,890 issued 21 August 2001.
[8] J. J. Fredberg, ''Pseudo-sound generation at athero-sclerotic constrictions in arteries,'' Bull. Math. Biol. **36**, 143–155 (1974).
[9] 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).

[10] K. F. Graff, *Wave Motion in Elastic Solids* (Ohio State University Press, Columbus, OH, 1975).
[11] A. I. Beltzer, *Acoustics of Solids* (Springer-Verlag, Berlin, 1988).
[12] W. M. Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).
[13] V. A. Babeshko, M. G. Seleznev, T. N. Selezneva, and V. P. Sokolov, ''On a method of studying steady-state oscillations of an elastic half-space containing a cavity,'' PMM USSR **47**, 88–93 (1984).
[14] T. G. Rumyantseva, T. N. Selezneva, and M. G. Seleznev, ''The three-dimensional problem of steady oscillations of an elastic half-space with a spherical cavity,'' PMM-J. Appl. Math. Mec. **50**, 497–501 (1986).
[15] A. Y. Sinyaev and A. L. Kal'ts, ''Stress waves in a half-space as the effect of an internal source,'' Sov. Min. Sci. **18**, 275–284 (1982).
[16] C. L. Pekeris, ''The seismic surface pulse,'' Proc. Natl. Acad. Sci. U.S.A. **41**, 469–480 (1955).
[17] C. C. Chao, ''Dynamical response of an elastic half-space to tangential surface loadings,'' J. Appl. Mech. **27**, 559–567 (1960).
[18] J. D. Achenbach, ''Calculation of surface wave motions due to a subsurface point force: An application of elastodynamic reciprocity,'' J. Acoust. Soc. Am. **107**, 1892–1897 (2000).
[19] R. Y. S. Pak and B. B. Guzina, ''Three-dimensional Green's functions for a multilayered half-space in displacement potentials,'' J. Eng. Mech. Div., Am. Soc. Civ. Eng. **128**, 449–461 (2002).
[20] 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 (1952).
[21] X. Zhang, T. J. Royston, H. A. Mansy, and R. H. Sandler, ''Radiation impedance of a finite circular piston on a viscoelastic half-space with application to medical diagnosis,'' J. Acoust. Soc. Am. **109**, 795–802 (2001).
[22] 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).

Prediction of backscatter coefficient in trabecular bones using a numerical model of three-dimensional microstructure

Frédéric Padilla[a)]
Laboratoire d'Imagerie Paramétrique, UMR CNRS 7623 Université Paris 6, 15 rue de l'Ecole de Médecine, 75006 Paris, France

Françoise Peyrin
CREATIS, UMR CNRS5515, INSA 502, 69621 Villeurbanne Cedex, France and ESRF, BP 220, 38043 Grenoble Cedex, France

Pascal Laugier
Laboratoire d'Imagerie Paramétrique, UMR CNRS 7623 Université Paris 6, 15 rue de l'Ecole de Médecine, 75006 Paris, France

(Received 23 February 2002; accepted for publication 1 November 2002)

A model of ultrasonic backscattering for cancellous bone saturated by water is proposed. This model assumes that scattering is caused by the solid trabeculae and describes the cancellous bone as a weak scattering medium. The backscatter coefficient is related to the spatial Fourier transform of bone microarchitecture and to the density and compressibility fluctuations between the solid trabeculae and the saturating fluid. The computations of the model make use of three-dimensional numerical images of bone microarchitecture, obtained by tomographic reconstructions with a 10 μm spatial resolution. With this model, the predictions of the frequency dependence and of the magnitude of the backscatter coefficient are reasonably accurate. The theoretical predictions are compared to experimental data obtained on 19 specimens. An accuracy error of approximately 1 dB was found (difference between the averaged experimental values and theoretical predictions). One limit of the model may come from inaccurate values of trabecular bone characteristics needed for the computations (density and longitudinal velocity), which are yet to be precisely determined for human trabecular bone. However, the model is only slightly sensitive to variations of bone material properties. It was found that an accuracy error of 2.2 dB at maximum resulted from inaccurate *a priori* values of bone material properties. A computation of the elastic mean free path in the medium suggests that multiple scattering plays a minor role in the working frequency bandwidth (0.4–1.2 MHz). It follows from these results that a weak scattering medium model may be appropriate to describe scattering from trabecular bone.
[DOI: 10.1121/1.1534835]

PACS numbers: 43.80.Ev, 43.80.Jz, 43.80.Qf, 43.80.Vj [FD]

I. INTRODUCTION

As the assessment of bone strength demands increasing noninvasive means to probe multiple bone properties, including bone mass and microarchitecture, ultrasonic backscattering has attracted the attention of a number of researchers for its potential to characterize bone microarchitecture. Cancellous bone is a highly porous and inhomogeneous medium, composed of a solid matrix (mineralized collagen) of interconnected trabeculae with diameter ranging from 50 to 200 μm filled with marrow. Usually, for *in vitro* ultrasonic experiments, the marrow is removed and bone specimens are filled with water. Strong ultrasonic scattering from the trabeculae is expected due to substantial mismatch in acoustic impedance between marrow (or water) and solid bone tissue. Cancellous bone loss due to aging or osteoporosis leads to increased porosity, thinning or even total disappearance of some trabecular elements and disruption of structure continuity. Assuming the trabeculae elements to be responsible for scattering of ultrasonic waves, measurement of backscattering might be an appropriate answer to address the above-mentioned issue of probing microarchitecture changes, because the backscatter cross section depends on scatterers elastic properties, spatial distribution, and size.

Measurements of ultrasonic backscattering from human calcaneus have been reported by different groups both *in vitro*[1–5] and *in vivo*.[6,7] The diagnostic sensitivity of integrated backscatter coefficient (so-called broadband ultrasonic backscatter in the bone-densitometry field) in discriminating osteoporotic patients from age-matched controls has been demonstrated.[8] The estimation of tissue microarchitectural features from backscatter data belongs to the class of inverse problems and has been successfully addressed in the field of soft tissue characterization.[9,10] This approach requires the development of specific appropriate scattering models: Estimates of scatterer characteristics, such as scatterer size or scattering strength, are obtained by a least-squares fit of the scattering model to experimental data. It has been successfully applied to estimate structural characteristics of normal and pathological human tissues such as the kidney,[11,12] the liver,[13,14] or the eye.[14] Recent developments of theoretical

[a)]Electronic mail: padilla@lip.bhdc.jussieu.fr

scattering models for cancellous bone have opened perspectives for the application of such approaches to noninvasive assessment of bone microarchitecture.

Two different approaches have been proposed to model ultrasonic scattering from cancellous bone, both assuming single scattering.

The first approach, proposed by Wear,[4,5] consists in solving the differential propagation equations in the ambient fluid and in the scatterers, and then to use the appropriate boundary conditions at the interface between fluid and scatterers. Analytical solutions can be derived for canonical geometries. In Wear's model, the trabecular network is represented by an assembly of randomly distributed identical cylindrical scatterers, aligned perpendicular to the direction of propagation of the incident wave. Assuming no multiple scattering, the backscatter coefficient of a collection of scatterers is obtained by the summation of the backscatter coefficient of each scatterer. The backscatter cross section of individual scatterers was computed using the analytical model of Faran.[15] The model proposed by Wear has given predictions of the frequency dependence of the backscatter coefficient which were in good agreement with experimental observations in the low frequency range.

The second approach consists in considering the medium as a fluid random continuum. The inhomogeneities are described as source terms which perturb the homogeneous wave equation in the ambient fluid. With the help of a Green function, the scattered pressure field is calculated by integrating the contribution of each source over a volume containing the scatterers. This scattered pressure can be expressed as a function of the spatial Fourier transform of a function describing the random inhomogeneities in density and compressibility.[9] This approach usually assumes weak scattering (Born approximation). For statistically homogeneous systems, one may perform an average of the values of the backscatter coefficient at different positions (equivalent to an average over different realizations of the process), and then relate the backscatter coefficient to the autocorrelation function of the medium. This method is classically used to describe scattering by inhomogeneities in soft tissues.[9] In order to apply such a model to bone, Strelitzki *et al.*[16] and Nicholson *et al.*[17] have suggested to model trabecular bone as a two-phase mixture. In their approach, fluctuations in sound speed accounted for scattering and an exponential autocorrelation function was used to describe the statistical properties of the random medium. The authors have shown that the order of magnitude of the model predictions was similar to some experimental values published by others, but no accurate comparison with experimental data was provided in their work.

Recently, we have developed a similar approach, but instead of using an exponential autocorrelation function to describe the medium, we have used data provided by high-resolution synchrotron radiation microcomputed tomography (μ-CT). First, these data allow the reconstruction of 2D and 3D bone microarchitecture with a spatial resolution of 10 μm. Second, an autocorrelation function was derived from two-dimensional microcomputed tomography (μ-CT) reconstructions of bone microarchitecture[1] and was used to compute the frequency dependence of the backscatter coefficient. This has resulted in close agreement between theoretical prediction of the frequency dependence of the backscatter coefficient and experimental data. One potential advantage of the random continuum approach over the discrete cylindrical model of Wear is a greater flexibility in representing the actual bone microarchitecture by using the data derived from the now available powerful high resolution microimaging techniques such as synchrotron radiation or x-ray μ-CT. In particular, the calculation can be achieved individually for each bone specimen, by taking into account the specific microarchitecture derived from μ-CT of each specimen under study.

This study is an extension of our previous work. In our former study, 2D numerical models of bone microarchitecture were used, thus imperfectly representing 3D microarchitecture. The objective of this work is to improve the theoretical prediction of the frequency dependence and to provide, for the first time, predictions of the magnitude of the backscatter coefficient using the random continuum model by implementing a calculation involving whole 3D numerical models of microarchitecture. In the present study, the microstructure of bone is not described in a statistical manner, but rather in a deterministic manner: direct computation of the spatial Fourier transform of the microstructure is performed to predict the backscatter coefficient.

II. EXPERIMENTAL PROCEDURE

The backscatter coefficient is defined as being the differential angular scattering cross section per unit volume of scatterers for the special case where the scattering angle equals π (scattering direction in the opposite of incident direction). The backscatter coefficient is obtained from the ratio of the scattered power to the incident intensity per unit solid angle and per unit volume of scatterers.

Backscatter coefficient measurements were performed on 19 human calcaneus samples removed from cadavers of age ranging from 75 to 90 years. Slices of pure trabecular bone with parallel faces and thickness of approximately 1 cm were cut from the whole calcaneal bone in the sagittal plane by removing the cortical shell. The specimens were defatted and refilled with water under vacuum for ultrasonic measurements. The experimental protocol of sample preparation has been described in a previous paper.[18] This preparation allows one to preserve the specimens for a long time, and to perform several measurements. Nicholson and Bouxsein[19] have reported that bone marrow influences quantitative ultrasound measurements in human calcaneus bone. Marrow significantly increased attenuation, attenuation slope, and backscatter coefficient compared to the water-saturated state. Their data indicate that the potential impact of marrow should be considered when interpreting clinical QUS measurements. However, this does not affect the validity of the model derived in the following, where a water-saturated state has been taken into account for the computations.

Measurements were performed using a substitution method.[20] Two focused transducers of 1 MHz central frequency were used. The frequency bandwidth was 0.4–1.2 MHz. The ultrasound beam axis was oriented perpendicular

to the surface of the samples (propagation in the medio-lateral direction as with clinical devices). The frequency-dependent attenuation coefficient was measured in transmission. Then, the backscatter coefficient was measured as follows. First, a reference echo was acquired on a plane reflector (steel plate) placed at a distance equals to the position of the scattering volume of the specimen under study. Then, an echo signal was acquired from the scattering of the incident pulse onto a bone specimen. This signal was time-weighted using a Hamming function in order to keep only part of the signal backscattered from a volume approximately 8 mm in length placed in the center of the specimen. Then the backscatter coefficient $\sigma(f)$ was calculated by computing the ratio of the frequency power spectrum of the time-gated echo signal to the power spectrum of the reference signal. Corrections were made to compensate for attenuation, Hamming gate function, and frequency-dependent scattering volume (diffraction). The detailed calculation may be found in a previous paper.[1] With this method, the intrinsic backscatter coefficient of the scattering volume is obtained, and it is independent of the characteristics of the measuring device and of experimental conditions.

Measurements of the backscatter coefficient on the specimen were performed on 2D scans with a step of 1 mm. At the scale of the whole calcaneus, the medium is strongly heterogeneous.[21] Heterogeneities in density and microarchitecture reflects the nonuniform spatial distribution of the stress applied to the heel. Therefore, 0.5 cm^2 (7×7 mm^2) square regions of interest (ROIs) were placed in homogeneous parts of the specimens. The values of attenuation and backscatter coefficients for one ROI were averaged over the 49 corresponding measurement points. It is usually considered that two *A* lines are uncorrelated if the overlap between two corresponding beam cross sections is less than 50%. Due to the important overlap between the ultrasound beam corresponding to two adjacent measurement points, these 49 *A* lines resulted in only 6 uncorrelated *A* lines. Since the sampling ROIs (0.5 cm^2) are much smaller than the calcaneal area, the medium was assumed to be statistically homogeneous at the scale of ROIs. Consequently, averaging the backscatter coefficient over the ROI potentially decreased the statistical variance compared to the variance of the backscatter coefficient derived from a single *A* line.

The 3D microarchitecture of the specimens was investigated using synchrotron radiation microtomography at the ESRF (European Synchrotron Radiation Facility, Grenoble, France). This technique provides 3D images of bone samples with a high spatial resolution (10 μm).[22] The field of view of the device was 1 cm^3 and only relatively small specimens could be investigated. Therefore, after ultrasonic testing, cylindrical cores of diameter 7 mm were cut from the slices, their axes being aligned along the medio-lateral axis (axis of propagation of the ultrasound), giving a total of 19 specimens. A volume of interest of $6.6\times6.6\times6.6$ mm^3, centrally located within each cylinder was reconstructed and analyzed. The voxel of the reconstructed images is cubic with a 10 μm side length (Fig. 1).

For each specimen, the tomographic reconstruction resulted in 3D numerical models of bone microarchitecture

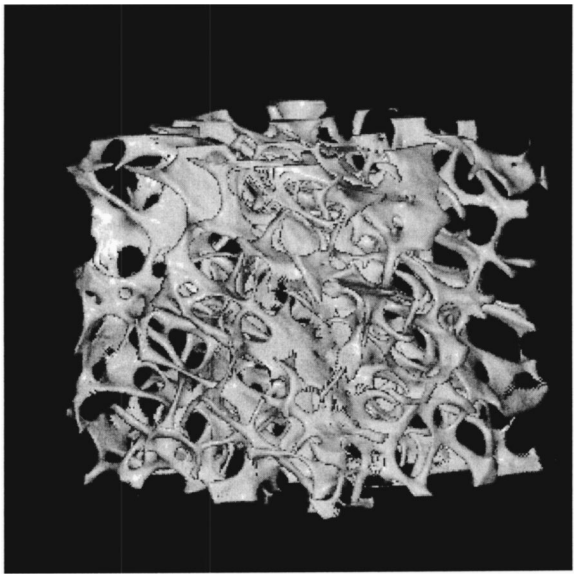


FIG. 1. 3D image of microarchitecture. The reconstructed volume is $6.6\times6.6\times6.6$ mm^3.

$I(x,y,z)$, where (x,y,z) are the Cartesian coordinates. The function $I(x,y,z)$ represents the linear attenuation coefficient values for the monochromatic energy of the x rays (set to 20 keV in this experiment). It is a gray level image (dynamic 8 bits) proportional to the density and to the mineral content of the bone trabeculae. A 3D display of a typical specimen is presented in Fig. 1. As we are only interested in microstructure, a function $I'(x,y,z)$ was derived from $I(x,y,z)$ such that

$$I'(\mathbf{r})=\begin{cases}1 & \text{for every voxel belonging to bone}\\ 0 & \text{for every voxel belonging to marrow.}\end{cases} \quad (1)$$

The function $I'(x,y,z)$ is a binary function derived from the 3D numerical image $I(x,y,z)$ by comparing it to a given threshold as follows: if $I(x,y,z)>$threshold then $I'(x,y,z)=1$, if $I(x,y,z)<$threshold then $I'(x,y,z)=0$. As synchrotron radiation microtomography provides highly contrasted and high resolution images of bone, the choice of a threshold was not a critical issue. The same threshold value was established for all the samples.

Photographs of the bones slices were taken after removal of the cylindrical cores. They were used to select on the ultrasonic backscatter 2D scans the ROI corresponding to the location of the cores.

III. MODEL OF BACKSCATTER COEFFICIENT

To model the backscatter coefficient, bone specimens (trabecular network saturated with water) are modeled as a medium composed of an ambient fluid and inhomogeneities. Water and bone are, respectively, characterized by their density ρ_f and ρ_b, and compressibility κ_f and κ_b [$\kappa=1/(\rho c^2)$] where c is the speed of sound of the longitudinal bulk waves: c_f for water and c_b for bone. Assumption is made that only longitudinal waves propagate and we neglect the possible longitudinal/transversal mode conversions at the solid–fluid interfaces.

This approach has been used successfully to model scattering by biological tissues,[10,9] or scattering by droplets in fog.[23] A classical derivation of the backscatter coefficient may be found in Ref. 23, and we will only recall what is necessary for our computations.

The propagation medium can be described by two functions called $\gamma_\rho(\mathbf{r})$ and $\gamma_\kappa(\mathbf{r})$ defined as

$$\gamma_\rho(\mathbf{r}) = \begin{cases} \dfrac{\rho_b(\mathbf{r}) - \rho_f}{\rho_b(\mathbf{r})} & \text{inside an inhomogeneity} \\ 0 \quad \text{outside,} \end{cases} \tag{2}$$

$$\gamma_\kappa(\mathbf{r}) = \begin{cases} \dfrac{\kappa_b(\mathbf{r}) - \kappa_f}{\kappa_f} & \text{inside an inhomogeneity} \\ 0 \quad \text{outside,} \end{cases} \tag{3}$$

where $\mathbf{r}$ is the spatial coordinate. These γ functions describe the variations in density and compressibility in bone compared to water. These variations cause scattering of the acoustic wave which propagates in the fluid with the wave number $k_f = \omega/c_f$. The scattered pressure field from an incident plane wave may be derived by considering these fluctuations as source terms which perturb the homogeneous wave equation in the fluid. The wave equation in the frequency domain (Helmholtz equation) can be solved with an integral formulation of the problem. Neglecting the absorption and multiple scattering effects (Born approximation), the backscatter coefficient $\sigma(f)$ may be written at an observation point $\mathbf{r}$ far enough from the scattering volume V as[23]

$$\sigma(f) = \frac{1}{V}\left| \frac{k_f^2}{4\pi} \int\int\int_V [\gamma_\kappa(\mathbf{r}_0) - \gamma_\rho(\mathbf{r}_0)] e^{-i\mathbf{K}.\mathbf{r}_0} d\mathbf{r}_0 \right|^2, \tag{4}$$

where $\mathbf{K} = -2k_f\hat{\mathbf{z}}$ with $\hat{\mathbf{z}}$ the unit vector in the direction of propagation of the incident wave, and $||$ designates the modulus.

It must be noticed from Eq. (4) that the backscatter coefficient is directly proportional to the spatial Fourier transform of the function $\gamma(\mathbf{r}) = \gamma_\kappa(\mathbf{r}) - \gamma_\rho(\mathbf{r})$. We will use this fundamental property to compute the backscatter coefficient from the reconstructed 3D numerical images of bone microarchitecture.

Equation (4) may now be decomposed into two parts describing two different effects: the coherent and the incoherent scattering. To do so, the γ function is written as

$$\gamma(\mathbf{r}) = \gamma'(\mathbf{r}) + \overline{\gamma}, \tag{5}$$

where $\overline{\gamma}$ is the mean value of $\gamma(\mathbf{r})$ over the volume V and $\gamma'(\mathbf{r})$ represents the fluctuations of $\gamma(\mathbf{r})$ around its mean value. Following the derivation of Morse and Ingard[23] to separate coherent and incoherent parts, we decomposed the backscatter coefficient as

$$\sigma(f) = \sigma'(f) + \overline{\sigma(f)}, \tag{6}$$

where

$$\begin{aligned} \sigma'(f) &= \frac{1}{V}\left| \frac{k_f^2}{4\pi} \int\int\int_V \gamma'(\mathbf{r}_0) e^{-i\mathbf{K}.\mathbf{r}_0} d\mathbf{r}_0 \right|^2, \\ \overline{\sigma(f)} &= \frac{1}{V}\left| \frac{k_f^2}{4\pi} \int\int\int_V \overline{\gamma} e^{-i\mathbf{K}.\mathbf{r}_0} d\mathbf{r}_0 \right|^2. \end{aligned} \tag{7}$$

The first term of the right member of Eq. (6) is the incoherent scattering contribution to $\sigma(f)$. It is generated by the variable part of the function $\gamma(\mathbf{r})$ around its mean value, and takes into account the interference between the waves scattered by different scatterers inside the volume V. This is the part of the backscatter coefficient we are interested in, because of the time signal processing on experimental data as explained in Sec. II.

The second term of the right member of Eq. (6) describes coherent backscattering: the volume V behaves as if it had uniform homogeneous density and compressibility, which differs from that in the ambient fluid outside V. This difference will produce backscattering of the wave incident on V. It is a cooperative effect from the scatterers which may be seen as a boundary effect of the volume V. For our purpose, we are not interested in this coherent contribution. Indeed, the experimental signals have been time gated to keep only the contribution from the inside of V, and the boundary effects are not taken into account. Moreover, the average value of $\gamma(\mathbf{r})$ on one ROI is approximately given by: $\overline{\gamma} = \gamma_0 V_f$ where $\gamma_0 = (\overline{\kappa_b} - \kappa_f)/\kappa_f - (\overline{\rho_b} - \rho_f)/\overline{\rho_b}$ (where $\overline{\kappa_b}$ and $\overline{\rho_b}$ are the average values of compressibility and density of the trabeculae over one ROI), and V_f is the volume fraction of bone in the ROI. As V_f is always small (the mean porosity of the samples is 0.93),[24] we will consider $\overline{\gamma}$ as negligible and we will assume that : $\gamma(\mathbf{r}) \approx \gamma'(\mathbf{r})$.

Finally, in the following of this paper, we consider the backscatter coefficient as

$$\sigma(f) \approx \sigma'(f). \tag{8}$$

To compute Eq. (8) from the numerical 3D images, we need one more assumption. We will consider that the values of density and compressibility are the same for every part of the solid matrix, and the same for every sample. The function $\gamma'(\mathbf{r})$ must be known for each sample. Assuming a uniform density and compressibility throughout bone tissue (i.e., $\rho_b(\mathbf{r}) = \rho_b$ and $\kappa_b(\mathbf{r}) = \kappa_b$), $\gamma(\mathbf{r})'$ is simplified to

$$\gamma(\mathbf{r})' = \begin{cases} \gamma_0 = \dfrac{\kappa_b - \kappa_f}{\kappa_f} - \dfrac{\rho_b - \rho_f}{\rho_b} \\ \qquad \text{for every voxel belonging to bone} \\ 0 \quad \text{for every voxel belonging to marrow.} \end{cases} \tag{9}$$

The function $\gamma(\mathbf{r})'$ can then be easily computed from the 3D reconstructed models of bone microarchitecture. Indeed, we can express $\gamma'(\mathbf{r})$ as

$$\gamma'(\mathbf{r}) = \gamma_0 I'(\mathbf{r}), \tag{10}$$

where

$$I'(\mathbf{r}) = \begin{cases} 1 \quad \text{for every voxel belonging to bone} \\ 0 \quad \text{for every voxel belonging to marrow,} \end{cases} \tag{11}$$

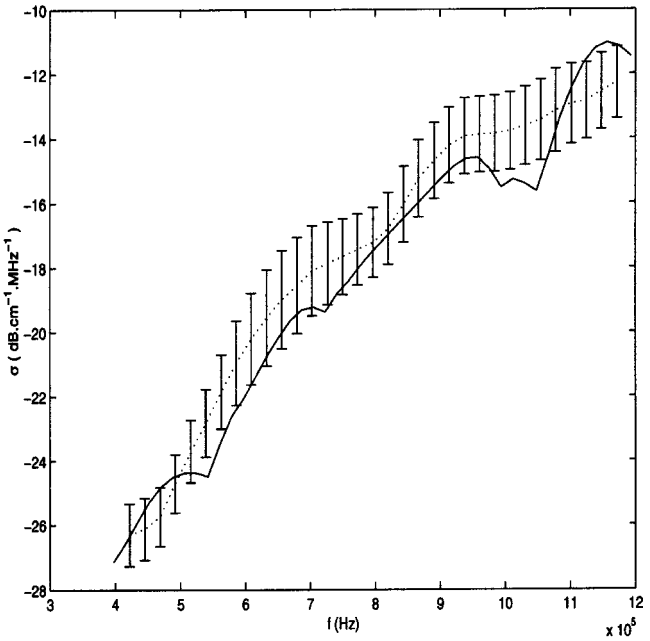


FIG. 2. Averaged backscatter coefficient over the 19 ROIs as a function of the frequency. Plain line: simulation, dashed line: experiment. Error bars: standard error.

as explained in Sec. II.

Let us now introduce a Cartesian coordinate system (x,y,z) such that the origin of the system is the center of the volume, and the z axis is directed along the direction of propagation of the incident wave. Let us define the function $B(z)$ by

$$B(z)=\int\int_S I'(x,y,z)dx\,dy, \tag{12}$$

where S is a section of the volume V normal to the z axis and including the point $(0,0,z)$. The backscatter coefficient may now be written as

$$\sigma(f)=\frac{1}{V}\left|\frac{k_0^2\gamma_0}{4\pi}\int_{z_{\min}}^{z_{\max}}B(z)e^{-2ik_0z}dz\right|^2, \tag{13}$$

where $z_{\min}$ and $z_{\max}$ represent the boundary of the volume V along the z axis. From Eq. (13), the calculation of the backscatter coefficient is now reduced to the computation of a one-dimensional (1D) spatial Fourier transform of a function $B(z)$. This function $B(z)$ represents the area of bone in a 2D section perpendicular to z axis. It should be noticed from Eqs. (12) and (13), that the theoretical backscatter coefficient is sensitive to the quantity of scatterers in a plane perpendicular to the direction of propagation of the ultrasound beam, but not sensitive to their spatial distribution in this plane.

To compute the backscatter coefficient using Eq. (13) we have proceeded as follows. The function $B(z)$ was first computed. Then, it was weighted by a Hamming function to avoid edge artifacts arising from the fact that $B(z)$ does not vanish at $z_{\min}$ and $z_{\max}$. A correction on the amplitude was applied to compensate for the Hamming gate function. Finally, the 1D Fourier transform was computed using a numerical algorithm with MATLAB software (MathWorks, Inc., Natick, MA). The results are presented in Sec. IV. The computation time is small. Once the thresholding of the images and the computation of the $B(z)$ function have been performed, the computation of the backscatter coefficient for one sample is quasi-instantaneous. It takes a few seconds with MATLAB and a PC (1.2 GHz).

IV. COMPARISON OF EXPERIMENTAL AND PREDICTED VALUES OF THE BACKSCATTER COEFFICIENT

For the computations, we have assumed the following values: $\rho_f=1000$ kg/m^3 and $c_f=1500$ m/s for water, $\rho_b=1960$ kg/m^3 and $c_b=3200$ m/s for trabeculae. The validity of these last values will be discussed later.

The group averaged backscatter coefficient over 19 bone specimens is plotted as a function of frequency in Fig. 2. The theoretical and experimental curves show a similar frequency variation. To assess the agreement between predicted values and experimental data, two parameters were used. First, a nonlinear fit $\sigma(f)\propto Af^n$ was performed on each sample to describe the frequency dependence. The averaged exponent n was found to be 3.33 ± 0.43 and 3.29 ± 0.17 for the model and the experimental data, respectively. This is in good agreement with previously reported results obtained by Chaffai *et al.*[1] with a 2D autocorrelation model and by Wear[5] using the cylinder model of Faran. As noted by Chaffai *et al.*,[1] a frequency dependence around $n=3.3$ lies between Rayleigh scattering by cylinders ($n=3$) and spheres ($n=4$). Second, the accuracy of the theoretical predictions was calculated as the absolute value of the difference between the experimental and theoretical curves, averaged over the frequency bandwidth. An accuracy error of 1 dB was found.

Results were averaged over the group of samples to decrease the statistical variance on backscatter estimates. Indeed, low variance estimates of backscatter usually requires ensemble averaging. In the field of biomedical ultrasound, ensemble averaging is usually obtained by spatial averaging. Statistical noise (speckle noise) is present in the numerical computations as well as in experimental data. However, the numerical estimates of backscatter take into account only one disorder realization for each specimen. The variance of numerical computation of backscatter is therefore larger than the variance on experimental estimates based on the average of six independent A lines. We are working on an improved computation from 3D distribution of bone in each sample to decrease the variance of this statistical noise, as is done experimentally, in order to obtain individual predictions.

Three assumptions of the model are now discussed: the values of ρ_b and c_b, the assumption of uniform ρ_b and c_b, and finally the assumption of single scattering.

One limit of the model for the predictions of the magnitude of the backscatter coefficient may come from the assumptions on *a priori* κ_b and ρ_b values that were used to compute the model. In Table I, the values of ρ_b and c_b used by different authors to model propagation in trabecular bone have been reported. These values have been used to compute the Biot's theory,[25–29] scattering by a collection of cylinders,[5] a multilayer model,[30,31] and a weak scattering medium model.[16] For the density, most of authors have assumed that

TABLE I. Values of ρ_b and c_b used by different authors to compute models of propagation through trabecular bone. (The last two papers propose models with no particular reference to any specie or skeletal site.)

	Species	Skeletal site	ρ_b (kg/m^3)	c_b (m/s)
Mc Kelvie and Palmer[a]	Human	Calcaneum	1800	3200
Wear[b]	Human	Calcaneum	3220	6790
Williams[c]	Bovine	Tibia	1960	3200
Lauriks *et al.*[d]	Bovine	Unknown	1720±360	···
Hosokawa and Otani[e,f]	Bovine	Femur	1960	3200
Hughes *et al.*[g]	Bovine	Femur, tibia	1960	3200
Padilla and Laugier[h]	···	···	1960	3200
Strelitzki *et al.*[i]	···	···	1800	3300

[a]Reference 25.
[b]Reference 5.
[c]Reference 26.
[d]Reference 27.
[e]Reference 28.
[f]Reference 29.
[g]Reference 30.
[h]Reference 31.
[i]Reference 16.

the density of the trabecular material was similar to the density of the cortical material.[16,25,26,28–31] Lauriks *et al.*[27] have obtained the density from measurements of apparent density and of porosity. For the velocity of bulk longitudinal waves in the solid, some authors have assumed the same value as in cortical bone,[16,25] Williams[26] has deduced it from a measure of the Young's modulus of the solid trabeculae and this value has been used by several authors.[28–31] Finally, Wear[5] has assumed that the properties of trabeculae were the same as those of hydroxyapatite (mineral content of trabeculae). For our computations, we have used the same values as in Refs. 26, 28–31, which were used to compute accurate predictions of phase velocity in trabecular bones. However, the accuracy of these values for human cancellous bone has yet to be demonstrated. Finally, any difference between the actual value of ρ_b and c_b and those which were used to compute Eq. (13) may result in a slight shift of the magnitude of theoretical predictions (for instance, modeling trabeculae as hydroxyapatite rather than cortical bone results in a shift of +2.2 dB of the backscatter coefficient).

Another possible source for the inaccuracy in the magnitude of the backscatter coefficient may come from the assumption of uniform density and compressibility in trabeculae and among different specimens. The assumption that ρ_b and c_b is uniform throughout the trabeculae is far from being realistic, as shown in Fig. 3, showing a 2D section of one bone specimen used in this study, which demonstrates spatial variation of density within the trabeculae. Similarly, it is likely that significant variations can be found between individuals. This may be the source for an additional inaccuracy in the prediction. One solution might be to use real data taking into account the spatial variations of density and velocity properties within the trabeculae to compute the backscatter coefficient rather than the binary 3D numerical model obtained with a threshold.

The scattering model assumes single scattering, although a severe mismatch in material properties exists at the interface between the fluid and the solid phases. The assumption of multiple scattering has been discussed by Wear[5] in a paper

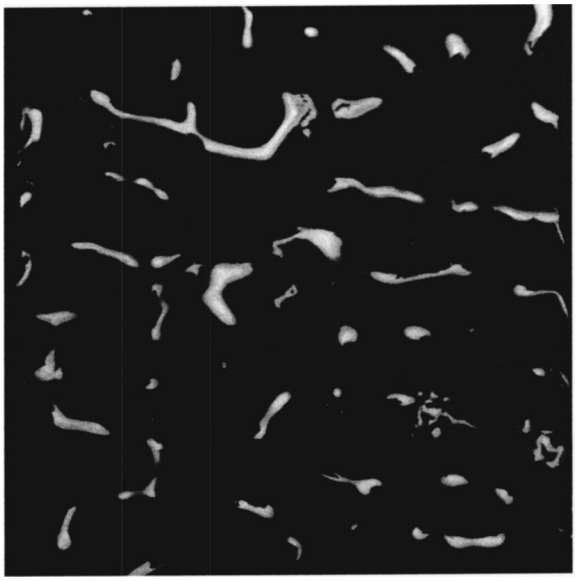


FIG. 3. Details of a 2D image of the model of reconstructed microarchitecture. The field of view is approximately 3×3 mm^2.

where trabecular microarchitecture is modeled as a collection of randomly distributed cylinders. Based upon the good agreement between predicted and experimental frequency dependence, it has been argued that multiple scattering should not play a major role in the frequency bandwidth.

In order to estimate whether substantial multiple scattering may be expected in our frequency range (0.4–1.2 MHz), a rough estimate of the elastic mean free path $l^*(f)$ of the medium has been performed. The elastic mean free path may be considered as an average distance between two subsequent scattering events.[32] The elastic mean free path is related in a complex manner to the elastic scattering strength of the medium. The readers will find detailed discussions about this issue in Sornette and Derode's papers.[32,33] One of the conditions that must be satisfied to observe multiple scattering is that the thickness of the sample $L(\approx 1$ cm) must be greater than the elastic mean free path. To obtain an approximation of the length $l^*(f)$, the bone microstructure was assumed to be similar to a collection of randomly positioned identical cylinders of infinite height, with axis oriented perpendicularly to the propagation direction of the ultrasound wave. As long as the number of scatterers per unit of volume is not too high, the elastic mean free path may be computed as[34,33]

$$l^*(f) = \frac{1}{N\sigma^*(f)}, \tag{14}$$

where N is the number of scatterers per unit area, and σ^* is the total scattering cross section of one cylinder.[23] According to Eq. (14), for a fixed frequency, the larger the mean size and number of scatterers, the smaller the mean free path. We have computed the mean free path for an extreme case, i.e., for the specimen with the highest bone volume fraction (lowest porosity of 0.75) and the larger mean trabecular thickness (150 μm). Under these conditions, the values of mean free path corresponding to the 0.4–1.2 MHz frequency band-

width falls in the range 9–1.2 cm and remains larger than the thickness of the specimens in the working frequency bandwidth. This suggests that multiple scattering should not be significant in the present experiments. Furthermore, the calculation of the mean free path suggests that substantial multiple scattering could be expected for frequencies greater than 1.3 MHz, which is above the frequencies used here. To be observed, it requires the absorption not to be too high. Given the complexity of these phenomena, this point should require a specific study.

V. CONCLUSION

A model of ultrasonic backscatter for cancellous bone was proposed, using for the first time 3D numerical models of bone microarchitecture. This model describes the cancellous bone as a weak scattering medium and the backscatter coefficient is related to the spatial Fourier transform of bone microarchitecture. With this model, the predictions of the frequency dependence and of the magnitude of the backscatter coefficient were reasonably accurate. An accuracy error of 1 dB was found. Improvement of the model might come from a better knowledge of trabecular bone characteristics (density and velocity), which are to be precisely determined for human trabecular bone. A computation of the elastic mean free path in the medium suggested that multiple scattering only plays a minor role in our working frequency bandwidth (0.4–1.2 MHz). Moreover, the computations from 3D distribution of bone in each sample has yet to be improved in order to obtain individual predictions.

It follows from these results that a weak scattering medium model may be appropriate to describe scattering from trabecular bone. With this model, there is a twofold perspective.

The first one is to improve bone characterization: to extract characteristics about bone microarchitecture and mineralization from backscatter measurements. As for soft tissue characterization, an appropriate solution might be to characterize bone microarchitecture using autocorrelation functions. This involves a two-step research. The first step is to validate the approach of the autocorrelation function using 3D numerical models of bone microarchitecture. The second step is to determine the autocorrelation model which gives the best prediction of the backscatter coefficient.

The second perspective is to compute ''virtual'' osteoporosis. By doing image processing on the numerical 3D models of microarchitecture, one may simulate the effects of osteoporosis on trabecular bone (by decreasing the trabecular thickness or by removing some trabeculae). Then, one may compute the backscatter coefficient with the modified microarchitecture. It should then be possible to determine which modifications of microarchitecture have the most influence on the backscatter coefficient.

ACKNOWLEDGMENT

The authors want to thank Dr. Arnaud Derode from LOA/University Paris 7 for useful discussions about multiple scattering.

[1] S. Chaffaï, V. Roberjot, F. Peyrin, G. Berger, and P. Laugier, ''Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results,'' J. Acoust. Soc. Am. **108**, 2403 (2000).
[2] V. Roberjot, P. Laugier, P. Giat, and G. Berger, ''Measurement of integrated backscatter coefficient of trabecular bone,'' in *Proceedings of the 1996 IEEE Ultrasonics Symposium*, pp. 1123–1126.
[3] K. A. Wear, ''The relationship between ultrasonic backscatter and bone mineral density in human calcaneus,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 777–780 (2000).
[4] K. A. Wear, ''Anisotropy of ultrasonic backscatter and attenuation from human calcaneus: Implications for relative roles of absorption and scattering in determining attenuation,'' J. Acoust. Soc. Am. **107**, 3474 (2000).
[5] K. A. Wear, ''Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment,'' J. Acoust. Soc. Am. **106**, 3659–3664 (1999).
[6] P. Laugier, P. Giat, C. Chappard, Ch. Roux, and G. Berger, ''Clinical assessment of the backscatter coefficient in osteoporosis,'' in Proceedings of 1997 IEEE Ultrasonics Symposium, pp. 1105–1104.
[7] K. A. Wear and B. S. Garra, ''Assessment of bone density using ultrasonic backscatter,'' Ultrasound Med. Biol. **24**, 689 (1998).
[8] C. Roux, V. Roberjot, R. Porcher, S. Kolta, M. Dougados, and P. Laugier, ''Ultrasonic backscatter and transmission parameters at the os calcis in postmenopausal osteoporosis,'' J. Bone Miner. Res. **16**, 1353–1362 (2001).
[9] *Ultrasonic Scattering in Biological Tissues*, edited by K. K. Shung and G. A. Thieme (CCR, London, 1993).
[10] D. K. Nassiri and C. R. Hill, ''The use of angular acoustic scattering measurements to estimate structural parameters of human and animal tissues,'' J. Acoust. Soc. Am. **79**, 2048–2054 (1986).
[11] M. F. Insana, ''Modeling acoustic backscatter from kidney microstructure using an anisotropic correlation function,'' J. Acoust. Soc. Am. **97**, 649–655 (1995).
[12] M. F. Insana, T. J. Hall, and J. L. Fishback, ''Identifying acoustic scattering sources in normal renal parenchyma for the anisotropy in acoustic properties,'' Ultrasound Med. Biol. **17**, 613–626 (1991).
[13] R. C. Waag, D. Dalecki, and P. E. Christopher, ''Spectral power determinations of compressibility and density variations in model media and calf liver using ultrasound,'' J. Acoust. Soc. Am. **85**, 423–431 (1989).
[14] F. L. Lizzi, M. Ostromogilsky, E. J. Feleppa, M. C. Rorke, and M. M. Yaremko, ''Relationship of ultrasonic spectral parameters to features of tissue microstructure,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 319–329 (1986).
[15] J. J. Faran, ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**, 405–418 (1951).
[16] R. Strelitzki, P. H. F. Nicholson, and V. Paech, ''A model for ultrasonic scattering in cancellous bone based on velocity fluctuations in a binary mixture,'' Physiol. Meas **19**, 189–196 (1998).
[17] P. H. F. Nicholson, R. Strelitzki, R. O. Cleveland, and M. L. Bouxsein, ''Scattering of ultrasound in cancellous bone: Predictions from a theoretical model,'' J. Biomech. **33**, 503–506 (2000).
[18] P. Laugier, P. Droin, A.-M. Laval-Jeantet, and G. Berger, ''*In vitro* assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative parametric imaging,'' Bone (N.Y.) **20**, 157–165 (1997).
[19] P. H. F. Nicholson and M. L. Bouxsein, ''Bone marrow influences quantitative ultrasound measurements in human cancellous bone,'' Ultrasound Med. Biol. **28**, 369–375 (2002).
[20] S. Chaffaï, F. Padilla, G. Berger, and P. Laugier, ''*In vitro* measurement of the frequency-dependent attenuation in cancelous bone between 0.2 and 2 MHz,'' J. Acoust. Soc. Am. **108**, 2403–2411 (2000).
[21] N. L. Jhamaria, K. B. Lal, M. Udawat, P. Banerji, and S. G. Kabra, ''The trabecular pattern of the calcaneum as an index of osteoporosis,'' J. Bone Jt. Surg., Br. Vol. **65**, 195–198 (1983).
[22] M. Salome, F. Peyrin, P. Cloetens, C. Odet, A.-M. Laval-Jeantet, J. Baruchel, and P. Spanne, ''A synchrotron radiation microtomography system for the analysis of trabecular bone samples,'' Med. Phys. **26**, 2194 (1999).
[23] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, 1986).
[24] S. Chaffaï, F. Peyrin, S. Nuzzo, R. Porcher, G. Berger, and P. Laugier, ''Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microarchitecture,'' Bone (N.Y.) **30**, 229–237 (2002).

[25] M. L. Mc Kelvie and S. B. Palmer, ''The interaction of ultrasound with cancellous bone,'' Phys. Med. Biol. **36**, 1331–1340 (1991).
[26] J. L. Williams, ''Ultrasonic wave propagation in cancellous and cortical bone: Prediction of experimental results by Biot's theory,'' J. Acoust. Soc. Am. **91**, 1106–1112 (1992).
[27] W. Lauriks, J. Thoen, I. van Asbroeck, G. Lowet, and G. Van der Perre, ''Propagation of ultrasonic pulses through trabecular bone,'' J. Phys. IV **4**, 1255–1258 (1994).
[28] A. Hosokawa and T. Otani, ''Ultrasonic wave propagation in bovine cancellous bone,'' J. Acoust. Soc. Am. **101**, 558–562 (1997).
[29] A. Hosokawa and T. Otani, ''Acoustic anisotropy in bovine cancellous bone,'' J. Acoust. Soc. Am. **103**, 2718–2722 (1998).
[30] E. R. Hughes, T. G. Leighton, G. W. Petley, and P. R. White, ''Ultrasonic propagation in cancellous bone: A new stratified model,'' Ultrasound Med. Biol. **25**, 811–821 (1998).
[31] F. Padilla and P. Laugier, ''Phase and group velocities of fast and slow compressional waves in trabecular bone,'' J. Acoust. Soc. Am. **108**, 1949–1952 (2000).
[32] D. Sornette, ''Acoustic waves in random media. I. Weak disorder regime,'' Acustica **67**, 199–215 (1989).
[33] A. Derode, A. Tourin, and M. Fink, ''Random multiple scattering of ultrasound. I. Coherent ballistic waves,'' Phys. Rev. E **64:036605**, 1–7 (2001).
[34] P. Sheng, *Introduction to Wave Scattering, Localization and Mesoscopic Phenomena* (Academic, San Diego, 1995).

Audiogram of a striped dolphin (*Stenella coeruleoalba*)

Ronald A. Kastelein[a] and Monique Hagedoorn
Harderwijk Marine Mammal Park, Strandboulevard-oost 1, 3841 AB Harderwijk, The Netherlands[b]

Whitlow W. L. Au
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Dick de Haan
Netherlands Institute for Fisheries Research (RIVO-DLO), P.O. Box 68, 1970 AB IJmuiden, The Netherlands

(Received 22 March 2002; accepted for publication 29 October 2002)

The underwater hearing sensitivity of a striped dolphin was measured in a pool using standard psycho-acoustic techniques. The go/no-go response paradigm and up–down staircase psychometric method were used. Auditory sensitivity was measured by using 12 narrow-band frequency-modulated signals having center frequencies between 0.5 and 160 kHz. The 50% detection threshold was determined for each frequency. The resulting audiogram for this animal was U-shaped, with hearing capabilities from 0.5 to 160 kHz (8 $\frac{1}{3}$ oct). Maximum sensitivity (42 dB *re* 1 μPa) occurred at 64 kHz. The range of most sensitive hearing (defined as the frequency range with sensitivities within 10 dB of maximum sensitivity) was from 29 to 123 kHz (approximately 2 oct). The animal's hearing became less sensitive below 32 kHz and above 120 kHz. Sensitivity decreased by about 8 dB per octave below 1 kHz and fell sharply at a rate of about 390 dB per octave above 140 kHz. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1532310]

PACS numbers: 43.80.Lb, 43.80.Ev, 43.80.Jz [FD]

I. INTRODUCTION

The striped dolphin (*Stenella coeruleoalba*; Meyen, 1833) is a medium-sized (1.8–2.5 m) odontocete found in tropical and warm-temperate waters around the world. It occurs primarily offshore and, when found close to land, usually in deep water. It is a very social species; school sizes range from a few tens to several thousand (Perrin *et al.*, 1994). This species is often incidentally captured in purse seine, gill net, and trawler fisheries (Perrin *et al.*, 1994; Aguilar, 2000). Possibly these bycatches could be reduced by deterring the dolphins from fishing nets with acoustic signals, which seems a promising method for reducing bycatch of harbor porpoises (*Phocoena phocoena*) in gill net fisheries (Anonymous, 2000). However, no information is available about the underwater hearing sensitivity of striped dolphins, nor about that of any other species of the genus *Stenella*. Striped dolphins are not usually available for testing, because they are rarely kept in captivity. However, the opportunity arose to conduct hearing experiments after a striped dolphin that had stranded on the Dutch coast had been rehabilitated at the Harderwijk Marine Mammal Park, The Netherlands. This was the second time in recorded history that this species stranded on the Dutch coast; the first stranding having occurred in 1967. Investigation of the hearing sensitivity of the striped dolphin increases the fundamental knowledge of hearing abilities in odontocetes, but also contributes to determining the optimal characteristics of acoustic deterring devices (e.g., frequency, source level) that might be used to reduce the bycatch of striped dolphins.

[a]Present address: Julianalaan 46, 3843 CC Harderwijk, The Netherlands. Electronic address: triessch@xs4all.nl
[b]Part of Grévin et Compagnie, France.

II. MATERIALS AND METHODS

A. Subject

The study animal was a female striped dolphin (code ScSH001) which stranded on the coast of the Netherlands on 9 December 1997 and had been rehabilitated at the Netherlands Cetacean Research and Rehabilitation Center at the Harderwijk Marine Mammal Park. The dolphin's estimated age at the time of stranding was between 3 (based on body length: 184 cm) and 4 years (based on body weight: 62 kg; Di-Méglio *et al.*, 1996).

During the experiment, the animal was estimated to be between 6 and 7 years old (maximum age is estimated to be 58 years by Perrin *et al.*, 1994), and had a body weight of 81 kg, a body length of 208 cm, and a girth anterior to the pectoral fins (at the auditory meatus) of 93 cm. Veterinary records showed that the animal had not been exposed to ototoxic medication. She had no previous experience with psycho-acoustic experiments.

The animal received approximately 3 kg of thawed fish (herring, *Clupea harengus*) per day, divided over five meals. The meal size during a hearing test session was disproportionally large (approximately 0.8 kg). The diet was supplemented by vitamins developed specially for marine mammals (Akwavit, Twilmij B.V., Stroe, The Netherlands).

B. Facility

The animal was kept in an indoor concrete oval pool [8.6 m(l)×6.3 m(w), 1.2 m deep; Fig. 1] at the Research and Rehabilitation Center at the Harderwijk Marine Mammal Park, The Netherlands. The water level was kept constant. The average water temperature was 19.5 °C and the average

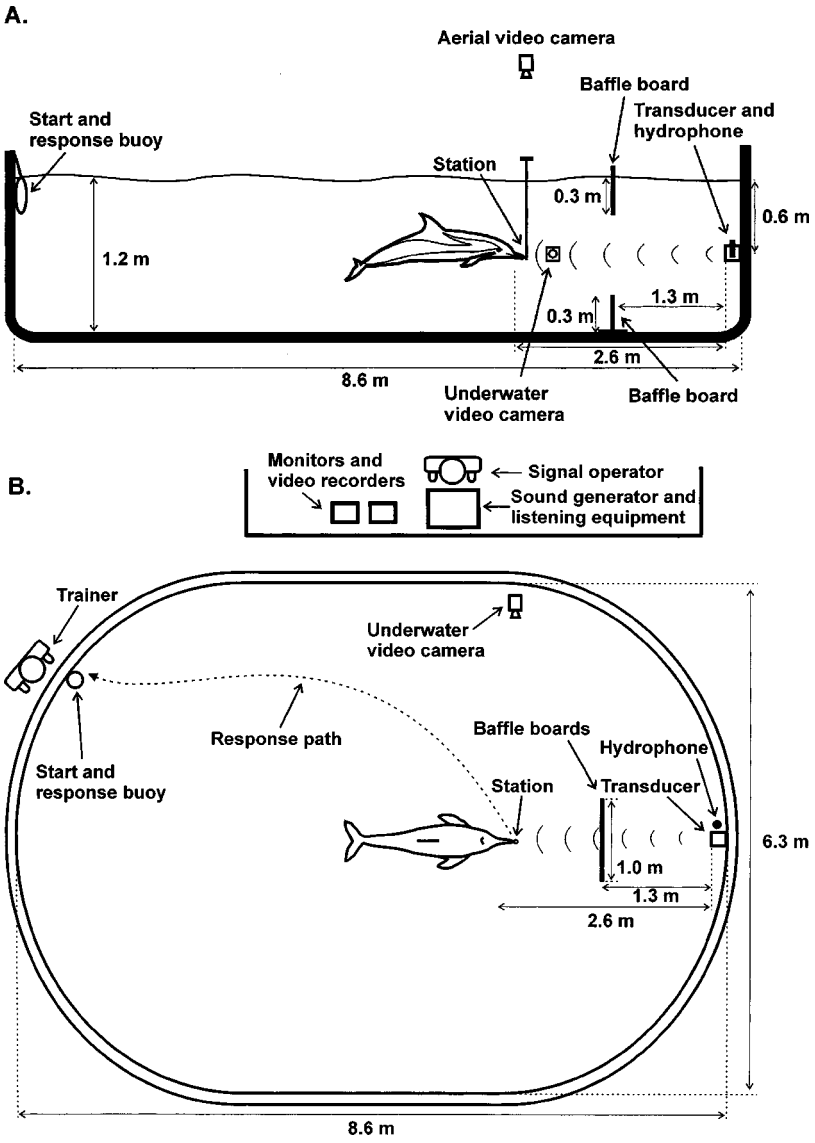


FIG. 1. The study area, showing the striped dolphin in the correct position at the listening station, and her response path: (a) side view and (b) top view.

salinity 2.2% NaCl. The water circulation pump (and the pump of the other pool in the building) was shut off 10 min before and during sessions. As well as reducing ambient noise, this ensured that no current was present in the pool during the experiments. During the study period the study animal shared the pool with a 3.5-year-old male harbor porpoise (PpSH047, *Phocoena phocoena*), and part of the time with an additional 2.5-year-old male harbor porpoise (PpSH052). Nonstudy animals were kept quiet and fed at the opposite end of the pool during the experimental sessions to eliminate any distractions and interferences during a session. An adjacent room (7 m from the station) served as the observation and data collection laboratory, where all the controlling electronics were housed and where the equipment operator was seated during the experiments (Fig. 1). The operation of the equipment was not visible to the dolphin.

C. Signals and signal generation

Signals were produced by a wave form generator (Hewlett Packard, model 33120A). Each acoustic stimulus consisted of a narrow-band sinusoidal frequency-modulated (FM) signal (wobble) of 2.0-s duration. The signals had 150-ms rise and fall times to prevent abrupt signal onset and offset transients. The steady-state portion of the signal thus was 1.7 s. The frequency-modulation range of each stimulus signal was $\pm 1\%$ of the center frequency (i.e., the frequency around which the signal fluctuated symmetrically), and the modulation frequency was 100 Hz. For example, if the signal's center frequency was 100 kHz, the frequency fluctuated 100 times per second (100 Hz) between 99 and 101 kHz ($\pm 1\%$). A custom-built signal shaper and attenuator was used to control the amplitude of the signals. The SPL of the signals at the dolphin's head while it was at the underwater listening station could be varied in 1-dB steps. Before each session, the voltage output level of the system at the input of the transducer or impedance matching transformer was verified with the calibrated levels (see Sec. II D) by using an oscilloscope (Dynatek, model 8120; Fig. 2).

The lower frequency signals (500 Hz to 32 kHz) were projected by an underwater LF piezoelectric transducer (Ocean Engineering Enterprise, USA, model DRS-8; 25-cm diameter) with an impedance matching transformer (Ocean Engineering Enterprise, USA). The 0.5- and 1-kHz signals were amplified with an audio amplifier (Technics SA-110K),

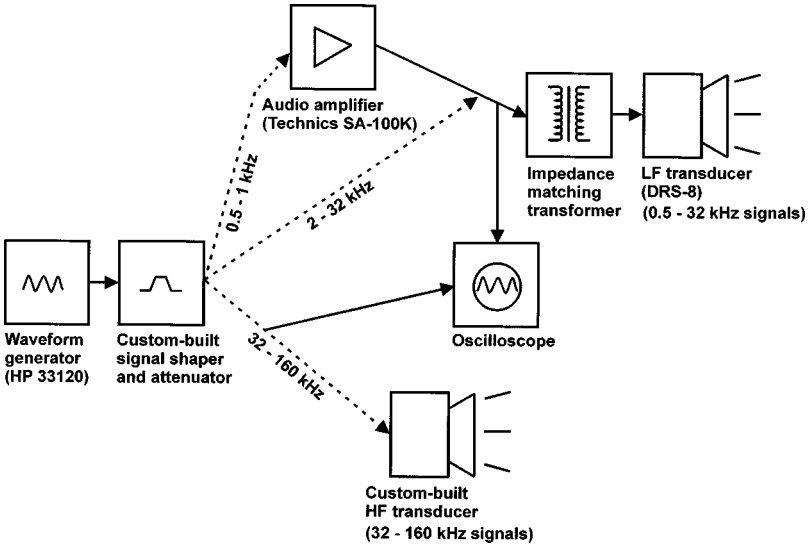


FIG. 2. Block diagram of the signal generation system used in the hearing study.

and the 2–32-kHz signals needed no additional amplification. The higher frequency signals (32–160 kHz) were projected by a custom-built transducer consisting of a circular disk of 1-3 composite piezoelectric material (Material Systems Inc., Littleton, MA), 0.64-cm thickness and 6.4-cm diameter and an effective aperture diameter of 4.5 cm. It was encapsulated in degassed polyurethane epoxy. No additional amplifier was needed for the 32–160-kHz signals. The dolphin's hearing threshold for the 32-kHz signal was tested with both transducers. The sound propagation beam of each transducer was always aligned with the animal's body axis while it was at the station.

The dolphin's hearing sensitivity was measured at center frequencies of 0.5, 1, 2, 4, 8, 16, 32, 40, 64, 120, 140, and 160 kHz. The low frequency cutoff of 500 Hz was determined by the limits of the sound production system, and was the lowest frequency the system could produce at sufficient (i.e., audible to the dolphin) amplitude without distortions. The high frequency cutoff of 160 kHz was determined during the study by the high frequency hearing limit of the animal. The −3-dB beamwidth at 160 kHz is approximately 12.7 degrees. At a distance of 2.6 m, the-3-dB beamwidth will cover a circle having a diameter of 58 cm, which is larger than the head of the dolphin. Since for a specific circular piston transducer the beam width is inversely proportional to frequency, the circle diameter increases as frequency decreases. The advantage of the transducers used in this study is their directionality, which results in low reflection from the sides of the pool. To reduce reflections from the water surface and the pool floor which might affect the signal SPL at the animal's head, two baffle boards (6 mm thick, 30×100 cm^2 aluminum plates covered with closed-cell neoprene) were placed perpendicular to the animal's axis, one breaking the water surface, and one at the pool floor (Fig. 1). The two boards were connected with nylon rope so they could be removed from the pool between sessions.

D. Signal calibration and monitoring

The root-mean-square (rms) SPL (dB *re* 1 μPa) for each test frequency was calibrated each month at the same position of the dolphin's head when it was at the listening station during test sessions (2.6 m from the transducer). The dolphin was not at the station during calibrations. The 10-cm grid measurements in all six directions showed only 2–4 dB differences up to 40 cm from the center of the grid (the location between the animal's ears). The position of the dolphin's head (while at the station) relative to the transducer was very consistent (within a few cm in all six directions), so the animal did not adjust its position to a spot with potentially higher (or lower) signal SPL. The calibration equipment used for all frequencies consisted of a hydrophone [Brüel & Kjaer (B&K) 8101], the calibration curve of which showed that its frequency response was flat up to 100 kHz, a conditioning amplifier (B&K, Nexus 2690), connected via a BNC-2090 (National Instruments) coaxial module to a computer with an analog input/output card (National Instruments, PCI-MIO-16E-1, 12-bit resolution). The system was calibrated with a pistonphone (B&K, 4223). For the calibration of signals above 100 kHz, the frequency response of the measurement system was taken into account. To confirm this approach, a second hydrophone (B&K 8103) with a flat response (+1 dB/−2 dB) up to about 120 kHz was used with the frequency response of the measurement system taken into account. The values obtained with the B&K 8103 matched the results obtained with the B&K 8101. During the calibration, the SPL was measured for levels that were above the threshold levels found in the present study. The linearity of the attenuator was checked frequently and was very precise.

The signals were digitized at a sample rate of 512 kHz and fast Fourier transformed (FFT) into the frequency domain using a Hanning window. The highest peak in the spectrum was selected to determine the SPL and five consecutive 0.2-s time blocks were used to calculate the average SPL. The SPLs of five consecutive 0.2-s time blocks were used to calculate the average SPL. Each month, the average SPL of each signal frequency was measured. During the calibration, the SPL was measured for levels that were about 12 dB above the 50% detection threshold levels found in the present study. The analysis of the signals in the frequency

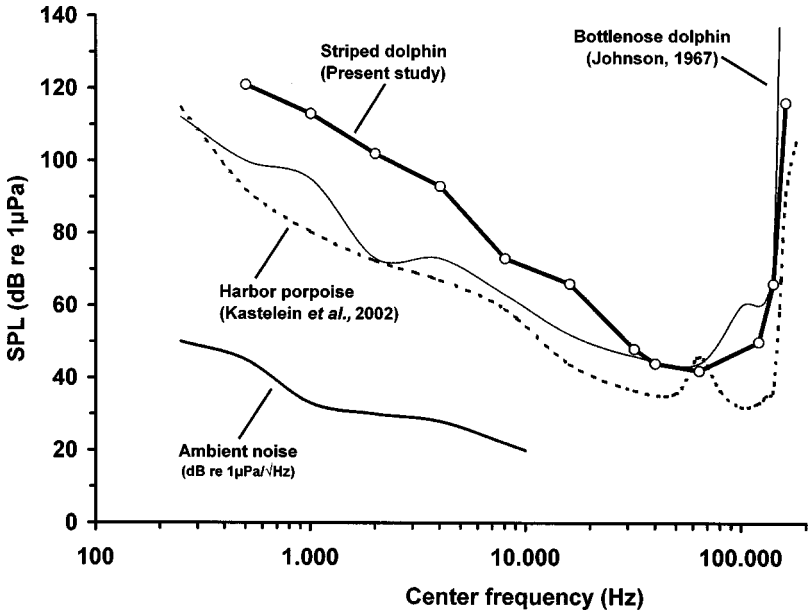


FIG. 3. The mean 50% detection thresholds (dB *re* 1 μ Pa, rms) of the striped dolphin in the present study for the 12 tested narrow-band FM signals (for details see Table I). Also shown is the underwater audiogram of a harbor porpoise (Kastelein *et al.*, 2002; obtained with the same equipment, very similar methodology, in the same study area and during overlapping study periods as that of the dolphin in the present study), and the audiogram of an Atlantic bottlenose dolphin (Johnson, 1967). The spectral level of the ambient noise in the pool up to 10 kHz is also shown (the level is expressed as dB *re* 1 μ Pa/$\sqrt{}$Hz; note that this is a different unit than the one along the *Y* axis).

domain was compared to rms analysis in the time domain, and the results matched. The linearity of the attenuator was also checked frequently.

Signals were analyzed with special attention to potential harmonics especially with the low frequency high amplitude signals (0.5 and 1 kHz). The LF signals produced harmonics with energy well below the hearing thresholds obtained for the frequencies of the harmonics. Calculation of average signal SPLs used for threshold determination was based on all monthly calibrations. The monthly calibrations varied within 4 dB.

Checks were often made for potential transients caused by pressing the signal button, both with and without the sound generator attached to the signal shaper/attenuator, or with the sound generator attached, but with the amplitude setting at 0. In each case, the animal did not respond to the action of pressing the signal button. Transient checks were performed for low and high amplitude settings of the sound generating system.

Before each session, the system was further verified by aurally monitoring the stimulus (usually at a higher amplitude than used during the session) via a hydrophone (LabForce 1BV) positioned directly adjacent to the transmitting transducer. The output of the monitoring hydrophone was connected to either an amplifier and loudspeaker for the frequencies up to 16 kHz, or to a bat-detector (Batbox III; Stag Electronics, Steyning, UK) for the signals with frequencies 32 kHz and above (maximum frequency possible was 120 kHz).

E. Background noise

Man-made noises in the vicinity of the test pool could have been directly coupled into the water, producing potentially significant masking to the test signals during sessions. Therefore, all indoor activities were stopped during sessions (nobody was allowed to move in the building). The underwater ambient noise level was measured up to 10 kHz (above this frequency, the ambient noise level was lower than the electronic noise level, and thus could not be measured) by using the B&K 8101 hydrophone and the earlier described acoustic amplifier and conversion equipment, under the same conditions as during the sessions. To allow comparison of the ambient levels and the hearing threshold levels, the recording and analysis methodology of the ambient levels was the same as for the stimuli calibrations described above. The recorded ambient signal was analyzed by FFT in ten blocks of 0.2 s. Of each block, the ambient levels of the tested frequencies were exported to a spreadsheet to calculate the average level over 2 s (ten blocks). The levels are plotted in Fig. 3.

F. Experimental procedure

Training the dolphin for the hearing experiment took two months (July and August 2000). Operant conditioning using positive reinforcement was used for all training. Due to the consistent behavior of the animal, no warm-up trials were conducted before an actual session began. A session began after the signal production equipment had been setup and the signal operator had set the frequency and the SPL for the first trial of the session. The amplitude of the signal for the first trial of the first session of each frequency was set at about 20 dB above the threshold reported for bottlenose dolphins (Johnson, 1967). A trial began with the animal stationed at the start and response buoy. When the trainer rang a bell, the animal swam to the listening station, which was the end of a 3-cm-diam water-filled PVC tube, and placed its head so that its auditory meatus was 2.6 m from the sound source, about 65 cm below the water surface (Fig. 1). She was trained to station with the tip of her rostrum at the station and her body axis in line with the beam of the transducer. Because the dolphin was expected to have directional hearing (Au and Moore, 1984; Schlundt *et al.*, 2002), a maximum deviation in the dolphin's position of only 5 degrees from the beam axis was accepted in all directions. Trials were canceled when the animal was not in the correct position. Two video cameras were used to check the animal's position at the sta-

tion, as well as to monitor her response (the images were visible in the signal operator's room). The dolphin was filmed from her left side by an underwater video camera (Mariscope, Micro, Kiel, Germany; Fig. 1) and from above by an aerial camera (Hapé, model CA 28) hung from the ceiling just above the water surface.

After the animal was correctly positioned at the station, two trial types could occur: signal-present trials and signal-absent trials. Signals were initiated following a random delay of 3 to 6 s after the dolphin was stationed. If the animal detected the sound it left the station (go response) at any time during the 2-s signal duration and returned to the start and response buoy [Fig. 1(b)]. The signal operator told the trainer that the response was correct (a hit), after which the trainer gave a vocal signal and the dolphin received a fish. If the animal did not respond to the sound (a miss), the signal operator would tell the trainer that the animal's response was incorrect. The trainer would then signal to the animal (by tapping on the side of the pool) that the trial had ended, thus calling the animal back to the start and response buoy. No reward was given. If the animal moved away before a signal was produced (a prestimulus response), the signal operator would indicate to the trainer to end the trial and not provide a reward to the animal.

For signal-absent trials, the signal operator told the trainer, after a random time period between 4 and 10 s after the dolphin had stationed, to end the trial. The trainer did this by blowing a whistle. In the case of a correct response (a correct rejection), the animal would return to the start buoy and receive a fish reward. If the dolphin left the station before the whistle was blown (a prestimulus response), the signal operator indicated to the trainer to end the trial and not reward the animal.

The amount of fish given as a reward for correct signal-present and signal-absent trials was the same. After a correct response trial, the next trial would start as soon as the dolphin voluntarily returned to the start buoy.

A single frequency was tested during each session. A modified up/down staircase psychometric technique was used (Robinson and Watkins, 1973), a variance of the method of limits, which results in a 50% correct detection threshold (Levitt, 1971).

If the animal heard a signal and responded to it (a hit), the next signal presented was 4 dB less intense. If the animal did not hear a signal and remained at the station (a miss), the following signal levels were increased in 4-dB steps until the animal detected the signal again. The starting SPL of a session, except for the first session of a particular frequency, was set at 12–16 dB above the threshold found during that frequency's previous session. Prestimulus responses did not result in a change in signal amplitude for the next signal trial. A session usually consisted of 12–25 trials and lasted for about 20 min. Because this animal refused to eat small fish and pieces of fish, the maximum possible number of trials per sessions was fewer than in similar studies with bottlenose dolphins *Tursiops truncatus* (Johnson, 1967) and false killer whales, *Pseudorca crassidens* (Thomas *et al.*, 1988). Each session consisted of 50% signal-present and 50% signal-absent trials based on a pseudo-random series table (Gellermann, 1933), with the modification that the first trial in a session was always a signal-absent trial and, to end in a positive way, the last trial was always either a signal-absent trial, or a detected signal, so the animal always received a reward after the last trial, unless that trial ended in a prestimulus response. Each session one of six different data collection sheets was used, each with a different Gellermann series. Sessions with more than 15% prestimulus responses were eliminated, because these usually coincided with restless behavior of the animal (i.e., swimming circles in the pool between trials and/or coming late to the start and response buoy). To avoid unintentional cueing, the trainer did not know before or during a trial whether a signal was present or absent (double blind presentation). When the dolphin left the station, the operator observed the animal's behavior on two monitors in the cabin [Fig. 1(b)], told the trainer whether or not to reward the dolphin, and recorded the animal's responses.

The order in which the frequencies were tested was randomized so that the effects of potential learning did not covary with test frequencies. The number of sessions per frequency was determined by the period that the animal and the pool were available for the study.

A switch in the dolphin's response from a detected signal amplitude (a hit) to a successive nondetected signal amplitude (a miss), and *vice versa*, is called a reversal. Amplitudes at which the animal reversed its response were taken as data points. On average 6 reversals per session were obtained (range 2–9). The mean 50% detection threshold was defined as the mean amplitude of all reversals obtained during eight sessions per frequency after the threshold had stabilized (i.e., until the mean session thresholds stopped descending), which usually occurred after two to three sessions per frequency.

Data were collected between September 2000 and April 2001. Generally, one session was conducted daily (5 days/week) between 1000 and 1100 h or between 1300 and 1400 h. No differences in average hearing thresholds were found between the morning and afternoon sessions. In total 1795 trials (12 frequencies $\times$ 8 sessions/frequency $\times$ on average 18 trials/session + 67 for the 32-kHz signals with the LF transducer) were analyzed.

III. RESULTS

The 50% detection thresholds for the 12 narrow-band frequency-modulated signals of the striped dolphin are listed in Table I. The resulting audiogram for this animal was U-shaped (Fig. 3), with hearing capabilities from 0.5 to 160 kHz ($8\frac{1}{3}$ octaves). Maximum sensitivity (42 dB *re* 1 μPa) occurred at 64 kHz (for details see Table I). The range of most sensitive hearing (defined as the frequency range with sensitivities within 10 dB of maximum sensitivity) was from 29 to 123 kHz (approximately 2 oct). The animal's hearing became less sensitive below 32 kHz and above 120 kHz. Sensitivity decreased by about 8 dB per octave below 1 kHz and fell sharply at a rate of about 390 dB per octave above 140 kHz. The average prestimulus response rate per frequency varied between 1% and 7% for all trials, and averaged 4% overall (Table I).

TABLE I. The average underwater 50% detection thresholds of a female striped dolphin for 12 narrow-band FM signals, session threshold range, number of reversals, number of sessions and prestimulus response rate over all (signal present and signal absent) trials.

Center frequency (kHz)	FM range (kHz)	Transducer	Mean 50% detection threshold (dB *re* 1 μPa)	Session threshold range (dB *re* 1 μPa)	No. of sessions (n)	Total no. of reversals	Prestimulus response rate (%)
0.5	0.495–0.505	LF	121	119–124	8	34	7
1	0.99–1.01	LF	113	112–116	8	31	3
2	1.98–2.02	LF	102	101–105	9	53	1
4	3.96–4.04	LF	93	88–98	8	43	2
8	7.92–8.08	LF	73	69–76	8	30	1
16	15.84–16.16	LF	66	63–71	8	36	4
32[a]	31.68–32.32	LF&HF	48	44–53	12	68	5
40	39.6–40.4	HF	44	40–46	10	55	6
64	63.36–64.64	HF	42	35–45	8	43	5
120	118.8–121.2	HF	50	45–54	8	45	4
140	138.6–141.4	HF	66	61–69	8	41	5
160	158.4–161.6	HF	116	116	8	50	5

[a]Based on four session thresholds obtained with the LF transducer and eight session thresholds with the HF transducer.

The 32-kHz average (over four sessions) threshold measured with the LF transducer was 49.2 dB, while the average (over eight sessions) threshold for the same frequency determined with the high frequency transducer was 47.2 dB. The mean thresholds were statistically similar (two-sample T-Test; $T=-0.93$, degrees of freedom$=4$, $P=0.403$). Therefore all 12 session thresholds were used in the calculation of average 50% detection threshold for the 32-kHz signal. The match between the 32-kHz thresholds obtained with the two transducers suggests that the shape of the audiogram is not influenced by differences in transducer characteristics. After the initial two to three sessions with each frequency, which were not analyzed, the animal's hearing sensitivity for each test frequency showed no ascending or descending trend over the 7-month study period.

IV. DISCUSSION AND CONCLUSIONS

A. Evaluation of the data

FM signals were used rather than pure tones. The advantage of using narrow-band FM signals is the reduction of propagation effects (multipath interferences) on the signals reaching the animal. However, narrow-band FM signals probably have a slightly higher arousal effect than pure tones, causing probably slightly lower thresholds. The use of narrow-band FM signals is therefore a trade-off: it provides a relatively stable sound pressure level (SPL) at the animal's head, but reduces comparability with previous studies on odontocete hearing.

During the first half of the present study, a similar audiogram study was conducted with a male harbor porpoise (*Phocoena phocoena*), under virtually matched conditions (i.e., the same equipment, the same pool, and, aside from slightly shorter delay times, the same methodology; Kastelein *et al.*, 2002). The porpoise's hearing thresholds between 0.5 and 120 kHz were 11 to 33 dB lower than those of the striped dolphin in the present study, except at 64 kHz when the difference was only 2 dB. In addition to the multiple calibrations, this information helps to confirm that the relatively low hearing sensitivity below 32 kHz of the striped dolphin in the present study is a property of the animal, and is not due to the equipment or methodology.

The animal in the present study had comparatively low prestimulus response rates, which is typical for marine mammals (Schusterman, 1974; Sauerland and Dehnhardt, 1998), although the prestimulus response rate of an animal in a psychophysical experiment is influenced by, among other factors, the way the animal is trained and the signal-present/signal-absent ratio. The low prestimulus response rate in the present study indicates that the 50% criterion was a valid choice. Normally low prestimulus response rates indicate conservative strategy of the subjects. However, in the present study, the dolphin probably only indicated the presence of a signal when it was very confident of perceiving one, since it took a long time to reach its decision. The dolphin reacted much more slowly than the harbor porpoise that was tested under matched conditions. Although it was not measured accurately, the striped dolphin waited about 500 ms before moving away from the station. This led to a much lower false response rate than the harbor porpoise, which usually responded within 80 ms (Kastelein *et al.*, 2002). The prestimulus response rate during the audiogram study with a Pacific white-sided dolphin (*Lagenorhynchus obliquidens*) which yielded a similar hearing sensitivity as the striped dolphin in the present study, was also low (Tremel *et al.*, 1998).

B. Comparison with hearing studies on other odontocetes

Underwater hearing thresholds have been determined in psychophysical tests for ten other odontocete species: the Atlantic bottlenose dolphin, *Tursiops truncatus* (Johnson, 1967, 1968; Ljungblad *et al.*, 1982; Fig. 3), harbor porpoise, *Phocoena phocoena* (Andersen, 1970; Kastelein *et al.*, 2002; Fig. 3), killer whale, *Orcinus orca* (Hall and Johnson, 1971; Szymanski *et al.*, 1999), Amazon river dolphin, *Inia geoffrensis* (Jacobs and Hall, 1972), beluga whale, *Delphinapterus leucas* (White *et al.*, 1978; Awbrey *et al.*, 1988;

Johnson, 1992; Klishin *et al.*, 2000), false killer whale, *Pseudorca crassidens* (Thomas *et al.*, 1988), baji (Chinese river dolphin), *Lipotes vexillifer* (Wang *et al.*, 1992), tucuxi, *Sotalia fluviatilis guianensis* (Sauerland and Dehnhardt, 1998), Risso's dolphin, *Grampus griseus* (Nachtigall *et al.*, 1995), and Pacific white-sided dolphin (Tremel *et al.*, 1998). Comparing the present study to other odontocete hearing studies is sometimes difficult. The absolute accuracy of the studies sometimes cannot be established due to a lack of information on the calibration methodology, threshold calculation, and variation in the thresholds. In addition, most studies have used different methodology and stimuli parameters such as signal type (pure tone versus FM) and signal duration. Also, the response effort of animal during a signal present trial probably influences the threshold level. In some studies the animals only needed to press a paddle next to them, while in the present study the dolphin had to swim about 4 m to the response buoy. The latter situation has probably led to a conservative strategy of the animal. Also the signal-present/signal-absent ratio influences the threshold level. Like the present study, most studies have been based on only one animal, and thus it cannot be established how representative the individual was for the species. In a species of which many animals have been tested, the bottlenose dolphin, age, sex, and prior environment seem to contribute to the differences in hearing sensitivity between individual animals (Ridgway and Carder, 1997).

Despite these comparison limitations, the hearing sensitivity of the striped dolphin in the present study was within the range of that of other odontocetes for frequencies above 32 kHz. However, below 32 kHz the study animal's hearing was less acute than that of the other odontocetes, but corresponded most closely to the hearing of a Pacific white-sided dolphin (Tremel *et al.*, 1998).

C. Comparison between the species' frequency range of best hearing and its phonation

A correspondence between a species phonation and hearing frequency range is a common characteristic of many mammal species. A precise match between the peak frequency of sonar signals and area of best hearing is found in the harbor porpoise (Kastelein *et al.*, 2002) and the greater horseshoe bat (*Rhinolophus ferrumequinum*; Long, 1977). These two mammalian species produce narrow-band echolocation signals. Also, in the bottlenose dolphin (Johnson, 1967) and the false killer whale (Thomas *et al.*, 1988) a match is found, but less strict. In these species, the peak frequency of echolocation signals is dependent on the amplitude of the signal (Au *et al.*, 1995), and the bandwidth of the signals is wider than in the harbor porpoise and the greater horseshoe bat. Animals may also adapt their signal spectrum to the ambient noise spectrum as was shown for false killer whales (Nester, 1999). Although very little is known about the acoustic capabilities of the striped dolphin, it probably produces wide band echolocation signals. The range of greatest hearing sensitivity found in the present study (29 to 123 kHz) overlaps the peak frequency (40 kHz) of echolocation signals of the study animal while in a pool (Dick de Haan, personal observation). The click frequencies, recorded from wild striped dolphins in the Mediterranean range from 0.3 kHz to over 100 kHz with a click repetition rate of 900 clicks/s (Zanardelli *et al.*, 1990).

A few social calls (whistles between 8 and 16 kHz) and sonar clicks of this species were presented by Schevill and Watkins (1962). The fundamentals of the whistles of a striped dolphin in distress from the Mediterranean lasted on average 0.9 s and had most energy between 8 and 12.5 kHz, with first harmonics >24 kHz (Busnel *et al.*, 1968). The fundamentals of whistles recorded from wild striped dolphins in the Mediterranean range from 3.5 to 28.5 kHz, with an average frequency emphasis near 10 kHz. Maximum recorded source level was 170 dB *re* 1 μPa (Zanardelli *et al.*, 1990). The increased hearing sensitivity between 10 and 32 kHz, found in the animal in the present study, suggests that striped dolphins, in addition to the fundamental frequency, also hear some of the harmonics of their own whistles.

D. Significance in reducing bycatch and suggestions for future research

Assuming that the hearing of the animal in the present study is representative of that of its species, the present study suggests that an acoustic alarm to prevent bycatch of striped dolphins should have most energy between 32 and 120 kHz, the frequency range of greatest hearing sensitivity of this species. Such signals are readily audible to striped dolphins even when their amplitude is low.

To be able to estimate the distance at which striped dolphins can hear each other, acoustic alarms on fishing nets, or vessels under varying ambient noise conditions, additional information is needed on how they hear in the presence of masking noise (critical ratios, critical bands), how they hear sounds of different duration, and how well they spatially resolve sounds coming from different directions (directivity index).

ACKNOWLEDGMENTS

We thank Jolanda Meerbeek, Lauro Marcenaro and Joyce Borrias for their help with training and data collection, Yvonne Vrugtman and Paulien Bunskoek for assisting with the data collection, and Rob Triesscheijn for making the graphs. We thank Gianni Pavan (Center of Bioacoustics, Pavia University, Italy) for supplying information on striped dolphin whistles. We also thank Jen Philips (Hawaii Institute of Marine Biology, USA), Jeanette Thomas (Western Illinois University, USA), William Watkins (Woods Hole Oceanographic Institution, USA), Alexander Supin (Institute of Ecology and Evolution, Moscow, Russia), and Nancy Vaughan (University of Bristol, UK) for their valuable constructive comments on this manuscript. This study was funded by the Harderwijk Marine Mammal Park, The Netherlands, and The North Sea Directorate (DNZ; through Wanda Zevenboom and Marcel Bommelé, Contract No. DNZ: 76/319471; 7154 kd 704020, 2001) of the Directorate-General of the Netherlands Ministry of Transport, Public Works and Water Management (RWS). The dolphin's training and testing were conducted under authorization of the

Netherlands Ministry of Agriculture, Nature Management and Fisheries, Department of Nature Management. Endangered Species Permit No. FEF27 06/2/98/0184.

Aguilar, A. (**2000**). ''Population biology, conservation threats and status of Mediterranean striped dolphins (*Stenella coeruleoalba*),'' J. Cetacean Res. Manage. **2**, 17–26.

Andersen, S. (**1970**). ''Auditory sensitivity of the harbour porpoise, *Phocoena phocoena*,'' in *Investigations on Cetacea*, edited by G. Pilleri (Institute for Brain Research, Bern), Vol. 2, pp. 255–259.

Anonymous 98, 51–59 (**2000**). ''Annex I Report of the Sub-Committee on Small Cetaceans,'' J. Cetacean Res. Manage. **2**(Suppl.), 235–243.

Au, W. L., and Moore, P. W. B. (**1984**). ''Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin *Tursiops truncatus*,'' J. Acoust. Soc. Am. **75**, 255–262.

Au, W. W. L., Pawloski, J., Nachtigall, P. E., Blonz, M., and Gisner, R. (**1995**). ''Echolocation signal and transmission beam pattern of a false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **98**, 51–59.

Awbrey, F. T., Thomas, J. A., and Kastelein, R. A. (**1988**). ''Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*,'' J. Acoust. Soc. Am. **84**, 2273–2275.

Busnel, R.-G., Pilleri, G., and Fraser, F. C. (**1968**). ''Notes concernant le dauphin *Stenella styx* Gray 1864,'' Mammalia **32**, 192–203.

Di-Méglio, N., Romero-Alvarez, R., and Collèt, A. (**1996**). ''Growth comparison in striped dolphins, *Stenella coeruleoalba*, from the Atlantic and Mediterranean coasts of France,'' Aquat. Mammals **22**, 11–21.

Gellerman, L. W. (**1933**). ''Chance orders of alternating stimuli in visual discrimination experiments,'' J. Gen. Psychol. **42**, 206–208.

Hall, J. D., and Johnson, C. S. (**1971**). ''Auditory thresholds of a killer whale,'' J. Acoust. Soc. Am. **51**, 515–517.

Jacobs, D. W., and Hall, J. D. (**1972**). ''Auditory thresholds of a fresh water dolphin, *Inia geoffrensis* Blainville,'' J. Acoust. Soc. Am. **51**, 530–533.

Johnson, S. C. (**1967**). ''Sound detection thresholds in marine mammals,'' in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. 2, pp. 247–260.

Johnson, S. C. (**1968**). ''Masked tonal thresholds in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **44**, 965–967.

Johnson, S. C. (**1992**). ''Detection of tone glides by the beluga whale.'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Y. Supin (Plenum, New York), pp. 241–247.

Kastelein, R. A., Bunskoek, P., Hagedoorn, M., Au, W. W. L., and de Haan, D. (**2002**). ''Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals,'' J. Acoust. Soc. Am. **112**, 334–344.

Klishin, V. O., Popov, V. V., and Supin, A. Y. (**2000**). ''Hearing capabilities of a beluga whale, *Delphinapterus leucas*,'' Aquat. Mammals **26**, 212–228.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Ljungblad, D. K., Scoggins, P. D., and Gilmartin, W. G. (**1982**). ''Auditory thresholds of a Captive eastern Pacific bottlenosed dolphin, *Tursiops* spp.,'' J. Acoust. Soc. Am. **72**, 1726–1729.

Long, G. R. (**1977**). ''Masked Auditory Thresholds from the Bat, *Rhinolophus ferrumequinum*,'' J. Comp. Physiol. **116**, 247–255.

Nachtigall, P. E., Au, W. W. L., Pawloski, J. L., and Moore, P. W. B. (**1995**). ''Risso's dolphin (*Grampus griseus*) hearing thresholds in Kaneohe Bay, Hawaii,'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden), pp. 49–53.

Nester, A. (**1999**). ''The underwater sound repertoire of wild false killer whales,'' MS thesis, Western Illinois University.

Perrin, W. F., Wilson, C. E., and Archer II, F. I. (**1994**). ''Striped dolphin, *Stenella coeruleoalba* (Meyen, 1833),'' in *Handbook of Marine Mammals, Volume 5, The First Book of Dolphins*, edited by S. H. Ridgway and R. Harrison (Academic, San Diego), pp. 129–159.

Ridgway, S. H., and Carder, D. A. (**1997**). ''Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin,'' J. Acoust. Soc. Am. **101**, 590–594.

Robinson, D. E., and Watkins, C. S. (**1973**). ''Psychophysical methods in modern Psychoacoustics,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 2, pp. 99–131.

Sauerland, M., and Dehnhardt, D. (**1998**). ''Underwater audiogram of a *Tucuxi* (*Sotalia fluviatilis guianensis*),'' J. Acoust. Soc. Am. **103**, 1199–1204.

Schevill, W. E., and Watkins, W. A. (**1962**). ''Whale and porpoise voices. A phonograph Record,'' Woods Hole Oceanogranic Institute, Woods Hole, MA.

Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002**). ''The effect of projector position on the hearing thresholds of dolphins (*Tursiops truncatus*) at 2, 8 and 12 kHz,'' in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (Univ. of Chicago, Chicago).

Schusterman, R. J. (**1974**). ''Low false-alarm rates in signal detection by marine mammals,'' J. Acoust. Soc. Am. **55**, 845–848.

Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (**1999**). ''Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms,'' J. Acoust. Soc. Am. **84**, 936–940.

Thomas, J. A., Chun, N., Au, W. W. L., and Pugh, K. (**1988**). ''Underwater audiogram of a false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **84**, 936–940.

Tremel, D. P., Thomas, J. A., Ramirez, K. T., Dye, G. S., Bachman, W. A., Orban, A. N., and Grimm, K. K. (**1998**). ''Underwater hearing sensitivity of a Pacific white-sided dolphin, *Lagenorhynchus obliquidens*,'' Aquat. Mammals **24**, 63–69.

Wang, D., Wang, K., Xiao, Y., and Sheng, G. (**1992**). ''Auditory sensitivity of a Chinese river dolphin, *Lipotes vexillifer*,'' in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Y. Supin (Plenum, New York), pp. 213–221.

White, Jr., M. J., Norris, J., Ljungblad, D. K., Baron K., and di Sciara, G. (**1978**). ''Auditory thresholds of two beluga whales (*Delphinapterus leucas*),'' HSWRI Tech. Rep. No. 78-109, Hubbs Marine Research Institute, San Diego, CA.

Zanardelli, M., Notarbartello de Sciara, G., and Pavan, G. (**1990**). ''Characteristics of underwater acoustic signals produced by the striped dolphin, *Stenella coeruleoalba*, in the central Mediterranean Sea,'' Abstracts of the Fourth Annual Conference of the European Cetacean Society, Palma de Mallorca, 2–4 March 1990, p. 69.

Discrimination of complex synthetic echoes by an echolocating bottlenose dolphin

David A. Helweg[a] and Patrick W. Moore
Space and Naval Warfare Systems Center, 53560 Hull Street, San Diego, California 92152

Lois A. Dankiewicz and Justine M. Zafran
Science Applications International Corp., 3990 Old Town Avenue, Suite 208A, San Diego, California 92110

Randall L. Brill
Space and Naval Warfare Systems Center, 53560 Hull Street, San Diego, California 92152

(Received 31 January 2002; revised 28 September 2002; accepted 25 October 2002)

Bottlenose dolphins (*Tursiops truncatus*) detect and discriminate underwater objects by interrogating the environment with their native echolocation capabilities. Study of dolphins' ability to detect complex (multihighlight) signals in noise suggest echolocation object detection using an approximate 265-μs energy integration time window sensitive to the echo region of highest energy or containing the highlight with highest energy. Backscatter from many real objects contains multiple highlights, distributed over multiple integration windows and with varying amplitude relationships. This study used synthetic echoes with complex highlight structures to test whether high-amplitude initial highlights would interfere with discrimination of low-amplitude trailing highlights. A dolphin was trained to discriminate two-highlight synthetic echoes using differences in the center frequencies of the second highlights. The energy ratio (ΔdB) and the timing relationship (ΔT) between the first and second highlights were manipulated. An iso-sensitivity function was derived using a factorial design testing ΔdB at -10, -15, -20, and -25 dB and ΔT at 10, 20, 40, and 80 μs. The results suggest that the animal processed multiple echo highlights as separable analyzable features in the discrimination task, perhaps perceived through differences in spectral rippling across the duration of the echoes. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1531175]

PACS numbers: 43.80.Lb, 43.66.Gf [WA]

I. INTRODUCTION

Bottlenose dolphins (*Tursiops truncatus*) detect and discriminate underwater objects by interrogating their environment with their native echolocation capabilities. *Tursiops* echolocation signals are clicks approximately 50–100 μs in duration, with peak frequencies typically ranging between 30–100 kHz and fractional bandwidths between 10%–90% of peak frequency (Au, 1980; Houser *et al.*, 1999). Although the outgoing echolocation signals are brief, echoes reflected from objects can be several milliseconds in duration and contain rich structure that encodes information about the object's shape, orientation, and internal composition (e.g., Chapman, 1971; Gaunaurd *et al.*, 1998; Neubauer, 1986; Urick, 1983). The diversity of complex time and frequency-domain structures includes great variability in the amplitude ratio of multiple echo components, called ''highlights'' or ''glints.'' The variance in echo structures between objects, and within objects in aspect-dependent shapes, immediately raises questions of how dolphins exploit the complex timing and relative amplitude of highlight structure to detect and identify objects.

Study of the dolphins' ability to detect multihighlight signals in noise has revealed a temporal integration time of approximately 265 μs (Au *et al.*, 1988; Moore *et al.*, 1984; Vel'min and Dubrovskiy, 1976). The energy of echo highlights appears to be summed within this window and contributes to signal detection, whereas stimulus highlights separated by more than this interval do not contribute to detection performance. Dolphins appeared to detect echoes using a 265-μs window sensitive to the echo region or highlight of highest energy, and low-amplitude echo highlights spaced more than a few hundred microseconds apart did not contribute to detection performance (Au *et al.*, 1988).

However, many large objects with complex structures generate echoes with highlight structure spaced over several milliseconds (e.g., Chapman, 1971; Gaunaurd *et al.*, 1998; Neubauer, 1986; Urick, 1983). For multiple highlights that fall within a single integration window, spectral models can describe discrimination performance. For example, Johnson and colleagues (1988) demonstrated that a dolphin could discriminate a signal with a high-amplitude followed by a low-amplitude highlight from one consisting of a low-amplitude followed by a high-amplitude highlight, even when both highlights appeared within the same putative integration window. Au and Pawloski (1992) demonstrated that a dolphin could discriminate metal cylinders with differences in the wall thickness. Inspection of cylinder echoes revealed multiple highlights within a single integration window, with interhighlight intervals proportional to wall thickness (in tens of μs). These studies indicate that the animal was not simply integrating over the integration window. Instead, spectral

[a]Electronic mail: david_helweg@usgs.gov

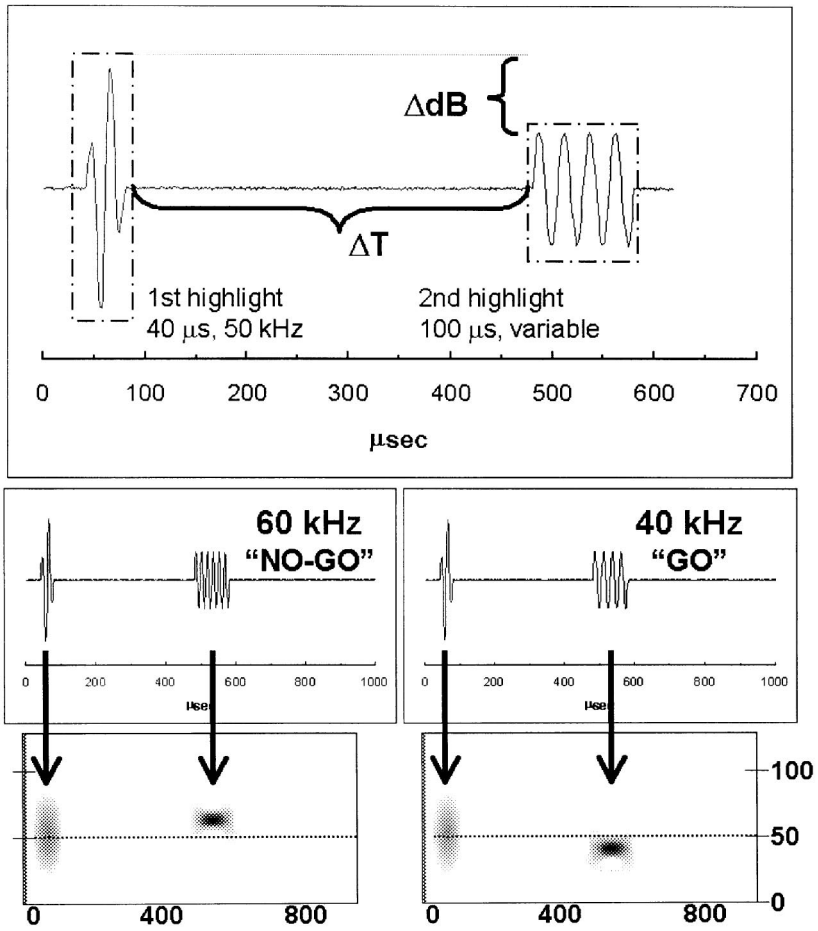


FIG. 1. The synthetic echo stimuli. The top panel illustrates the relationship between the stimulus waveform components and independent variables (ΔdB, ΔT). In this figure, ΔT is 400 μs and ΔdB (energy flux) is zero. The center panels show the ''NO-GO'' waveform on the left and the ''GO'' waveform on the right. They differ only in the frequency of the second highlight, which was 60 kHz for the NO-GO stimulus and 40 kHz for the GO stimulus. The bottom panels are Gabor spectrograms of the stimuli, with frequency on the vertical axis and time aligned with the waveforms.

characteristics generated by the amplitude and timing of multiple highlights were possible acoustic features that may have controlled the dolphin's performance. Johnson *et al.* (1988) demonstrated that the temporal order of click pairs could be discriminated by relative timing of spectral rippling, which was revealed using short-time Fourier transform of the signals. Similarly, Au and Pawloski (1992) suggested that the cylinders of different wall thickness could be discriminated based on differences in spectral rippling within the temporal integration time. Likewise, Moore *et al.* (1984) conducted a backward-masking experiment to replicate the work of Vel'min and Dubrovskiy (1976), and reported results which appeared to support the notion of the critical interval. They suggested, however, that time separation pitch (TSP) might be the underlying mechanism instead of a ''critical interval'' in dolphin hearing. Thus, for multiple highlights that fall within a single integration window, spectral models can describe discrimination performance.

In contrast to within-265-μs mechanisms, the work by Au *et al.* (1988) raises the question of the degree to which dolphin auditory processes are sensitive to information contained in low-amplitude highlights that lie in different temporal integration windows. We investigated this question using synthetic echo stimuli and a computerized echo generator. The use of synthetic echoes allowed absolute experimental control over the amplitude, timing, and spectral relationships among multiple highlights within the synthetic echoes. The work by Au *et al.* (1988) suggests that dolphins may not attend to trailing highlights more than 6 dB below a larger highlight if the time separation is more than about 265 μs. Thus, we tested the dolphin's ability to discriminate two-highlight stimuli differing in the spectra of the trailing highlight, while manipulating the time separation and amplitude ratio of the two highlights.

II. METHODS

A. Subject

The subject was CAS, a 16-year-old female Atlantic bottlenose dolphin housed with several other dolphins in a floating pen complex at the Space and Naval Warfare Systems Center facility in San Diego Bay. CAS had over 5 years of experience as a pyschoacoustical research subject coming into the current study. Based on routine assessments, her hearing was considered normal (Brill *et al.*, 2001).

B. Synthetic echo stimuli

A pair of ''synthetic echoes'' was designed to test the research hypotheses. Sample waveforms and Gabor spectrograms are presented in Fig. 1. The waveforms consisted of two highlights. The initial highlight of both stimuli was a 40-μs 50-kHz sinusoid passed through a triangular window. The second highlight was 100 μs in duration, at 60 kHz for the ''NO-GO'' stimulus and 40 kHz for the ''GO'' stimulus (''NO-GO'' and ''GO'' are behavioral response categories

and are described below). The 20-kHz difference in frequency was substantial compared with frequency limens reported for bottlenose dolphins in a wide range of paradigms (Jacobs, 1972; Thompson and Herman, 1975); thus, the stimuli were discriminable based on the frequency of the second highlight alone. To control for the effects of ambient noise and to provide a uniform noise background across the frequency range of the test stimuli, the noise floor was controlled by adding 95 dB *re*: 1 Vrms of white noise to the stimuli.

Two variables were manipulated. One, manipulation of the energy flux ratio of the second to the first highlight, permitted evaluation of discrimination performance as the ratio of the two stimulus highlights increased. The relative energy ratio was termed ''ΔdB,'' use of energy flux was based on the assumption that dolphin echo detection is energy based rather than pressure based (Au *et al.*, 1988). The amplitude of the first highlight (50 kHz) was held constant at 135 dB *re*: 1 Vrms. The amplitude of the second highlight (40 or 60 kHz) was manipulated to create the specified ΔdB. A ΔdB of zero meant that the energy flux of the initial highlight was equal to the energy flux of the second highlight. As the amplitude of the second highlight was experimentally decreased, the ΔdB value became more negative. Thus, a stimulus with a ΔdB of −10 dB would have a higher-amplitude second highlight than a stimulus with ΔdB of −20 dB. The ΔdB of the NO-GO and GO stimuli were equated; thus, any change made to the NO-GO stimulus also was applied to the GO stimulus and vice versa. This eliminated energy cues that may have confounded the dolphin's second-highlight frequency discrimination performance if the highlights were summed. Again, note that the ΔdB refers to the ratio of the energy flux of the first and second highlights *within each synthetic echo*, not an amplitude relationship between the GO and NO-GO stimuli.

The second variable that was manipulated was the timing relationship between the initial and second highlights (ΔT). Manipulation of ΔT permitted evaluation of discrimination performance around the 265-μs temporal integration time (Moore *et al.*, 1984; Vel'min and Dubrovskiy, 1976). ΔT ranged from 10 to 400 μs. The initial highlight was 40 μs in duration, and the second highlight was 100 μs in duration. Thus, both highlights were inside the 265-μs temporal energy integration window when ΔT was set to $\leqslant$125 μs. Any ΔT change made to the NO-GO stimulus also was applied to the GO stimulus and visa-versa.

C. Apparatus

1. Digital synthetic echo system

A synthetic echo system (SES) was constructed to detect outgoing echolocation clicks and transmit a single stimulus waveform per detected click. The SES, graphic user interface, data collection parameters, and trial scheduling information were controlled by a LABVIEW Virtual Instrument running a National Instruments PCI MIO-16E-1 multifunction board hosted on a Pentium PC. The digital synthetic echo was generated prior to the start of each trial, mixed with white noise, and stored in RAM. Information available to the dolphin was held constant by permitting only 20 synthetic echoes per trial, regardless of how many clicks the dolphin emitted.

CAS was trained to position her head in a hoop 1 meter below the surface. An acoustically opaque screen (sheet PVC covered with closed-cell neoprene) was placed between the dolphin and the echo projector, which prevented CAS from echolocating the apparatus until the screen was removed. At the start of a trial, the screen was raised. Outgoing echolocation clicks were detected using a Reson TC4013 omnidirectional broadband hydrophone placed 0.5 m from the dolphin's melon. The click channel was bandpass filtered from 16–200 kHz with 40 dB of gain by a DL Electronics 4302 filter/amplifier and cabled to the analog input of the MIO board. When the click exceeded 170 dB *re*: 1 μPa, a digital trigger was sent to the SES software. The trigger generated analog output of a single synthetic echo stored in RAM on board the MIO board. Thus, one echo was projected per click emitted by the dolphin. A target range of 14 m was simulated using a delay of 18 ms between reception of an echolocation click trigger and analog output of the synthetic echo. The synthetic echo was bandpass filtered from 20–100 kHz with 40 dB of gain by a DL Electronics 4302 filter/amplifier and projected to the dolphin with an International Transducer Corporation 5446 transducer located 1.4 m from the dolphin. The digital waveforms were matched to the transmit response of the ITC 5446. Multipath echoes were prevented from reaching the dolphin using a floating horsehair mat placed just below the water surface at the surface reflection point. Prior to data collection, the system was calibrated by projecting synthetic dolphin clicks through the ITC 5446 and measuring received synthetic echoes with a calibrated ITC 6030 omnidirectional hydrophone mounted in the dolphin's stationing hoop.

D. Threshold estimation methodology

Data were collected using two methods. In phase one, the ΔdB threshold was measured using an up–down staircase method of threshold titration similar to that used by Moore and Schusterman (1987). For phase two, ΔdB was held constant at 75%-correct level, and the boundaries of ΔT were measured using a modified method of constants (Green and Swets, 1966). Finally, in phase three ΔdB and ΔT were jointly manipulated in a 4×4 factorial design using the modified method of constants.

1. Titration paradigm (phase one)

A standard titration method (Green and Swets, 1966) was used to evaluate the ΔdB threshold—the largest ΔdB that the dolphin would tolerate. The amplitude of the initial highlight was held constant at 135 dB *re*: 1 μPa. At the start of each session, ΔdB was set well above the subject's previous threshold (ΔdB was proportional to the energy in the second highlight; thus, more positive values of ΔdB resulted in higher second-highlight amplitudes). After every correct response the ΔdB was decreased by 2 dB, thereby driving the amplitude of the second highlight down (recall that any given ΔdB setting was applied to both ''NO-GO'' and ''GO'' stimuli). Once the dolphin made an error, the first reversal

was said to have occurred and the ΔdB was increased by 1 dB. ΔdB were increased in 1-dB steps until the dolphin produced a correct response, the second reversal. The ΔdB were then decreased in 1-dB steps until she produced another error, the third reversal. The session was continued until ten reversals were elicited. The ΔdB threshold was estimated as the average of the values at the ten reversals; thus, each session yielded one threshold estimate. After five training sessions, ΔdB titration sessions were conducted until thresholds within 3 dB were reached on two successive sessions.

2. Method of constants paradigm

Phase two and three testing was accomplished using the method of constant stimuli (Green and Swets, 1966). Each session consisted of a block of ten warm-up trials, followed by four ten-trial test blocks. When practicable, sessions also were terminated with a set of cool-down trials.

First, ΔT was manipulated while holding ΔdB constant at the 75%-correct choice level from the phase one data. This value was selected to allow CAS to demonstrate either increased or decreased choice performance as ΔT was manipulated, while providing a ΔdB level that would assure a good rate of reinforcement. A running estimate of percent correct was calculated for each session using a ten-trial sliding window, and the 75%-correct point(s) were tabulated. The median and semi-interquartile range were derived (Blalock, 1979), and ΔdB was set to the third quartile of the pooled 75%-correct choice data. A set of six ΔT values was tested per session. The dolphin's performance was measured as percent correct for each combination of ΔdB and ΔT.

In the last phase of testing, ΔT and ΔdB were manipulated in a factorial design using ranges for ΔT and ΔdB determined in the first two phases. With $4\Delta T \times 4$ ΔdB levels in the factorial design matrix, and four ten-trial blocks of data per session, four sessions were required to generate one ten-trial block for each level in the 4×4 matrix. Order was counterbalanced across the four sessions. Thus, 28 sessions were run in order to collect seven ten-trial blocks of data for each level. The values of d' were calculated for each ten-trial block, the minimum and maximum values discarded, and an average d' and β were calculated for the pooled 50 trials that remained.

The results of the factorial experiment were analyzed using signal detection parameters d' and β (Green and Swets, 1966), adjusted using an unequal variance model (Hautus, 1995). The receiver sensitivity metric d' is zero at chance performance, i.e., 50%-correct choice in this two-alternative task. To account for unequal variance in responding, threshold was estimated at d' of 1.0 (Green and Swets, 1966). A value of zero for the natural log of the receiver response bias metric β [ln(β), henceforth β] indicates unbiased responding.

E. Behavioral paradigm

The data collection sessions began with CAS facing the trainer, touching her rostrum against an intertrial station (foam pad) located just above the water surface. Upon presentation of a hand cue, the dolphin would submerge and position her head in the test station hoop. The trainer removed the acoustically opaque screen as a computer operator activated the SES. A 4-s trial period followed, during which time CAS would freely echolocate, receiving up to 20 stimuli in return, and respond. A correct ''GO'' response was made if she swam out of the hoop and touched a nearby paddle. A correct ''NO-GO'' response was made if she stayed in the hoop for the 4-s trial duration. Both correct responses were reinforced by a bridging stimulus and a consistent fish reward. Data were collected using a modified Gellermann series (Gellermann, 1933) that had been counterbalanced in ten-trial blocks. Each session was initiated with a ten-trial block of warm-up trials. If CAS's performance was less than 80% correct, the session was terminated and revisited later in the day. One session was run per day.

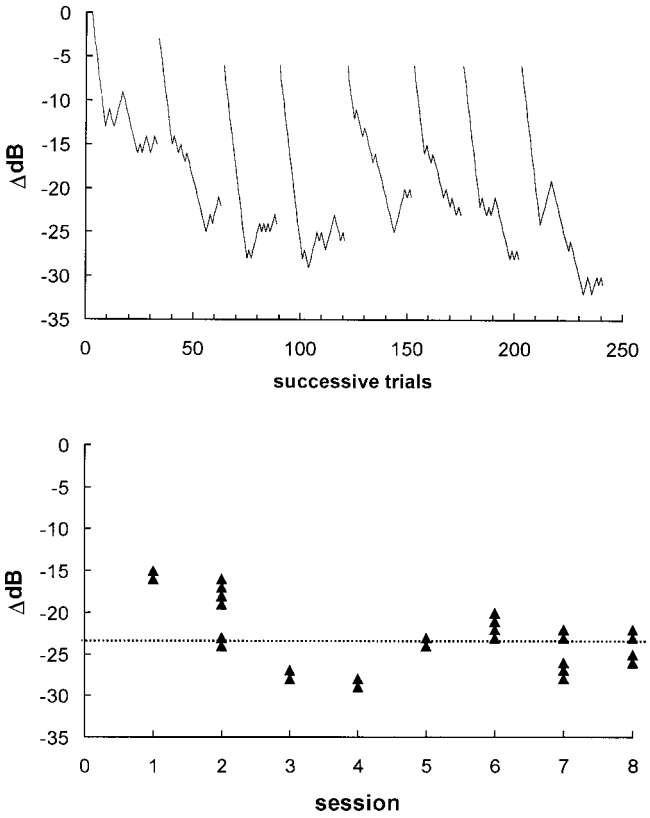


FIG. 2. Determining ΔdB threshold by titration. The top panel shows the raw titration data for each session, plotted as a function of trial number. The bottom panel shows the ΔdB values at the 75%-correct threshold for each session. The median (−22 dB) is indicated by the dotted line.

III. RESULTS

A. Assessment of ΔdB threshold (phase one)

The first phase of measurement was assessment of the ΔdB threshold. ΔT was held constant at 400 μs, which placed the two highlights in separate 265-μs integration windows. Eight ΔdB titration sessions were run and the ΔdB threshold session results are presented in Fig. 2. The top panel illustrates the ΔdB values at which the reversals occurred for each session. CAS's minimum ΔdB was −32 dB. This corresponds to a value of 96.5 dB *re*: 1 Vrms for the second highlight, approximately 1.5 dB above the white-noise floor. A sliding ten-trial window was passed over the

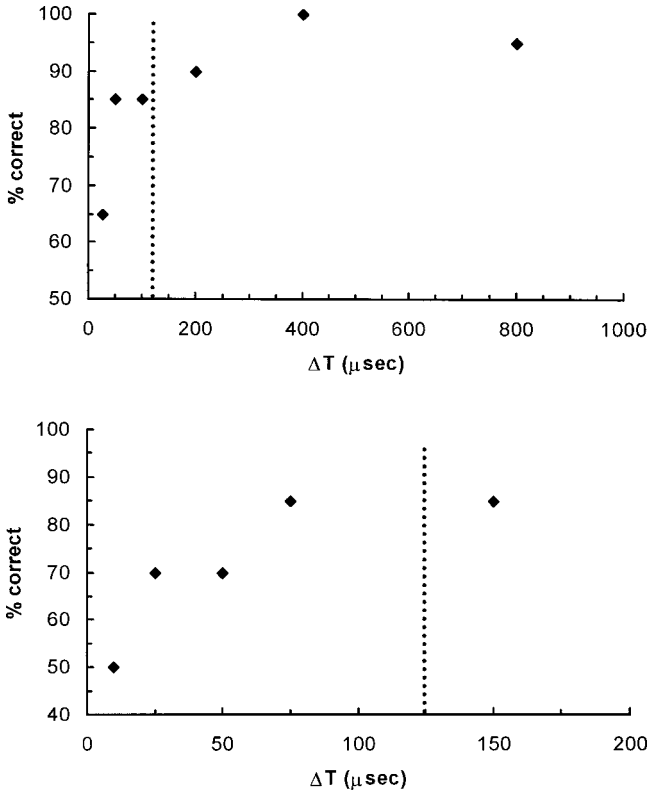


FIG. 3. Determining the limits of ΔT. The top panel summarizes CAS's performance on the first set of ΔT values ($n=20$ per value), and the bottom panel summarizes her performance on the second set ($n=20$ per value). ΔdB was held constant at -19 dB. In each panel, the dotted vertical line indicates the approximate ΔT transition from single to multiple (nonoverlapping) 265 μs temporal integration windows.

data for each session, and the ΔdB values at the 75%-correct threshold were extracted, presented in the bottom panel of Fig. 2. Overall, the median threshold was -22 dB, with a semi-interquartile range of 3 dB.

B. Assessment of ΔT boundaries (phase two)

In the second phase of measurement, we held ΔdB constant at -19 dB (third quartile), and manipulated ΔT to determine the dolphin's performance boundaries. Warm-up blocks were run with ΔT at 400 μs, and two ten-trials blocks were run for ΔT at 25, 50, 100, 200, 400, and 800 μs. The overall percentage of correct responses for each session in phase two is presented in the top panel of Fig. 3. CAS's performance at the 50-μs level was well above chance; thus, we ran a second set of blocks with the warm-up ΔT at 150 μs, and tested at 10, 25, 50, and 75 μs. The results are summarized in the bottom panel of Fig. 3. With ΔdB held constant at -19 dB, CAS's performance approached chance level as ΔT was decreased below 50 μs, but performance remained at or above 85% correct above 75 μs. Recall that ΔT less than 125 μs placed both highlights within a single separate integration window. CAS's results clearly indicate no significant decrement in performance as the highlights transitioned between separate and single critical intervals.

TABLE I. Values of d' for each combination of ΔdB and ΔT ($n=25$ per cell).

		ΔT (μsec)			
		10	20	40	80
ΔdB	-10	1.68	2.90	3.16	3.81
	-15	1.06	2.58	2.93	3.05
	-20	0.74	0.96	1.31	2.81
	-25	0.18	0.41	0.70	1.29

C. Factorial test: ΔdB vs ΔT (phase three)

The results of phases one and two provided estimates of ΔdB and ΔT that described the boundaries of CAS's discrimination performance. In the third phase, we conducted a factorial experiment to evaluate CAS's performance within these limits. For the warm-up block in each session, ΔdB was set at -19 dB and ΔT at 160 μs. ΔdB was tested at -10, -15, -20, and -25 dB. ΔT was tested at 10, 20, 40, and 80 μs. Average d' and β were calculated for the pooled 50 trials for each factorial level. Results of the factorial testing will be described using the combination of$\{\Delta\text{dB},\Delta T\}$. CAS's performance on the warm-up and cool-down trials $\{-19, 160\}$ was near perfect, with a d' of 3.2 and virtually no response bias ($\beta=0.06$). For test blocks, her response bias remained minimal and nonsystematic, with an average false-alarm probability of 0.17 and β of -0.02. The test results are presented in Table I and in graphical form in the top panel of Fig. 4. The horizontal line in the top panel of Fig. 4 indicates a d' threshold of 1.0. Sensitivity was highest

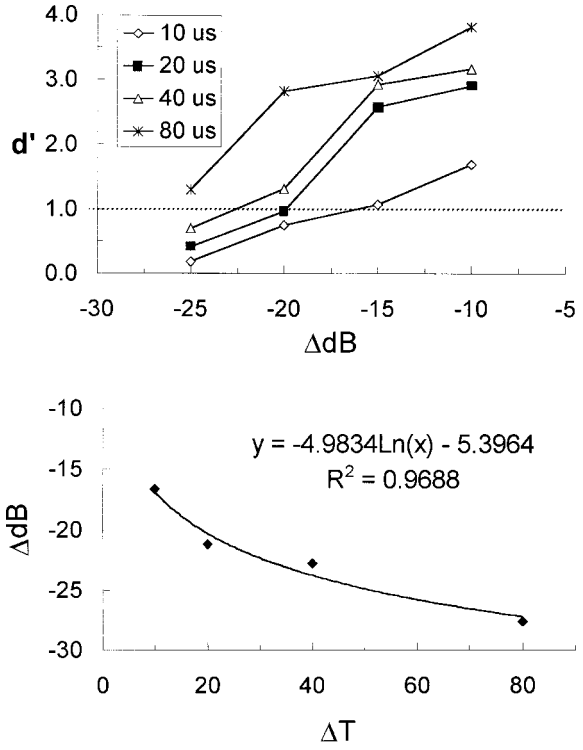


FIG. 4. Derivation of an iso-sensitivity function for $\Delta\text{dB}\times\Delta T$. The top panel shows the results of the factorial experiment in which ΔdB and ΔT were jointly manipulated. ΔdB was estimated for each ΔT curve at d' equal to 1.0. The iso-sensitivity function is presented in the bottom panel, with the best-fit exponential curve.

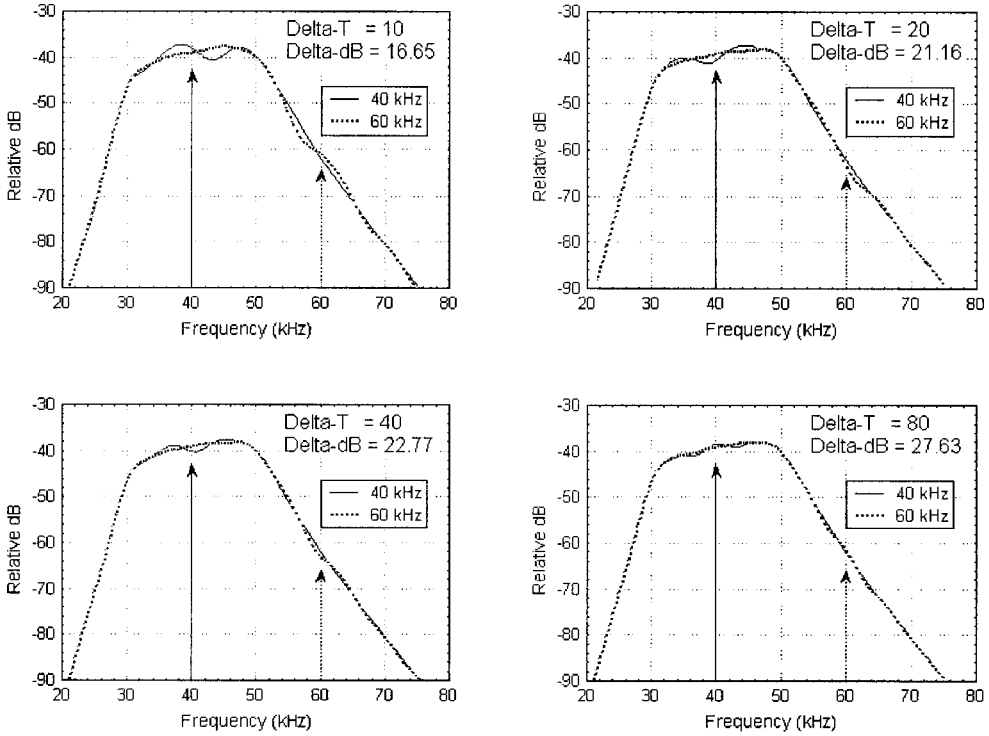


FIG. 5. Spectra of the stimuli, one for each combination of ΔdB and ΔT predicted from the iso-sensitivity function derived in the factorial experiment. Frequency resolution was 488 Hz per FFT bin. The NO-GO spectra (60-kHz second highlight) are represented by dotted lines, and the GO echoes (40-kHz second highlight) by solid lines.

at $\{-10, 80\}$ and lowest at $\{-25, 10\}$. To fuse the results into a single function, we estimated the ΔdB value at d' equal to 1.0 by linear fit to each ΔT curve. The resulting iso-sensitivity function is presented in the bottom panel of Fig. 4. The function is well-behaved, described well by a natural logarithmic function ($\Delta\text{dB} = -4.9834 \cdot \ln(\Delta T) - 5.3964, R^2 = 0.969$). The results clearly demonstrate the relationship between the energy and timing features of the synthetic echoes, with the dolphin requiring increasing separation between the first and second highlight to maintain discrimination sensitivity as the energy in the second highlight decreased.

IV. DISCUSSION

The first phase of measurement was an assessment of the ΔdB threshold. ΔT was held constant at 400 μs, with the initial stimulus component held constant at 135 dB *re*: 1 Vrms and the white-noise floor at 95 dB *re*: 1 Vrms. CAS's median threshold was −22 dB. Her maximum ΔdB was −32, which corresponds to a value of 96.5 dB for the second highlight, approximately 1.5 dB above the white-noise floor. Thus, ΔdB was limited by the white-noise floor and not by the amplitude relationship of the first and second echo highlights. This contrasts with the detection results reported by Au *et al.* (1988), which would have predicted that CAS's choice performance would decline at ΔdB of about −6 dB since the initial highlight would have "captured" the 265-μs temporal integration window, reducing attention to low-amplitude trailing highlights.

In the second phase of measurement, ΔdB was held constant at −19 dB, and ΔT was manipulated to determine the dolphin's performance boundaries. Discrimination performance approached chance level as ΔT was decreased below 50 μs, but performance remained at or above 85% correct from 75–800 μs. The results clearly indicate no significant decrement in performance as the highlights transitioned between multiple or single temporal integration intervals. These results suggest an echo-feature discrimination window that in some sense can operate independently of the energy integration detection process.

The third phase was a factorial study with ΔdB tested at −10, −15, −20, and −25 dB, and ΔT tested at 10, 20, 40, and 80 μs. No evidence of response bias was observed. Sensitivity was highest at $\{-10, 80\}$ and lowest at $\{-25, 10\}$. The data supported a well-behaved iso-sensitivity function indicating that the dolphin required increasing energy in the second highlight within each echo to maintain discrimination sensitivity as the separation between the first and second highlight decreased.

The dolphin's ability to discriminate the synthetic echoes was a function of her sensitivity to the center frequency of the second echo highlight. At 40 kHz, the frequency limens of the bottlenose dolphin auditory system is at most 1% (or about 400 Hz; see Thompson and Herman, 1975), thus the 40-versus 60-kHz discrimination was straightforward. The acoustical feature(s) of the stimuli that controlled her choice performance are unknown. Time separation pitch (Au and Pawloski, 1992) likely was not a cue, because the time separation (ΔT) between the highlights was equated for the GO and NO-GO stimulus waveforms.

The distribution of spectral energy contains differences that could have cued her responses (Au and Pawloski, 1992;

Hammer and Au, 1980; Johnson *et al.*, 1988; Moore *et al.*, 1984). Using frequency cues, a parsimonious description of CAS's decision rule is an ''A versus Not-A'' detection—that is, perform a paddle press (GO) if a 40-kHz signal is detected, otherwise remain in the hoop (NO-GO). To illustrate this concept, we applied a symmetric filter with center frequency of 40 kHz and Q of approx. 2.2 (see Au and Moore, 1990) to stimuli created using the iso-sensitivity function generated in the factorial experiment. The filtered spectra are presented in Fig. 5, depicting four combinations of ΔdB and the ΔT predicted from the natural logarithmic fit to the experimental data $[\Delta\text{dB} = -4.9834 \cdot \ln(\Delta T - 5.3964)]$. For purposes of illustration, the spectral bandwidth was set to 488 Hz, to be consistent with the frequency limens reported by Thompson and Herman (1975). Notice the spectral ripple centered around 40 kHz. This ripple is most pronounced in the $\{\Delta T = 10\ \mu\text{s},\ \Delta\text{dB} = -16.65\ \text{dB}\}$ waveform and gradually attenuates towards the average level as ΔT was increased and ΔdB was decreased. The ΔT and ΔdB values were derived from an iso-sensitivity function, however, so the ripple should have remained more constant to persist as the sole cue.

In summary, unlike the energy integration observed in the *detection* thresholds of complex stimuli (Au *et al.*, 1987; Vel'min and Dubrovskiy, 1976), it appears that in a *discrimination* task the animal may perceive the within-echo components as separable analyzable features. The dolphin's performance was high with multiple highlights both within a single integration window or distributed across several integration windows (e.g., with ΔT greater than 125 μs). Temporal smearing of features, implicit in an energy integrator, did not appear to limit discrimination performance because the dolphin was able to discriminate low-amplitude highlights in close proximity to uninformative high-amplitude highlights. Moreover, as separation between highlights increased, sensitivity to lower-amplitude highlights increased, thereby improving the likelihood that the animal could detect lower-amplitude trailing echo features, such as those generated by target resonance (Gaunaurd *et al.*, 1998). Thus, the energy integration detection mechanism does not necessarily ''lock on'' to high-amplitude features at the expense of reduced sensitivity to lower-amplitude features in trailing integration windows, as can be inferred from detection of complex echoes (Au *et al.*, 1988).

Based on the results provided here, dolphins can isolate and process brief acoustic features that lie within and between energy integration windows of the echo detection system. Such performance would permit the dolphin auditory system to attend to lower-amplitude echo features (unmasked by ambient noise) related to objects of interest while rejecting higher-amplitude features related to reverberation and clutter, an adaptive capability in the high-clutter high-reverberation littoral niche occupied by bottlenose dolphins.

ACKNOWLEDGMENTS

We thank the Office of Naval Research Code 321US, for sponsorship of Project RN15C44.

Au, W. W. L. (**1980**), ''Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncatus*) in open waters,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Au, W. W. L., and Moore, P. W. B. (**1990**). ''Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **88**, 1635–1638.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. A. (**1988**). ''Detection of complex echoes in noise by an echolocating dolphin,'' J. Acoust. Soc. Am. **83**, 662–668.

Au, W. W. L., and Pawloski, D. A. (**1992**). ''Cylinder wall thickness discrimination by an echolocating dolphin,'' J. Comp. Physiol., A **72**, 41–47.

Blalock, H. M. (**1979**). *Social Statistics, 2nd ed., rev.* (McGraw-Hill, New York).

Brill, R. L., Moore, P. W. B., and Dankiewicz, L. A. (**2001**). ''Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones,'' J. Acoust. Soc. Am. **109**, 1717–1722.

Chapman, S. (**1971**). ''Size, shape, and orientation of sonar targets measured remotely,'' Am. J. Phys. **39**, 1181–1190.

Gaunaurd, G. C., Brill, D., Huang, H., Moore, P. W. B., and Strifors, H. C. (**1998**). ''Signal processing of the echo signatures returned by submerged shells insonified by dolphin ''clicks:'' Active classification,'' J. Acoust. Soc. Am. **103**, 1547–1557.

Gellermann, L. W. (**1933**). ''Chance orders of alternative sitmuli in visual discrimination experiments,'' J. Genet. Psychol. **42**, 206–208.

Green, D. M., and Swets, J. A. (**1996**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hammer, C. E., Jr., and Au, W. W. L. (**1980**). ''Porpoise echo recognition: An analysis of controlling target characteristics,'' J. Acoust. Soc. Am. **68**, 1285–1293.

Hautus, M. J. (**1995**). ''Corrections for extreme proportions and their biasing effects on estimated values of d','' Behav. Res. Methods Instrum. Comput. **27**, 46–51.

Houser, D. S., Helweg, D. A., and Moore, P. W. (**1999**). ''Classification of dolphin echolocation clicks by energy and frequency distributions,'' J. Acoust. Soc. Am. **106**, 1579–1585.

Jacobs, D. W. (**1972**). ''Auditory frequency discrimination in the Atlantic bottlenose dolphin *Tursiops truncatus* Montagu: a preliminary report,'' J. Acoust. Soc. Am. **53**, 696–698.

Johnson, R. A., Moore, P. W. B., Stoermer, M. W., Pawloski, J. L., and Anderson, L. C. (**1988**). ''Temporal order discrimination within the dolphin critical interval,'' in *Animal Sonar: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 317–321.

Moore, P. W. B., Hall, R. W., Friedl, W. A., and Nachtigall, P. E. (**1984**). ''The critical interval in dolphin echolocation: What is it?'' J. Acoust. Soc. Am. **76**, 314–317.

Moore, P. W. B., and Schusterman, R. J. (**1987**). ''Audiometric assessment of Northern fur seals, *Callorhinus ursinus*,'' Marine Mammal Sci. **3**, 31–53.

Neubauer, W. G. (**1986**). *Acoustic Reflection from Surfaces and Shapes* (Naval Research Lab, Washington, DC).

Thompson, R. K. R., and Herman, L. M. (**1975**). ''Underwater frequency discrimination in the bottlenose dolphin (1–140 kHz) and the human (1–8 kHz),'' J. Acoust. Soc. Am. **57**, 943–948.

Urick, R. J. (**1983**). *Principles of Underwater Sound, 3rd ed.* (McGraw-Hill, New York).

Vel'min, V. A., and Dubrovskiy, N. A. (**1976**). ''The critical interval of active hearing in dolphins,'' Sov. Phys. Acoust. **2**, 351–352.

Development of form and function in peripheral auditory structures of the zebrafish (*Danio rerio*)[a)]

Dennis M. Higgs,[b)] Audrey K. Rollo, Marcy J. Souza,[c)] and Arthur N. Popper
Department of Biology, University of Maryland, College Park, Maryland 20742

(Received 10 May 2002; revised 3 November 2002; accepted 18 November 2002)

Investigations of the development of auditory form and function have, with a few exceptions, thus far been largely restricted to birds and mammals, making it difficult to postulate evolutionary hypotheses. Teleost fishes represent useful models for developmental investigations of the auditory system due to their often extensive period of posthatching development and the diversity of auditory specializations in this group. Using the auditory brainstem response and morphological techniques we investigated the development of auditory form and function in zebrafish (*Danio rerio*) ranging in size from 10 to 45 mm total length. We found no difference in auditory sensitivity, response latency, or response amplitude with development, but we did find an expansion of maximum detectable frequency from 200 Hz at 10 mm to 4000 Hz at 45 mm TL. The expansion of frequency range coincided with the development of Weberian ossicles in zebrafish, suggesting that changes in hearing ability in this species are driven more by development of auxiliary specializations than by the ear itself. We propose a model for the development of zebrafish hearing wherein the Weberian ossicles gradually increase the range of frequencies available to the inner ear, much as middle ear development increases frequency range in mammals. © *2003 Acoustical Society of America.*
[DOI: 10.1121/1.1536185]

PACS numbers: 43.80.Lb, 43.64.Ri, 43.64.Tk [WA]

I. INTRODUCTION

A comparative approach to studies of auditory processing can be informative both for questions of human hearing deficits and for questions of auditory evolution. This is particularly true from a developmental perspective, as even small changes in auditory structure can have profound effects on hearing ability (Werner and Gray, 1998). Most of the work done thus far on development of hearing structure and function (reviewed in Werner and Gray, 1998) has been in mammals (e.g., Ehret and Romand, 1981; Walsh *et al.*, 1986a; Geal-Dor *et al.*, 1993; Hill *et al.*, 1998) and a few species of birds (e.g., Gray and Rubel, 1985; Dmitrieva and Gottlieb, 1992; Gray, 1993; Brittan-Powell and Dooling, 2000), with less attention paid to other vertebrates. These studies have shown that as mammals and birds develop, responses are found first to low and middle frequencies and only later do responses to higher frequencies develop (e.g., Moore and Irvine, 1979; Ehret and Romand, 1981; Gray and Rubel, 1985; Brittan-Powell and Dooling, 2000), despite the fact that morphological development proceeds from high frequency to low frequency regions of the cochlea (Pujol and Marty, 1970; Rubel, 1978). In mammals this apparent discrepancy has been linked to the opening of the external ear canal (Hill *et al.*, 1998) and formation of the middle ear bones (Ehret and Romand, 1981; Geal-Dor *et al.*, 1993), both of which are necessary to transmit higher frequency information to the inner ear. Mammals and birds also show a developmental decrease in the latency of brainstem response to auditory stimulation (e.g., Walsh *et al.*, 1986b; Kuse and Okaniwa, 1993; Hill *et al.*, 1998; Brittan-Powell and Dooling, 2000) and a developmental increase in amplitude of brainstem response (e.g., Walsh *et al.*, 1986c; Kuse and Okaniwa, 1993; Brittan-Powell and Dooling, 2000), perhaps due to changes in myelination of neurons in the auditory system, innervation of the sensory cells of the ear, and cochlear mechanics (Walsh *et al.*, 1986b, c). Thus, correlation between development of auditory performance and structure can be used to construct hypotheses on the role of different portions of the auditory system in hearing ability. The ability to test evolutionary hypotheses is constrained, however, by the relatively limited focus on birds or mammals of previous studies.

Apart from a few studies during metamorphosis of frogs (e.g., Schofner and Feng, 1981; Boatwright-Horowitz and Megala Simmons, 1995, 1997) the only other developmental studies of auditory function of which we are aware are a few done in fishes. In the ray (*Raja clavata*), there is an increase in the sensitivity of the ramus neglectus nerve, stimulated as an isolated ear preparation, with development, and it has been suggested that this increased sensitivity is due to an increase in the number of sensory hair cells (Corwin, 1983). In contrast, no change in auditory sensitivity with growth has been found in the juvenile and adult stages of goldfish (*Carassius auratus*) using heart rate conditioning (Popper, 1971) and zebrafish (*Danio rerio*) using evoked brainstem responses (Higgs *et al.*, 2002a) despite significant increases in the number of sensory hair cells (Platt, 1977; Higgs *et al.*, 2002a). In other teleosts there are either large increases in auditory sensitivity over the entire range of detectable fre-

[a)]Portions of this work were presented at the annual meeting of the Association for Research in Otolaryngology, 2001.
[b)]Current address: Department of Biology, University of Windsor, Windsor, ON N9B 3P4, Canada. Electronic mail: dhiggs@uwindsor.ca
[c)]Current address: North Carolina State University, College of Veterinary Medicine, Raleigh, NC.

quencies [using behavioral conditioning, damselfish, *Pomacentrus* spp. (Kenyon, 1996)] or small improvements in sensitivity over a much narrower range of audible frequencies [Red Sea bream, *Pagrus major*, with heart rate conditioning (Iwashita *et al.*, 1999); gourami, *Trichopsis vittata*, with brainstem responses (Wysocki and Ladich, 2001)] during the juvenile and adult stages. Behavioral work has shown increases in responsiveness to a broadband auditory stimulus during the larval and juvenile periods of fish [Atlantic herring, *Clupea harengus* (Blaxter and Batty, 1985); red drum, *Sciaenops ocellatus* (Fuiman *et al.*, 1999)] and in herring this increased responsiveness has been correlated to inflation of the auditory bullae, gas-filled chambers directly connected to the inner ear in this species (Blaxter and Batty, 1985).

The purpose of the current study was to examine developmental changes in auditory structure and function in zebrafish. Zebrafish are an important model species for many aspects of vertebrate biology and are particularly useful for auditory work because they belong to the superorder Ostariophysi, a group of fish known as hearing specialists due to their broad range of detectable frequencies and specialized Weberian apparatus connecting the swim bladder to the ear (von Frisch, 1938; Fay and Popper, 1974). While there has been some examination of the morphology of the adult (Platt, 1993) and developing (Waterman and Bell, 1984; Haddon and Lewis, 1996; Riley *et al.*, 1997; Bang *et al.*, 2001) zebrafish ear, there has been no examination of the development of zebrafish auditory function except for our previous work on hearing in juveniles and adults (Higgs *et al.*, 2002a).

II. MATERIALS AND METHODS

A. Animal supply

We examined auditory abilities and morphological development in zebrafish from 10 to 45 mm total length (TL). The zebrafish used in this study were bred and reared in our fish colony at the University of Maryland. Adults used as broodstock were purchased from a local pet store, kept in a 38 L aquarium over marbles, and fed several times each day. Embryos were collected by siphoning from the bottom of the tank. Larvae were reared in small net baskets in a 38 L aquarium until they reached approximately 15 mm total length TL, at which point they were placed loose into a tank and kept in uncrowded conditions (see Higgs *et al.*, 2002a). Ages of fish used were not determined because length is a better indicator of developmental state than age for fish (Fuiman *et al.*, 1998; Higgs *et al.*, 2002a). All animal rearing and experimental methods were approved by the Institutional Animal Use and Care Committee at the University of Maryland.

B. Auditory physiology

We used the auditory brainstem response (ABR) to examine changes in hearing ability during the larval, juvenile, and adult period of zebrafish to ascertain how hearing function may change in this species. The use of ABR has become common in studies of auditory ability in a wide variety of vertebrates (e.g., Corwin *et al.*, 1982; Klein, 1984; Walsh *et al.*, 1986a; Brittan-Powell and Dooling, 2000), including fishes (e.g., Corwin *et al.*, 1982; Kenyon *et al.*, 1998; Yan and Curtsinger, 2000; Higgs *et al.*, 2002a), and is particularly suited to developmental investigations as it requires no training of the animal and can be performed noninvasively. This last attribute was essential for success in our very small zebrafish larvae. The methods used to measure auditory abilities in the current study are similar to those in Higgs *et al.* (2002a) but the animals were considerably smaller in the current study.

A total of 31 zebrafish from 10 to 45 mm TL were used for ABR, with all testing conducted in a sound attenuating chamber (Industrial Acoustics Company, New York). Animals were wrapped in a small mesh rectangle so that the entire fish was surrounded by mesh. The mesh was then clipped onto a holder and lowered into a 20 L water-filled bucket until the fish was completely submerged. This arrangement was loose enough to allow the fish to accelerate with the sound wave while remaining still enough for electrode placement. Fine positioning of the fish was controlled with a micromanipulator attached to the net holder. At final position the animal was approximately 25 cm above an underwater speaker (UW-30, Underwater Sound Inc., Oklahoma City, OK) and approximately 5 cm under the water surface. No muscle relaxants or anesthetics were needed for these experiments. Temperature of the water in the bucket ranged from 21 °C to 23 °C. To control for possibly spurious responses, three dead adult fish were also tested in our apparatus. At no time did a dead fish give a ''response'' in any way similar to those seen for the experimental animals.

Presentation of auditory stimuli was controlled using a Tucker-Davis Technologies (TDT, Gainesville, FL) physiology apparatus controlled by a computer running SigGen and BioSig software (TDT). Stimuli were played from the computer to the UW-30 underwater speaker and consisted of tone bursts of 100, 200, 400, 600, 800, 1000, 2000 or 4000 Hz. No frequencies above 4000 Hz were presented because a previous study (Higgs *et al.*, 2002a) showed that adult zebrafish never respond to higher frequencies. Calibration of output intensity for each frequency was accomplished using a hydrophone with precalibrated amplifier (calibration sensitivity of -195 dB nominal re: 1V/μPa; 0.2–10 kHz, omnidirectional, InterOcean Systems, San Diego, CA). Use of this calibration technique revealed that our thresholds previously published for adult zebrafish (Higgs *et al.*, 2002a) were in error (see erratum Higgs *et al.*, 2002b) and results in thresholds approximately 30 dB lower than those used in the previous study. Tone bursts had a 5-ms duration with a 2-ms rise/fall time and were gated through a Hanning window. Despite large sidebands to the stimulus at frequencies below 800 Hz, the level of the second harmonic was at least 15 dBV below the fundamental output frequency for all frequencies used.

Auditory responses to presented stimuli were collected using two stainless steel electrodes (Rochester Electro-Medical Inc., Tampa, FL) resting on the surface of the fish head. The recording electrode was positioned on the dorsal

midline of the fish just posterior to the operculum using a micromanipulator. The reference electrode was placed, also using a micromanipulator, on the dorsal midline just behind the eyes. All exposed surfaces of the electrode tip that were not in direct contact with the fish were coated with fingernail polish for insulation. Care was taken not to penetrate the skin of the fish with the electrodes since this hampered survival. A total of 400 responses (200 from stimuli presented at 90 degrees and 200 from stimuli presented at 270 degrees to cancel stimulus artifacts) were averaged together for each sound level at each frequency, after going through a 60-Hz notch filter to remove electrical noise.

Sound intensity at each frequency was increased in 5-dB steps until a stereotypical ABR was seen and then continued at least two steps (10 dB) higher to examine suprathreshold responses. Threshold was defined as the lowest level at which a clear response could be seen. This visual detection method is commonly employed in ABR studies (e.g., Walsh *et al.*, 1986a; Hall, 1992) and gives identical results to those achieved using more statistical approaches (Mann *et al.*, 2001).

For measurement of latency and amplitude of auditory responses we used responses that occurred at 5 dB above threshold for each animal examined above. A value of 5 dB above threshold was used to standardize across animals because of the variation between individuals in the level necessary for auditory stimulation. We did not use traces at a higher suprathreshold level because at some of the higher sound levels the responses were overwhelmed by stimulus artifact. Latency of the response was defined as the time between arrival of the stimulus (calculated as the time of stimulus onset minus 0.17 ms to account for travel time, assuming a speed of sound in water of $14\,872.6\ \mathrm{m\cdot s^{-1}}$ and a travel distance of 25 cm) and the maximum position of the first trough on the ABR waveform [Fig. 1(a)]. Amplitude was defined as the amplitude of the first trough relative to the background noise level just preceding the trough [Fig. 1(a)].

C. Morphology

To determine what morphological structures might be driving changes in auditory physiology we examined the number of saccular and lagenar sensory hair cells, the size of anterior and posterior regions of the saccule, the size of the swim bladder, and the development of Weberian ossicles in fish from 10 to 45 mm TL. Before fixation, fish were heavily anesthetized in MS-222 and the total length was measured. Fish were then fixed in 4% paraformaldehyde, except for those animals in which the swim bladder was measured. Swim bladders were removed for measurement from unfixed but anesthetized animals and immediately viewed under a dissecting microscope connected to a digital camera. The camera was connected to a computer with the MagnaFire (Optronics, Inc., Goleta, CA) imaging system. The lengths of the anterior and posterior chambers of the swim bladder were measured using NIH image software.

For hair cell counts, the saccules and lagenae of 12 fish from 15 to 45 mm TL were dissected free from the ear and stained with 2.5% Oregon-green conjugated phalloidin (Molecular Probes, Eugene, OR), an actin specific label that has been used to stain hair cell stereocilia in previous work (Higgs *et al.*, 2002a). Whole mounts of stained epithelia were coverslipped with Prolong antifade (Molecular Probes) and viewed under a Zeiss epifluorescence microscope. Digital images were taken at 400× magnification across the surface of the epithelium and then compiled into one image reconstructing the entire epithelial surface using Photoshop 6.0 (Adobe Systems, Inc., San Jose, CA). Counts of the total hair cell number were then taken either directly from the computer screen or, more often, from printouts of these images.

Images of saccules stained with phalloidin were also used to measure saccule size. Images of entire saccular epithelia taken at 100× magnification were used in NIH image software to estimate the perimeter of both the anterior and posterior halves of the saccule for comparisons of differential growth of these two regions. Simple linear regression was used to examine changes in hair cell number and sizes of saccular regions with development. To compare growth rate of the two different saccular regions, the regression coefficients of saccular perimeter estimates (anterior versus posterior) were compared using the Student's *t*-test (Zar, 1984).

To estimate progression of Weberian ossicle development, eight animals from 5 to 20 mm TL were cleared and stained following the protocol of Dingerkus and Uhler (1977). Animals were fixed in 4% paraformaldehyde, rinsed in distilled water for 2–3 days and, for larger animals, the skin was carefully removed to ensure penetration of the various chemicals. Animals were then placed in a mixture of alcian blue: 95% ethanol: glacial acetic acid for 24 h, rinsed through an ethanol series into distilled water, and placed into a solution of aqueous sodium borate with trypsin until the flesh was cleared and the bones were visible as blue structures underneath (approximately 15–17 days). Cleared specimens were then placed in an aqueous KOH solution with approximately 2–4 grains of alizarin red for 24 h and transferred to glycerin for storage. Images of stained fish were captured under a Wild dissecting scope with imaging capabilities. Detailed description of Weberian development was not attempted as this work is near completion in a different laboratory (Grande and Young, submitted) and would therefore have represented a duplication of effort. Only enough animals were examined to provide a general picture of Weberian ossicle development.

D. Statistical analyses

Because of the difficulty of performing physiological recordings on the small animals measured in the current study, fish were grouped into size classes to perform statistical comparisons of functional development. Based on similarity of physiological responses, animals were grouped into size classes of 10–13 mm TL ($n=4$), 15–16 mm TL ($n=3$), 17–20 mm TL ($n=8$), and animals over 20 mm TL ($n=6$). As it was not possible to obtain measurements of fish TL before running an ABR due to stress of handling, it was not deemed efficient to continue running trials until each size class contained the exact same number of animals. Variability in responses was similar across size classes so we feel that more trials would have yielded the same results. For

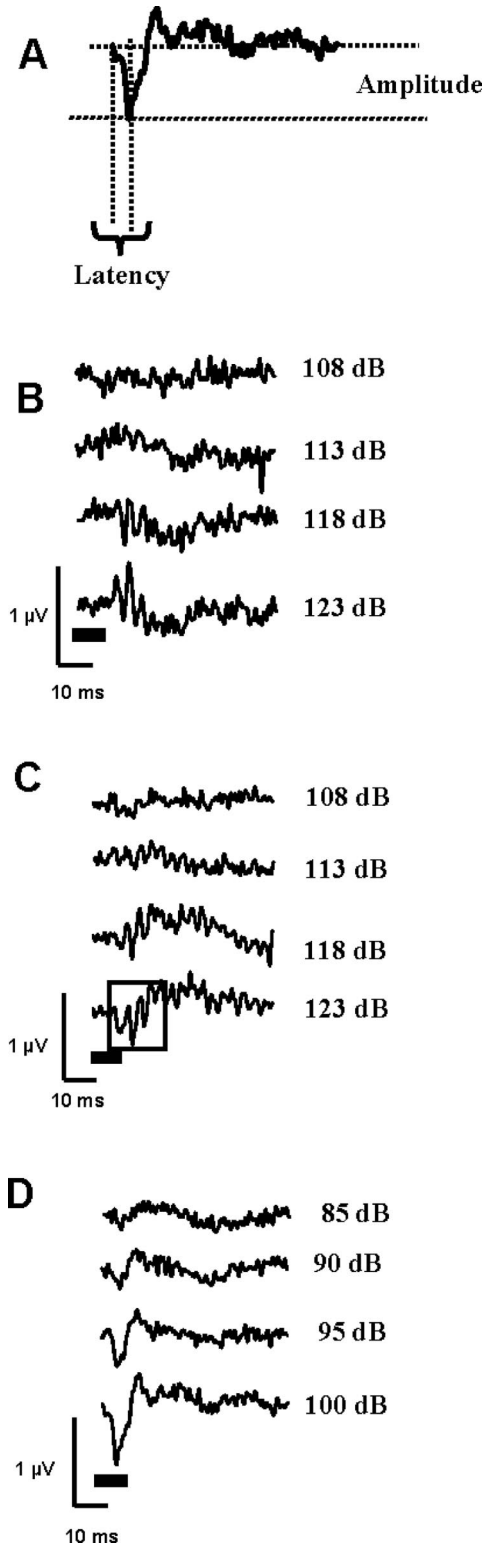


FIG. 1. (a) An example response waveform (to an 800-Hz stimulus) showing measurement parameters for latency and amplitude of the response. There were no qualitative differences in the shape of ABR waveforms in response to 100-Hz tone bursts across sizes, shown here for a 13.5-mm total length zebrafish larva (b) and a 42-mm total length zebrafish larva (c). The box in (c) shows the waveform region containing the initial response with the apparent frequency doubling seen at 100 and 200 Hz for all fish tested. The ABR responses to 200-Hz tone bursts looked identical to those shown here for 100 Hz. Above 200 Hz, all ABR waveforms looked like those shown here, for example, at 800 Hz in a 42-mm larva (d). All intensity values are dB *re* 1 μPa. The bars under waveforms in (b)–(d) represent stimulus timing. Waveforms were band-pass filtered between 30 and 1000 Hz for presentation.

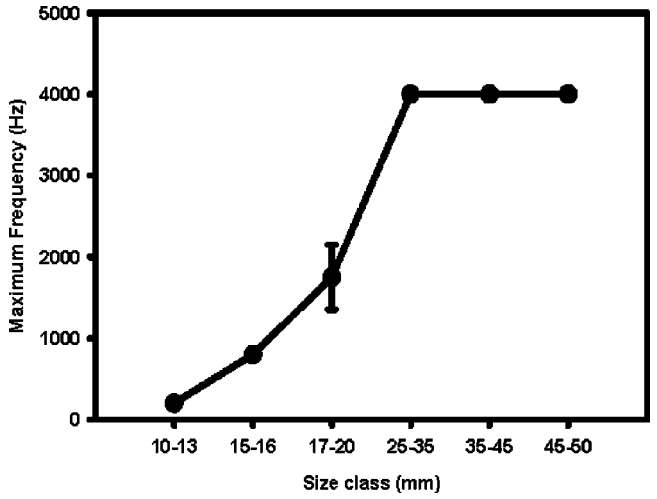


FIG. 2. The maximum frequency to which zebrafish showed an ABR gradually increased from 200 Hz in 10–13-mm larvae up to 4000 Hz in larvae larger than 20 mm. The >20 mm size class has been subdivided to visually demonstrate that maximum frequency of detection plateaus at 4000 Hz for zebrafish. Symbols represent mean ±1 s.e. Numbers of animals used are given in text.

comparisons of threshold, latency, and amplitude of the response two-way ANOVAs were run with frequency and size class as the independent variables. When significant interactions of frequency*size class were found, individual ANOVAs were conducted across size class for each frequency to focus on the comparisons of interest, although this inflates the probability of a Type I error (Zar, 1984). Significance level for individual ANOVAs was therefore set to $\alpha/n-1$, where $n=8$ (the number of possible comparisons). This gives a critical α of 0.006 for individual frequency comparisons of threshold, latency, and amplitude. Morphological measures of hair cell number, saccule size and swimbladder size were conducted as simple linear regression, using $P<0.05$ as the critical level.

III. RESULTS

A. Physiology

The shape of the ABR waveform differed depending on the frequency of the tone burst presented. For responses to 100- and 200-Hz tone bursts, there were three waves within the first 15 ms of tone presentation with what appeared to be a frequency doubling response [Figs. 1(b) and (c)]. For tone bursts of 400 Hz and above, there was one large trough in response to the tone burst, with waveforms quickly returning to background levels after the response [Fig. 1(d)]. Within a given frequency, there was no apparent change in the shape of the waveforms over development in zebrafish [Figs. 1(b) and (c)].

There was an increase in maximum frequency to which animals responded over development (Fig. 2). Animals from 10–13 mm ($n=4$) all responded to 100- and 200-Hz tone bursts but never responded to any tone bursts above 200 Hz. All animals from 15–16 mm ($n=3$) responded up to 800 Hz but never above. Animals from 17–20 mm ($n=8$) responded to tone bursts up to 2000 Hz with the mean maxi-

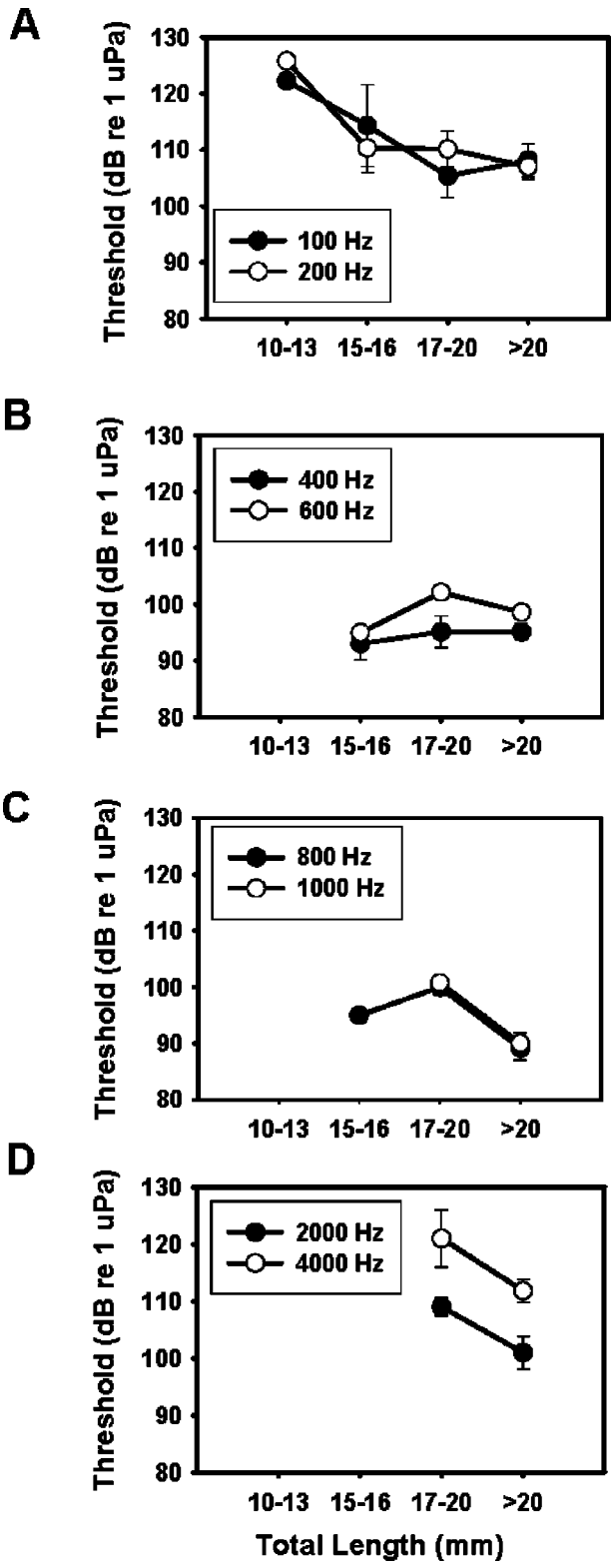


FIG. 3. Auditory threshold shows no consistent differences with growth of zebrafish larvae across frequencies. (a) 100 and 200 Hz, (b) 400 and 600 Hz, (c) 800 and 1000 Hz, and (d) 2000 and 4000 Hz.

mum frequency for the size class being 1750 (± 399.5 SE) Hz. All animals larger than 20 mm ($n=6$) responded to tone bursts up to and including 4000 Hz (Fig. 2).

The threshold at which animals responded to specific frequencies showed no consistent changes with development (Fig. 3). While there was a significant frequency*size interaction ($P<0.001$) in the ANOVA for threshold, there were no consistent growth effects on threshold. At 100 Hz animals responded to tones between approximately 105 and 125 dB (re: 1 μPa) with no significant differences ($P>0.05$) between size classes [Fig. 3(a)]. At 200 Hz all animals responded between 105 and 125 dB (re: 1 μPa) with no consistent differences between size classes, although the smallest size class (10–13 mm) did tend to have higher thresholds than the three groups (15–16, 17–20, and >20 mm) of larger animals [Fig. 3(a)]. As frequency increased, fewer animals responded but there was no difference in threshold between sizes among fish that did respond [Figs. 3(b)–(d)]. At 800 Hz, the best frequency of adult animals, threshold ranged from 90 to 100 dB (re 1 μPa) for all responding animals regardless of size [Fig. 3(c)].

There was a significant frequency*size interaction ($P<0.001$) in the ANOVA for latency but no frequencies showed a significant difference after adjusting for multiple comparisons (Fig. 4). The only frequencies over which all animals responded (100 and 200 Hz) showed no significant differences ($P>0.05$) in response latency over development [Fig. 4(a)]. There tended to be a higher latency of response to 100- and 200-Hz tone bursts [overall mean latency 10–12 ms, Fig. 4(a)] than to higher frequencies [overall mean latency 6–8.5 ms, Figs. 4(b)–(d)] but it is not clear if the responses at 100–200 Hz are comparable to those at higher frequencies (see below).

Within each frequency, there was no difference ($P>0.05$) in response amplitude over development (Fig. 5). At 100 and 200 Hz, the only frequencies at which all fish responded, all responses at 5 dB above threshold were between -0.3 and -0.8 μV with no consistent changes with size [Fig. 5(a)]. As frequency increased fewer size classes of fish responded to the stimulus, but, when fish did respond, the amplitude of the response was independent of fish size [Figs. 5(b)–(d)].

B. Morphology

There was a significant increase in the total number of saccular ($P<0.001$, $r^2=0.84$) and lagenar ($P<0.001$, $r^2=0.70$) hair cells with development in zebrafish [Figs. 6(a) and (b)]. Saccular hair cell number increased from approximately 700 in the smallest animals examined (14 mm TL) up to 2000 in the largest fish [37 mm TL, Fig. 6(a)]. Lagenar hair cell number underwent a similar increase, from approximately 700 lagenar hair cells at 15 mm TL up to approximately 2500 at 36 mm TL and 3500 at 48 mm TL [Fig. 6(b)].

There was a significant increase in the perimeter of both the anterior ($r^2=0.49$, $P<0.01$) and posterior ($r^2=0.79$, $P<0.001$) regions of the saccule with development (Fig. 7). For both regions of the saccule, the perimeter of the sensory area went from approximately 0.5 mm at 14–15 mm TL to approximately 0.9 mm at 37 mm TL. There was no significant difference ($P>0.05$) in the rate of increase of the perimeter between the anterior and posterior saccule (anterior: $Y=0.02X+0.28$; posterior: $Y=0.02X+0.10$), showing isometric growth of the two saccular regions relative to one another (Fig. 7).

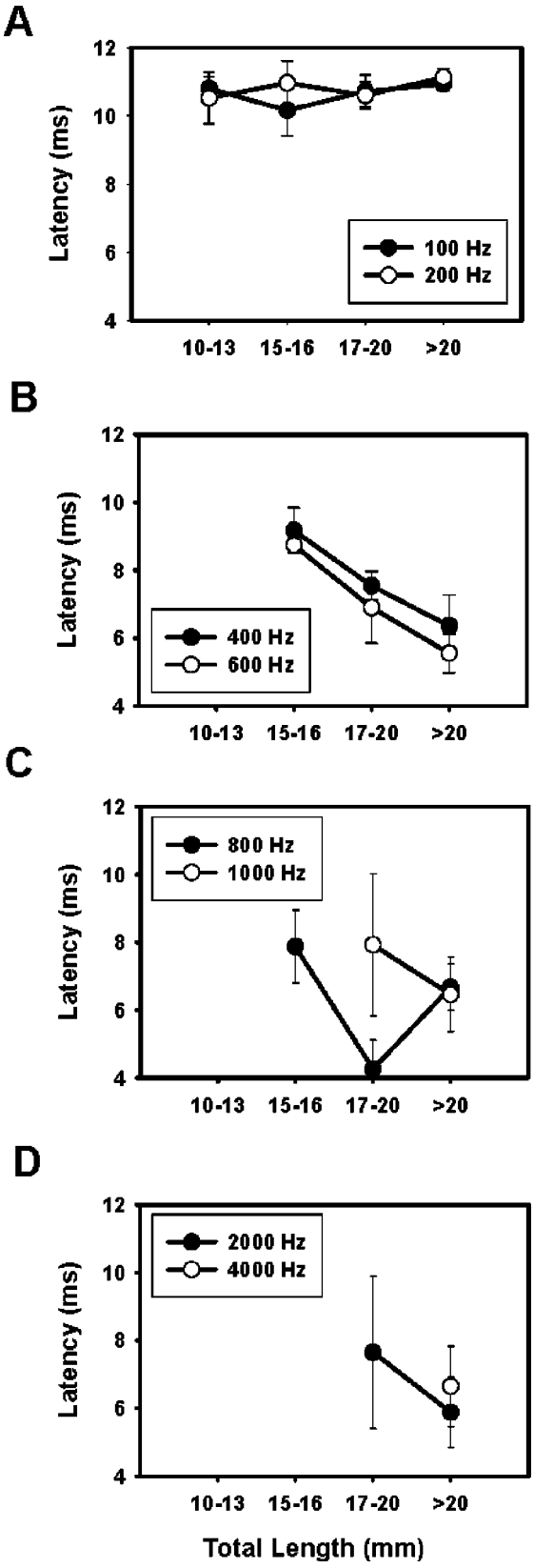


FIG. 4. The latency to response (time from arrival of stimulus to location of ABR trough) shows no consistent differences with growth of zebrafish larvae for the range of frequencies showing a response. (a) 100 and 200 Hz, (b) 400 and 600 Hz, (c) 800 and 1000 Hz, and (d) 2000 and 4000 Hz.

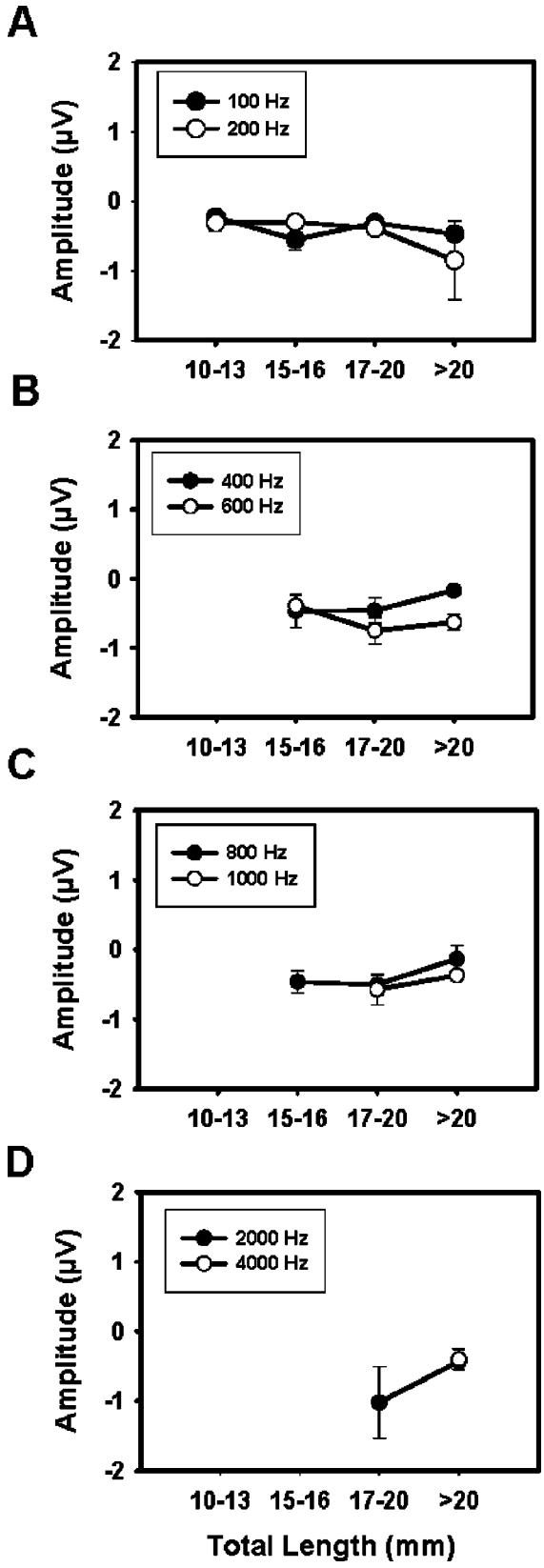


FIG. 5. The amplitude of the response (the size of the first trough relative to background noise levels) shows no consistent differences with growth of zebrafish larvae for the range of frequencies showing a response. (a) 100 and 200 Hz, (b) 400 and 600 Hz, (c) 800 and 1000 Hz, and (d) 2000 and 4000 Hz.

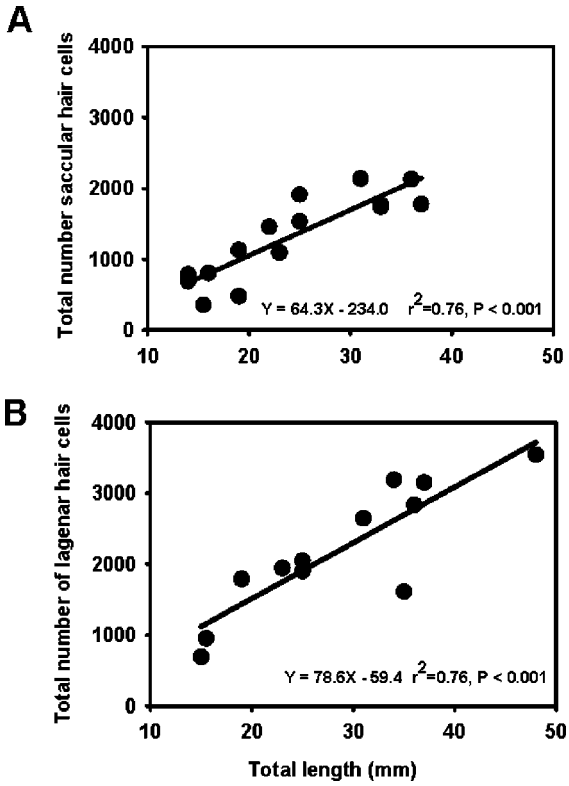


FIG. 6. There was a significant increase in the total number of saccular (a) and lagenar (b) sensory hair cells with growth of zebrafish.

The swim bladder first showed clear division into anterior and posterior chambers at 10 mm TL. Both anterior and posterior swim bladder chambers showed significant ($r^2 = 0.69$ and 0.86 for anterior and posterior chambers respectively, $P \ll 0.001$ for both) increases in length over development (Fig. 8). The anterior chamber tended to be more spherical than the posterior, with the posterior becoming more elongate as fish grew.

The first evidence of Weberian ossicle formation was seen at 7 mm TL [Fig. 9(a)]. At this size, the ossicles were quite small and had large gaps between ossicular elements. By 13 mm TL, the size of the individual ossicles had increased and the supraoccipital bone first became evident but

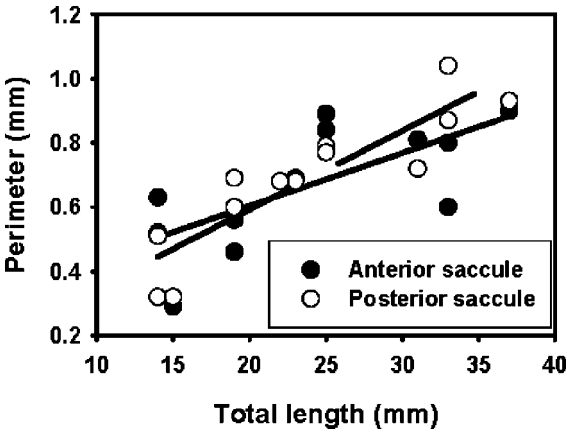


FIG. 7. The perimeter length of the sensory area of the anterior and posterior saccules increased significantly with growth but there was no significant difference in the rate of increase between these two saccular areas.

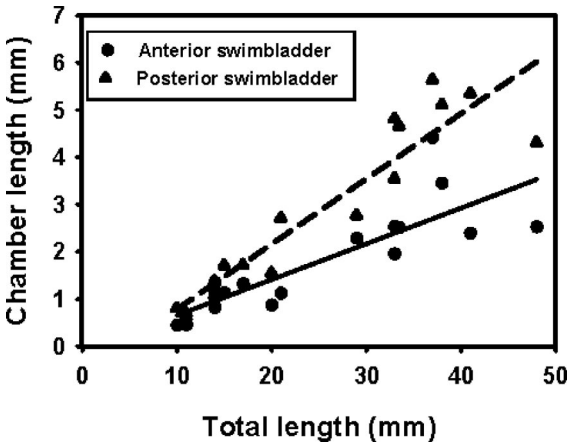


FIG. 8. The length of both the anterior and posterior swim bladder chambers showed significant increases with growth of zebrafish larvae.

there remained large gaps between individual ossicular elements [Fig. 9(b)]. Ossicle size increased but in fish at 17 mm TL there were still large spaces between individual ossicles and there was a prominent gap between the supraoccipital bone and the supraneurals of the Weberian apparatus [Fig. 9(c)]. By 19.5 mm TL, the ossicles were well formed and there was no gap between the supraoccipital bone and the supraneural elements of the Weberian apparatus, forming an unbroken chain of ossicles from the swimbladder to the inner ear [Fig. 9(d)].

IV. DISCUSSION

Before discussing the actual results of any physiological study, it is important to realize the potential limitations on the stimulus delivery and resulting responses. All sound stimuli contain both pressure and displacement information and, in our setup, with the speaker in the water, there is

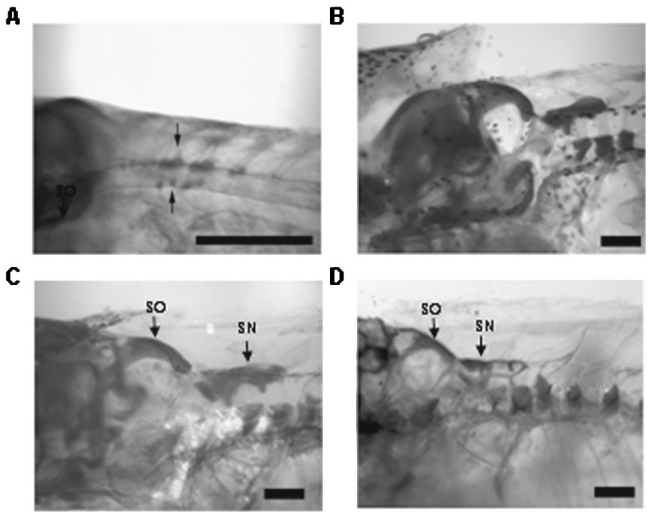


FIG. 9. Weberian ossicles are first evident at 7 mm total length in zebrafish [arrows in (a)] but are very small and poorly connected. By 13 mm (b) the ossicles are larger but large gaps remain between individual elements. By 17 mm TL (c) the dorsal plate has expanded but there are still gaps between individual elements and the supraoccipital (SO) is not connected to the supraneural (SN) Weberian elements. By 19.5 mm TL (d) the supraoccipital bone is well attached to the supraneurals, forming an unbroken chain from the Weberian apparatus to the inner ear. Scale bars = 0.1 mm.

probably quite a bit of displacement information present at the lowest frequencies used. Since the main purpose of this study was to examine changes in auditory ability between animals under constant experimental conditions, this does not cause a problem in the current study but must be kept in mind.

The differences in waveform shape at all sizes between responses to low (100–200 Hz) and middle to high (400–4000 Hz) frequencies suggest that perhaps different systems may be involved in detection of these frequencies. The fish should be well within the near-field domain for 100–200 Hz in the current setup (Rogers and Cox, 1988) so the lateral line system could also be stimulated by displacement effects of the presented sound stimuli. The multiple waveforms seen in response to 100- and 200-Hz stimuli therefore could represent a combination of lateral line and auditory responses, whereas higher frequencies would be expected to cause less stimulation to the lateral line (Rogers and Cox, 1988). Responses to tone bursts at 400 Hz and above should consist of mainly auditory contributions. Alternatively, the waveforms in response to 100- and 200-Hz stimulation might be the frequency doubling seen by Flock (1965), with the higher frequencies just representing temporal integration of the signal. There has as yet been no study published detailing how changes in waveform shape may relate to sensory structures in fish, as has been detailed so well in mammals (Hall, 1992). Analysis of this question may provide valuable insights on pathways of auditory transduction in fishes.

The increase of maximum detectable frequency seen in the current study has not been reported before for fishes, but is similar to data for mammals and birds. The development of the middle ear in mammals and birds allows transmission and therefore detection of higher frequency information in the inner ear (Ehret and Romand, 1981; Saunders *et al.*, 1983; Geal-Dor *et al.*, 1993; Hill *et al.*, 1998). In the current study, development of the Weberian ossicles coincides with expansion of auditory bandwidth. Fish in the 10–13 mm size class never responded to tone bursts above 200 Hz and their Weberian ossicles were small with large gaps between individual elements. The 15–16 and 17–20 mm size classes showed a gradual increase in detectable frequencies coincident with increases in size and connectivity of the Weberian elements and in the size of the swimbladder. By 20 mm the ossicles formed a continuous chain between a well developed swimbladder and the inner ear and those animals responded to pure tones up to 4000 Hz. In adult fish, it has long been hypothesized that the Weberian apparatus and swimbladder are responsible for transmitting higher frequency auditory information to the inner ear (von Frisch, 1938; Fay and Popper, 1974), and deflation of the swim bladder results in a reduction in high frequency sensitivity in ostariophysans such as zebrafish (Fay and Popper, 1974; Yan *et al.*, 2000). Our results are consistent with these observations. As the ossicles developed and became more highly connected to one another in zebrafish, and as the swim bladder increased in size, we saw a gradual shift in maximum detectable frequency from 200 Hz up to 4000 Hz. This then suggests that the ossicles and swimbladder are essential for detection of high frequency information in ostariophysan fishes.

It is also possible that the changes we saw in maximum detectable frequency are due to selective addition of high frequency hair cells in the saccule. Fish in the family Cyprinidae (to which zebrafish and goldfish belong) may have some degree of frequency coding in the saccule, such that higher frequencies are detected in the anterior saccule and lower frequencies are detected in the posterior saccule (Furukawa and Ishii, 1967; Fay, 1978; Moeng and Popper, 1984), although this still remains unclear. If selective addition of higher frequency hair cells were occurring with development, we would have expected to see differential growth of the saccule in the anterior-posterior plane. We did not see this but instead saw both regions growing at the same rate. There are also no differences in density distributions of saccular hair cells in zebrafish over development (Higgs *et al.*, 2002a), so measuring saccular size should be a good indicator of changes in hair cell distributions. Thus the increase in maximum detectable frequency is apparently not explained by selective addition of higher frequency hair cells.

The fact that there was no change in auditory sensitivity is interesting. Previous reports in teleosts have found either no change in auditory sensitivity with growth of adults in hearing specialists (i.e., a species with extra-aural hearing specializations; Popper, 1971; Higgs *et al.*, 2002a), a drastic improvement in sensitivity in a hearing generalist (i.e., a species with no extra-aural hearing specializations; Kenyon, 1996), or small changes over a restricted size range of fish in the two other teleost species tested (Iwashita *et al.*, 1999; Wysocki and Ladich, 2001). In the current study, we saw an increase in the number of auditory hair cells (increase in the number of sensory receptors) but no change in auditory sensitivity, at least not at the level of the ABR.

Measuring the physiological sensitivity of the eighth cranial nerve during development of an elasmobranch (the ray *Raja clavata*), Corwin (1983) found an increased sensitivity in conjunction with an increase in number of auditory hair cells. That Corwin (1983) found an increase in sensitivity and we did not may be due to a difference in techniques used between his studies and ours, or simply due to the wide disparity in species examined (elasmobranch versus teleost). Moreover, recordings from the eighth cranial nerve measure a different attribute of hearing than the synchrony required for an ABR response (Hall, 1992), so perhaps an increase in sensory receptors causes a different response in these two auditory measures. Alternatively, the response of the auditory system to an increase in hair cell number may be dependent on the auditory specializations in the studied species. Other studies that have found changes in auditory sensitivity with growth in fish have been conducted on hearing generalists (Corwin, 1983; Kenyon, 1996; Iwashita *et al.*, 1999) or on a hearing specialist with a specialization quite different from that seen in zebrafish and goldfish (Wysocki and Ladich, 2001). The form of auditory specializations may influence the developmental pattern of auditory sensitivity, although many more species will need to be examined before this can be determined.

A word of caution must be issued concerning comparison of absolute threshold values between laboratories, even when using the same species. The thresholds reported here for zebrafish are up 5–30 dB higher than those reported for goldfish by Yan *et al.* (2000) using ABR, even though our previous work (Higgs *et al.*, 2002a) showed little difference in threshold between goldfish and zebrafish in our setup. Previous work (Popper *et al.*, 1973; Fay, 1978) has shown a 30–50-dB difference in thresholds in goldfish between laboratories, even when similar methods were used. There is currently no standard method for testing hearing in fish and there are even large differences in technique between laboratories using ABR [e.g., we test fish under water while Yan *et al.* (2000) tested fish at the surface interface with an airborne speaker]. These methodological differences will make it impossible to perform interspecific comparisons using data from different laboratories. We propose that all laboratories presenting audiograms for a new species also include an audiogram of goldfish tested in the same system to better facilitate interspecific comparisons.

We postulate the following model for the development of hearing in zebrafish, and perhaps other ostariophysan fishes. By 10 mm TL, the ear appears quite well developed but the Weberian apparatus is not. As the swim bladder and Weberian ossicles develop and improve connections along the apparatus, more high frequency information can be passed along the ossicles to the inner ear. Once auditory information reaches the ear, the ear can process the information in the larvae as well as in the adult. While hair cells continue to be added to the inner ear throughout the life of the fish (Corwin, 1981, 1983; Popper and Hoxter, 1984; Lombarte and Popper, 1994; Higgs *et al.*, 2002a; current study), we suggest that this addition does not improve sensitivity, at least in zebrafish, but instead is used to keep pace with growth of the ear. This is supported by the fact that regional differences in hair cell density are maintained during development (Higgs *et al.*, 2002a) and by the fact that the different saccular regions grow at the same rate (current study). This also fits the predictions of a model that suggests that hair cell addition is necessary for stable hearing thresholds as the distance between the ear and peripheral structures such as the swimbladder increase (Popper *et al.*, 1988; Rogers *et al.*, 1988; Fineran and Hastings, 2000).

If our model of the development of zebrafish hearing is correct, this represents one more example of how similar the fish auditory system is to those of mammals and birds (see Fay and Popper, 2000). Just as mammals and birds seem to need the development of the middle ear for detection of higher frequencies (Ehret and Romand, 1981; Saunders *et al.*, 1983; Geal-Dor *et al.*, 1993), so too do at least zebrafish need development of the Weberian ossicles to transmit higher frequency information to the inner ear for detection. While it was initially thought that fish could not even hear (von Frisch, 1938), it is becoming increasingly obvious that the auditory system of many species of fish is quite advanced and possesses many of the attributes seen in amniotes (e.g., Fay and Popper, 2000). Fish can contain several types of auditory hair cells (Chang *et al.*, 1992; Popper *et al.*, 1993; Lanford *et al.*, 2000), have sharply tuned auditory filters (e.g., Fay, 1978), can detect sound direction and may be able to localize sounds (e.g., Schuijf and Buwalda, 1975; Hawkins and Sand, 1977; Lu and Popper, 2001), and can also perform complex auditory stream segregation necessary for auditory scene analysis (Fay, 2000). Thus rather than thinking of ''the fish'' auditory system as a rather general and unspecialized vertebrate ear, it is better to realize that the auditory systems of all vertebrates have many aspects in common and that examination of processes in the ear of a variety of fish species can tell us much about the evolution of the vertebrate auditory system in general (Fay and Popper, 2000).

ACKNOWLEDGMENTS

We thank Kirsten Poling, Beth Brittan-Powell, and Olivia Haine for critical review of the manuscript and discussions of the data in this study. We would also like to thank Alison Krupin, Sonya Rosenfeld, and Justina Efobi for assistance with the zebrafish developmental morphology. The manuscript was further improved by the comments of two anonymous reviewers. This work was funded as a grant from the NIH-NIDCD No. (DC04502-01) to DMH. Additional support was provided by grants to ANP (NIH-NIDCD No. DC-039036 and NIH-NIA No. AG-015681) and a NIDCD training grant (No. DC-00046).

Bang, P. I., Sewell, W. F., and Malicki, J. J. (**2001**). ''Morphology and cell type heterogeneities of the inner ear epithelia in adult and juvenile zebrafish (*Danio rerio*),'' J. Comp. Neurol. **438**, 173–190.

Blaxter, J. H. S., and Batty, R. S. (**1985**). ''The development of startle responses in herring larvae,'' J. Mar. Biol. Assoc. U.K. **65**, 737–750.

Boatwright-Horowitz, S. S., and Megela Simmons, A. (**1995**). ''Postmetamorphic changes in auditory sensitivity of the bullfrog midbrain,'' J. Comp. Physiol., A **177**, 577–590.

Boatwright-Horowitz, S. S., and Megela Simmons, A. (**1997**). ''Transient 'deafness' accompanies auditory development during metamorphosis from tadpole to frog,'' Proc. Natl. Acad. Sci. U.S.A. **94**, 14877–14882.

Brittan-Powell, E. F., and Dooling, R. J. (**2000**). ''Development of auditory sensitivity in budgerigars,'' J. Acoust. Soc. Am. **107**, 2785.

Chang, J. S. Y., Popper, A. N., and Saidel, W. M. (**1992**). ''Heterogeneity of sensory hair cells in a fish ear,'' J. Comp. Neurol. **324**, 621–640.

Corwin, J. T. (**1981**). ''Postembryonic production and aging of inner ear hair cells in sharks,'' J. Comp. Neurol. **201**, 541–553.

Corwin, J. T. (**1983**). ''Postembryonic growth of the macula neglecta auditory detector in the ray, *Raja clavata*: continual increases in hair cell number, neural convergence, and physiological sensitivity,'' J. Comp. Neurol. **217**, 345–356.

Corwin, J. T., Bullock, T. H., and Schweitzer, J. (**1982**). ''The auditory brain stem response in five vertebrate classes,'' Electroencephalogr. Clin. Neurophysiol. **54**, 629–641.

Dingerkus, G., and Uhler, L. D. (**1977**). ''Enzyme clearing of alcian blue stained whole small vertebrates for demonstration of cartilage,'' Stain Technol. **52**, 229–232.

Dmitrieva, L. P., and Gottlieb, G. (**1992**). ''Development of brainstem auditory pathway in mallard duck embryos and hatchlings,'' J. Comp. Physiol., A **171**, 665–671.

Ehret, G., and Romand, R. (**1981**). ''Postnatal development of absolute auditory thresholds in kittens,'' J. Comp. Physiol. Psychol. **95**, 304–311.

Fay, R. (**1978**). ''Coding of information in single auditory-nerve fibers of the goldfish,'' J. Acoust. Soc. Am. **63**, 136–146.

Fay, R. R. (**2000**). ''Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*),'' JARO **01**, 120–128.

Fay, R. R., and Popper, A. N. (**1974**). ''Acoustic stimulation of the ear of the goldfish (*Carassius auratus*),'' J. Exp. Biol. **61**, 243–260.

Fay, R. R., and Popper, A. N. (**2000**). ''Evolution of hearing in vertebrates: the inner ears and processing,'' Hear. Res. **149**, 1–10.

Fineran, J. J., and Hastings, M. C. (**2000**). ''A mathematical analysis of the peripheral auditory system mechanics in the goldfish (*Carassius auratus*),'' J. Acoust. Soc. Am. **108**, 1308–1321.
Flock, Å. (**1965**). ''Electron microscopic and electrophysiological studies on the lateral line canal organ,'' Acta Oto-Laryngol., Suppl. **199**, 1–90.
Fuiman, L. A., Poling, K. R., and Higgs, D. M. (**1998**). ''Quantifying developmental progress for comparative studies of larval fishes,'' Copeia **1998**, 602–611.
Fuiman, L. A., Smith, M. E., and Malley, V. N. (**1999**). ''Ontogeny of routine swimming speed and startle responses in red drum, with a comparison of responses to acoustic and visual stimuli,'' J. Fish Biol. **55**, 215–226.
Furukawa, T., and Ishii, Y. (**1967**). ''Neurophysiological studies on hearing in goldfish,'' J. Neurophysiol. **30**, 1377–1403.
Geal-Dor, M., Freeman, S., Li, G., and Sohmer, H. (**1993**). ''Development of hearing in neonatal rats: air and bone conducted ABR thresholds,'' Hear. Res. **69**, 236–242.
Grande, T., and Young, B. (**Submitted**). ''Ontogeny of the Weberian apparatus in the zebrafish Danio rerio (Ostariophysi, Cypriniformes),'' Can J. Zool.
Gray, L. (**1993**). ''Developmental changes in chickens' masked thresholds,'' Dev. Psychobiol. **26**, 447–457.
Gray, L., and Rubel, E. W. (**1985**). ''Development of absolute thresholds in chickens,'' J. Acoust. Soc. Am. **77**, 1162–1172.
Haddon, C., and Lewis, J. (**1996**). ''Early ear development in the embryo of the zebrafish, *Danio rerio*,'' J. Comp. Neurol. **365**, 113–128.
Hall, J. W. (**1992**). *Handbook of Auditory Evoked Responses* (Allyn and Bacon, Boston).
Hawkins, A. D., and Sand, O. (**1977**). ''Directional hearing in the median vertical plane by the cod,'' J. Comp. Physiol. [A] **122**, 1–8.
Higgs, D. M., Souza, M. J., Wilkins, H. R., Presson, J. C., and Popper, A. N. (**2002a**). ''Age- and size-related changes in the inner ear and hearing ability of the adult zebrafish (*Danio rerio*),'' JARO **3**, 174–184.
Higgs, D. M., Souza, M. J., Wilkins, H. R., Presson, J. C., and Popper, A. N. (**2002b**). ''Age- and size-related changes in the inner ear and hearing ability of the adult zebrafish (*Danio rerio*) ERRATUM,'' JARO **3**, 222.
Hill, K. G., Cone-Wesson, B., and Liu, G.-B. (**1998**). ''Development of auditory function in the tammar wallaby *Macropus eugenii*,'' Hear. Res. **117**, 97–106.
Iwashita, A., Sakamoto, M., Kojima, T., Watanabe, Y., and Soeda, H. (**1999**). ''Growth effects on the auditory threshold of Red Sea bream,'' Nippon Suisan Gakkaishi **65**, 833–838.
Kenyon, T. N. (**1996**). ''Ontogenetic changes in the auditory sensitivity of damselfishes (Pomacentridae),'' J. Comp. Physiol., A **179**, 553–561.
Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). ''A comparative study of hearing ability in fishes: the auditory brainstem response approach,'' J. Comp. Physiol. **182**, 307–318.
Klein, A. J. (**1984**). ''Frequency and age-dependent auditory evoked potential thresholds in infants,'' Hear. Res. **16**, 291–297.
Kuse, H., and Okaniwa, A. (**1993**). ''Postnatal development of the auditory brainstem response (ABR) in beagles,'' Exp. Anim. **42**, 377–382.
Lanford, P. J., Platt, C., and Popper, A. N. (**2000**). ''Structure and function in the saccule of the goldfish (*Carassius auratus*): a model of diversity in the non-amniote ear,'' Hear. Res. **143**, 1–13.
Lombarte, A., and Popper, A. N. (**1994**). ''Quantitative analyses of postembryonic hair cell addition in the otolithic endorgans of the inner ear of the European hake, *Merluccius merluccius* (Gadiformes, Teleostei),'' J. Comp. Neurol. **345**, 419–428.
Lu, Z., and Popper, A. N. (**2001**). ''Neural response directionality correlates of hair cell orientation in a teleost fish,'' J. Comp. Physiol., A **187**, 453–465.
Mann, D., Higgs, D., Tavolga, W., Souza, M., and Popper, A. N. (**2001**). ''Ultrasound detection by clupeiform fishes,'' J. Acoust. Soc. Am. **109**, 3048–3054.
Moeng, R., and Popper, A. N. (**1984**). ''Auditory responses of saccular neurons of the catfish, *Ictalurus punctatus*,'' J. Comp. Physiol. **155**, 615–624.
Moore, D. R., and Irvine, I. F. (**1979**). ''The development of some peripheral and central auditory responses in the neonatal cat,'' Brain Res. **163**, 49–59.
Platt, C. (**1977**). ''Hair cell distribution and orientation in goldfish otolith organs,'' J. Comp. Neurol. **172**, 283–298.
Platt, C. (**1993**). ''Zebrafish inner ear sensory surfaces are similar to those in goldfish,'' Hear. Res. **65**, 133–140.
Popper, A. N. (**1971**). ''The effects of fish size on auditory capacities of the goldfish,'' J. Aud Res. **XI**, 239–247.
Popper, A. N., and Hoxter, B. (**1984**). ''Growth of a fish ear: 1. Quantitative analysis of hair cell and ganglion cell proliferation,'' Hear. Res. **15**, 133–142.
Popper, A. N., Chan, A. T. H., and Clarke, N. L. (**1973**). ''An evaluation of methods for behavioral investigations of teleost audition,'' Behav. Res. Methods Instrum. **5**, 470–472.
Popper, A. N., Saidel, W. M., and Chang, J. S. Y. (**1993**). ''Two types of sensory hair cell in the saccule of a teleost fish,'' Hear. Res. **64**, 211–216.
Popper A. N., Rogers P. H., Saidel W. M., and Cox, M. (**1988**). ''The role of the fish ear in sound processing,'' in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga (Springer-Verlag, New York), pp. 687–710.
Pujol, R., and Marty, R. (**1970**). ''Postnatal maturation in the cochlea of the cat,'' J. Comp. Neurol. **139**, 115–126.
Riley, B. B., Zhu, C., Janetopoulos, C., and Aufderheide, K. J. (**1997**). ''A critical period of ear development controlled by distinct populations of ciliated cells in the zebrafish,'' Dev. Biol. **191**, 191–201.
Rogers, P. H., and Cox, M. (**1988**). ''Underwater sound as a biological stimulus,'' in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga (Springer-Verlag, New York), pp. 131–150.
Rogers, P. H., Popper, A. N., Cox, M., and Saidel, W. M. (**1988**). ''Processing of acoustic signals in the auditory system of bony fish,'' J. Acoust. Soc. Am. **83**, 338–349.
Rubel, E. W. (**1978**). ''Ontogeny of structure and function in the vertebrate auditory system,'' in *Handbook of Sensory Physiology Vol. IX*, edited by M. Jacobson (Springer-Verlag, New York), pp. 135–237.
Saunders, J. C., Relkin, E. M., Rosowski, J. J., and Bahl, C. (**1983**). ''Changes in middle-ear input admittance during postnatal auditory development in chicks,'' Hear. Res. **24**, 227–235.
Schofner, W. P., and Feng, A. S. (**1981**). ''Post-metamorphic development of the frequency selectivities and sensitivities of the peripheral auditory system of the bullfrog *Rana catesbiana*,'' J. Neurophysiol. **93**, 181–196.
Schuijf, A., and Buwalda, R. J. A. (**1975**). ''On the mechanism of directional hearing in cod (*Gadus morhua* L.),'' J. Comp. Physiol. **98**, 333–343.
von Frisch, K. (**1938**). ''The sense of hearing in fish,'' Nature (London) **141**, 8–11.
Walsh, E. J., McGee, J., and Javel, E. (**1986a**). ''Development of auditory-evoked potentials in the cat. I. Onset of response and development of sensitivity,'' J. Acoust. Soc. Am. **79**, 712–724.
Walsh, E. J., McGee, J., and Javel, E. (**1986b**). ''Development of auditory-evoked potentials in the cat. II. Wave latencies,'' J. Acoust. Soc. Am. **79**, 725–744.
Walsh, E. J., McGee, J., and Javel, E. (**1986c**). ''Development of auditory-evoked potentials in the cat. III. Wave amplitudes,'' J. Acoust. Soc. Am. **79**, 745–754.
Waterman, R. E., and Bell, D. H. (**1984**). ''Epithelial fusion during early semicircular canal formation in the embryonic zebrafish, *Danio rerio*,'' Anat. Rec. **210**, 101–114.
Werner, L. A., and Gray, L. (**1998**). ''Behavioral studies of hearing development,'' in *Development of the Auditory System*, edited by E. W. Rubel, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 12–79.
Wysocki, L. E., and Ladich, F. (**2001**). ''The ontogenetic development of auditory sensitivity, vocalization and acoustic communication in the labyrinth fish *Trichopsis vittata*,'' J. Comp. Physiol., A **187**, 177–187.
Yan, H. Y., and Curtsinger, W. S. (**2000**). ''The otic gasbladder as an ancillary auditory structure in a mormyrid fish,'' J. Comp. Physiol., A **186**, 595–602.
Yan, H. Y., Fine, M. L., Horn, N. S., and Colón, W. E. (**2000**). ''Variability in the role of the gasbladder in fish audition,'' J. Comp. Physiol., A **186**, 435–445.
Zar, J. H. (**1984**). *Biostatistical Analysis*, 2nd ed. (Prentice–Hall, Englewood Cliffs, NJ).

The effect of a low-frequency sound source (acoustic thermometry of the ocean climate) on the diving behavior of juvenile northern elephant seals, *Mirounga angustirostris*

Daniel P. Costa[a)]
Department of Ecology and Evolutionary Biology and the Institute of Marine Sciences, University of California, Santa Cruz, California 95064

Daniel E. Crocker
Department of Biology, Sonoma State University, Rohnert Park, California 94928-3609

Jason Gedamke
Department of Ocean Sciences and the Institute of Marine Science, University of California, Santa Cruz, California 95064

Paul M. Webb
Biology Department, Roger Williams University, One Old Ferry Road, Bristol, Rhode Island 02809

Dorian S. Houser
BIOMIMETICA, 5750 Amaya Drive, La Mesa, California 91942

Susanna B. Blackwell
Greeneridge Sciences, Inc., 1411 Firestone Road, Goleta, California 93117

Danielle Waples, Sean A. Hayes, and Burney J. Le Boeuf
Department of Ecology and Evolutionary Biology and the Institute of Marine Sciences, University of California, Santa Cruz, California 95064

(Received 4 March 2002; revised 26 July 2002; accepted 20 August 2002)

Changes in the diving behavior of individual free-ranging juvenile northern elephant seals, *Mirounga angustirostris*, exposed to the acoustic thermometry of the ocean climate (ATOC) sound source were examined using data loggers. Data loggers were attached to the animals and measured swim speed, maximum depth of dive, dive duration, surface interval, descent and ascent rate, and descent and ascent angle along with sound pressure level (SPL). The ATOC sound source was at a depth of 939 m and transmitted at 195 dB *re*: 1 μPa at 1 m centered at 75 Hz with a 37.5-Hz bandwidth. Sound pressure levels (SPL) measured at the seal during transmissions averaged 128 dB and ranged from 118 to 137 dB *re*: 1 μPa for the 60–90 Hz band, in comparison to ambient levels of 87–107 dB within this band. In no case did an animal end its dive or show any other obvious change in behavior upon exposure to the ATOC sound. Subtle changes in diving behavior were detected, however. During exposure, deviations in descent rate were greater than 1 s.d. of the control mean in 9 of 14 seals. Dive depth increased and descent velocity increased in three animals, ascent velocity decreased in two animals, ascent rate increased in one animal and decreased in another, and dive duration decreased in only one animal. There was a highly significant positive correlation between SPL and descent rate. The biological significance of these subtle changes is likely to be minimal. This is the first study to quantify behavioral responses of an animal underwater with simultaneous measurements of SPL of anthropogenic sounds recorded at the animal. © *2003 Acoustical Society of America.* [DOI: 10.1121/1.1538248]

PACS numbers: 43.80.Nd [WA]

I. INTRODUCTION

The enhanced transmission of sound and the poor conduction of light in the ocean compel marine mammals to rely heavily on acoustics as their primary sensory modality for many aspects of their life history. Marine mammals produce sounds to transmit information about location, intention, age, sex, reproductive status, and identity of the caller (Thompson *et al.*, 1979; Tyack and Whitehead, 1983; Watkins *et al.*, 1987, 2000; Van Parijs *et al.*, 1999; Tyack and Clark, 2000). The odontocete cetaceans use high-frequency biosonar to locate prey and to interrogate their environment (Au *et al.*, 1974). Most marine mammal species use passive listening to aid in location of prey, avoidance of predators, and navigation (Tyack and Clark, 2000). Given our understanding of how sound is used by marine mammals, there is reason for concern over the potential impact that anthropogenic noise may have upon them. Human generated underwater noise originates from a variety of sources, but the dominant sources are ship traffic and offshore industrial activities. The

a)Electronic mail: costa@biology.ucsc.edu

dominant sources of ship traffic include merchant vessels, icebreakers, naval activities, fishing fleets, and scientific research. Offshore industrial activities include seismic exploration, construction work, drilling, and oil and gas production (Green *et al.*, 1994; Richardson *et al.*, 1995; Gordon and Moscrop, 1996).

An additional, though somewhat infrequent, source of anthropogenic noise comes from oceanographic research. Although high-intensity low-frequency sound (LFS) sources have been used to study the ocean, concern was not raised about the effects of such sound sources on ocean fauna, particularly marine mammals, until the Heard Island Feasibility Test (Munk and Forbes, 1989; Munk *et al.*, 1994). The Acoustic Thermometry of the Ocean Climate (ATOC) experiment was a follow up to the Heard Island study and generated considerable debate (Potter, 1994). ATOC tested the feasibility of using changes in the speed of sound in the ocean measured across long distances (3000–5000 km) to estimate integrated ocean temperatures (Munk and Forbes, 1989; ATOC Consortium, 1998; Worcester *et al.*, 1999). Achieving this task was done through the introduction of a high intensity, low-frequency sound (broadband source level 195 dB *re*: 1 μPa at 1 m) into the deep sound channel (SOFAR channel) at two locations, off California and Hawaii. In order to determine whether these sounds affected marine mammals, a Marine Mammal Research Program (MMRP) was developed as part of the overall ATOC research program. The overall objective of the MMRP was to determine what species would be exposed to the operational ATOC sound source and assess the effects of that exposure.

The MMRP identified the northern elephant seal, *Mirounga angustirostris*, as a species of particular concern because these seals naturally migrate past the site of the California ATOC sound source, they are relatively abundant near the source, and they have the best low-frequency hearing capability of any pinniped measured to date (Kastak and Schusterman, 1998). This species is also one of the few marine mammals that frequent the deep sound channel in the northeastern Pacific where the California ATOC source was located (Le Boeuf *et al.*, 2000a). Finally, elephant seals are ideal research subjects for tag-based behavior studies because a wealth of background information exists on their population dynamics, general biology, and at-sea behavior and distribution (Le Boeuf and Laws, 1994).

Recent studies of the free-ranging diving behavior using attached data loggers show that northern elephant seals, *M. angustirostris*, as well as southern elephant seals, *M. leonina*, are consummate divers with few equals in the marine environment (e.g., Le Boeuf *et al.*, 1988; Hindell *et al.*, 1991; Stewart and DeLong, 1991). Northern elephant seals dive continuously, day and night, for periods at sea lasting from 2 to 8 months. They spend 90% of their time at sea submerged with an average of 20 min per dive (with maximum dive durations of over 1 $\frac{1}{2}$ h) followed by less than 3 min at the surface between dives. Northern elephant seals dive to modal depths of 300–600 m with maximum dives exceeding 1500 m. The diving behavior, periods at sea, and migratory paths are known at a general level for juveniles, subadult and adult males, and pregnant and nonpregnant females (Le Boeuf *et al.*, 1996, 2000a). Males travel from California to foraging areas along the continental slope. These foraging areas include the state of Washington north to the upper reaches of the Gulf of Alaska across to the eastern Aleutian Islands. Females disperse more widely across the Northeastern Pacific to as far as 150 °W (approx. due north of the Hawaiian Islands), in the range 44–52 °N (DeLong *et al.*, 1992; Le Boeuf *et al.*, 2000a). One-and-a-half-year-old juveniles of both sexes take similar paths as adult females (Le Boeuf *et al.*, 1996). Elephant seals have good hearing within the range of the ATOC sound source (Kastak and Schusterman, 1998) and frequently dive to the deep sound channel. Their long pelagic migrations are also more likely to cause them to be exposed to deep ocean noise as compared to more coastal species.

A significant problem with studying the effect of a particular sound on any animal is identifying the sound pressure level (SPL) to which the animal is exposed. Previous studies have estimated the intensity of sound exposure by correlating the animal's position with empirical measurements or by using model-based predictions of the sound field (Richardson *et al.*, 1995; Au *et al.*, 1997; Frankel and Clark, 1998, 2000; Erbe and Farmer, 2000). Although general features of a low-frequency sound source can be predicted and/or empirically measured on the large scale, there is much variability at the local scale of an individual animal, including variability in received levels with depth. Previous work has relied on surface measurements of animal location, without data on the animal's depth. The recent development of acoustic data loggers, which can be attached directly to free-ranging animals, allows accurate measurements of the SPL to which the animal is exposed (Fletcher *et al.*, 1996; Burgess *et al.*, 1998).

Simultaneous attachment of instruments that collect information on movements, dive depth and duration, duration at surface, swimming speed, and acoustic exposure enables measurements of received SPLs at the elephant seal to be correlated with simultaneous changes in dive behavior and location (Costa, 1993; Fletcher *et al.*, 1996; Burgess *et al.*, 1998). Changes in these behaviors should reflect the animals' response to the ATOC sound source. For example, if the source were aversive we would predict that animals might respond by spending more time at the surface, making shallower dives, remaining immobile, increasing their ascent rate and velocity when returning to the surface, or in the extreme completely ceasing to dive. Instruments such as data loggers are easily attached to the animals, and attachment to the pelage is secure even after 8 months at sea (Le Boeuf *et al.*, 2000a). Since elephant seals return reliably to the rookery, the instrument recovery rate is high (about 90% for adults and juveniles) and archival data loggers can be used instead of the more expensive, unrecoverable, and more limited telemetry transmitters that cetacean studies require (Mate *et al.*, 1999). Because these animals are free-ranging, the data gathered by these instruments can provide information on their underwater behavior, allow monitoring at great distances, and provide information that could not be obtained in any other way.

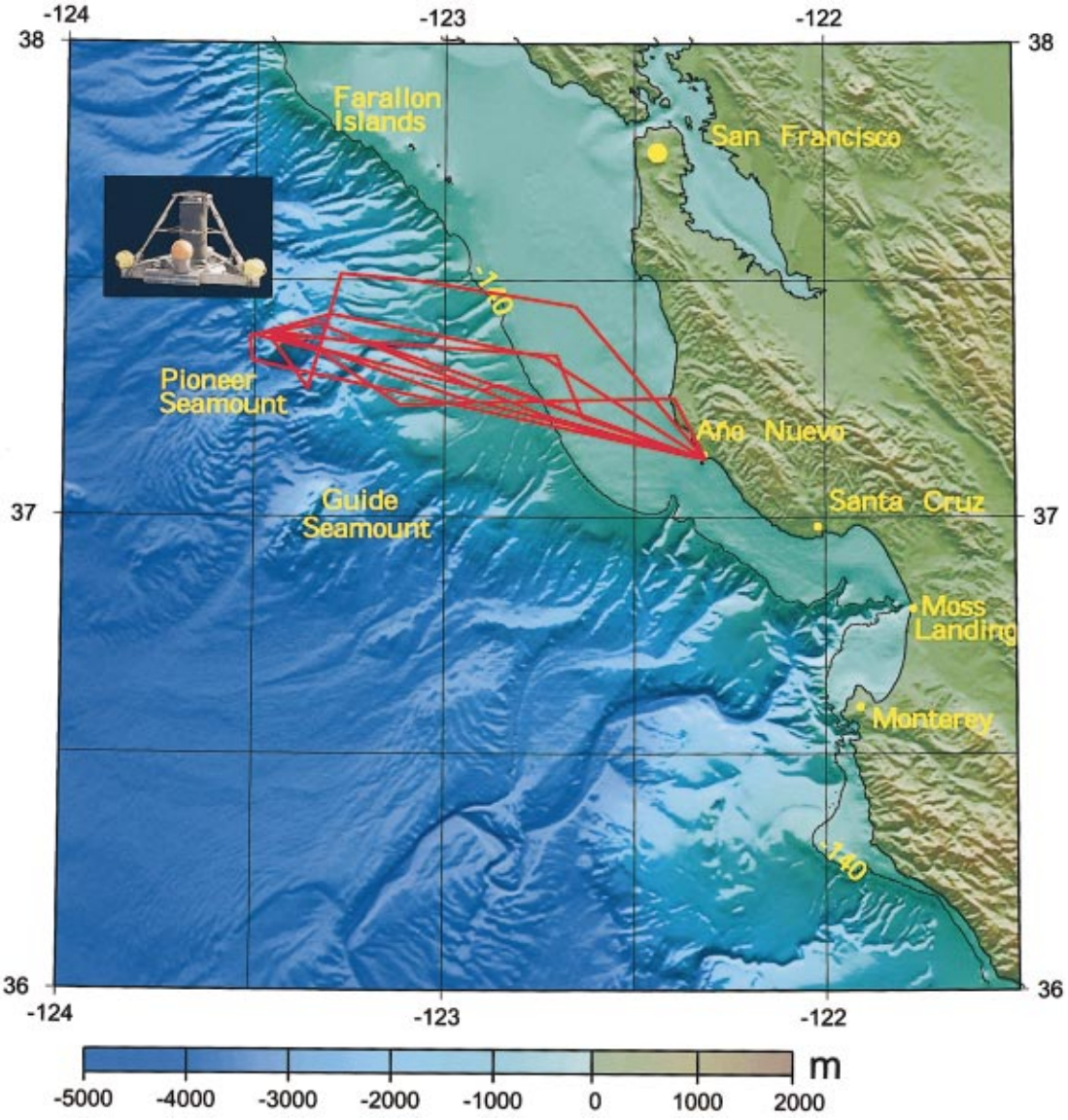


FIG. 1. A chart of the study area is shown with the location of the PIONEER SEAMOUNT relative to the Año Nuevo elephant seal rookery 102 kilometers away. The insert is the ATOC sound source. Red lines are representative elephant seal tracks recorded from the ARGOS satellite transmitters. ARGOS locations do not provide precise transit information as only a few locations, of variable quality (1–10 km), are received each day.

II. MATERIALS AND METHODS

A. ATOC sound source

The ATOC sound source (Alliant Techsystems, HX-554) was deployed in late October 1995 on the PIONEER SEAMOUNT (37°20.555′N, 123°26.7117′W), 100 km west of Half Moon Bay, California, at a depth of 939 m (Howe, 1996) (Fig. 1). As the focus of this study was to examine the effects of an operational ATOC source, the acoustic power, duty cycle, and frequency used during the trials followed those that would be used for the thermometry studies. The source signal was an M-sequence phase-modulated carrier that consisted of a center frequency of 75 Hz, a bandwidth of 37.5 Hz, and transmitted power of 260 W (195 dB *re*: 1 μPa at 1 m). The M-sequence modulated signal allowed matched filtering to be used to increase the precision of the time of arrival measurements while operating at significantly lower power than would be required without this coding (Munk *et al.*, 1994; Worcester *et al.*, 1999). This coding gives the ATOC sound source higher bandwidth and thus increases the likelihood of detection by a marine mammal relative to a 75 Hz pure tone (Au *et al.*, 1997). In an attempt to reduce the effect of the sound source on marine mammals, the source was started at 165 dB *re*: 1 μPa at 1 m and ''ramped up'' by 6-dB steps each minute for 5 min until reaching and maintaining full power at 195 dB for 20 min. Transmissions occurred 6 times a day at 4-h intervals for a maximum of 4 days. Ramp up was initiated 5 min prior to the hour, and full power achieved on the hour.

B. Approach

The response of elephant seals that were captured at the Año Nuevo rookery, translocated, and released beyond the source site was examined experimentally. With the selected protocol, there was a relatively high probability that they would pass over or near the source (i.e., within the 120-dB sound field) in returning to the Año Nuevo rookery. The

response of these animals to the ATOC sound source was examined using a variety of instruments that provided information on the diving behavior, return track, and the SPL at the animal. ARGOS satellite tags provided information on approximate location at sea. Information on dive behavior was collected using archival tags that, upon recovery, provided the animals' time-marked swim speed and dive depth (STDR) or dive depth alone (TDR). Finally, acoustic data loggers were used to provide information on the animals' ambient acoustic environment as well as the actual SPL of the ATOC signal as received at the animal.

We took advantage of the natural tendency of juvenile elephant seals to return to the site of capture (Oliver *et al.*, 1998) to experimentally expose individuals to the ATOC sound source. These ''translocation'' animals could be released in such a way that there was a high probability of exposure to the ATOC sound source (Fig. 1). This manipulation had a significant advantage over studies with adult animals because the whole manipulation could be completed in less than 1 week, whereas studies that utilized the natural migration of adults would require at minimum 3.5 months and up to 8 months to complete (Le Boeuf *et al.*, 2000a). Further, we were able to control the distance the animals were released from the source. As we knew that they would return directly to the Año Nuevo rookery, this would increase the probability that they would cross over the source on their return. Moreover, 2-year-old juveniles dive to the same mean depths and exhibit the same mean dive duration as adult males and females (Le Boeuf *et al.*, 1996).

C. Acoustic data loggers

Two forms of acoustic data loggers were used. The first was a DAT tag, which consisted of a Sony TCD-D8 digital audio tape recorder (DAT) (9 Hz–16 kHz±3 dB) enclosed in an aluminum housing (17.1×12.7×6.7 cm) with an external hydrophone (High Tech, Inc. HT1-SSQ-41b; 10 Hz–30 kHz±3 dB) that received the ATOC signal (Fletcher *et al.*, 1996). Units could be turned on either manually with a magnet just prior to release, or the DAT could be programed to automatically turn on approximately $\frac{1}{2}$ hour before animals were released in the water. These units sampled at 32 kHz, and recorded for a total of 4–8 hours depending on the tape used. Calibration tones of known rms voltage were recorded onto each tape to allow actual received SPL to be calculated from the recording.

The other instrument was called a compact acoustic probe or CAP tag (Burgess *et al.*, 1998). CAP tags used the same hydrophone as the DAT tag. Within the housing (36-cm long, 10-cm diameter cylinder) was a programmable Tattletale 7 data acquisition system (Onset Computer Corp., Pocasset, MA) and a 340-MB hard disk. The CAP tags also sampled pressure and temperature. Like the programmable DATs, these units were programed to turn on approximately $\frac{1}{2}$ hour before release. These units were programed to sample at 2000 Hz and could record for up to 8 days with a frequency response of 10–1000 Hz. The known gain of each step of the digital acquisition process allowed calibrated received SPLs to be calculated.

The DAT and CAP units were recovered after the animals returned to the beach. Acoustic data from the DAT packs were transferred digitally to an SGI workstation, and then low-pass filtered (4 kHz) and resampled (8 kHz) in order to minimize file size. The output sound files and calibration tones were then recorded onto a CD-ROM. Data from the CAP tags were transferred directly from the units to a Macintosh and recorded onto CD-ROM.

D. Animal manipulation

We used this experimental system to record the behavior of 29 juvenile (1.8 to 2.4 years old) elephant seals that were instrumented with ARGOS satellite tags and separate data loggers (3 during Fall 1995, 11 during Spring 1996, 11 during Fall 1996, 4 during Spring 1997) and released near the ATOC sound source. Each seal had the following instruments attached to the dorsum: (1) a custom-made swim speed, time-depth recorder (STDR) (B-H Mk1, Santa Cruz, CA) for recording the diving pattern; (2) an acoustic data logger (DAT or CAP); (3) a VHF transmitter (Advanced Telemetry Systems, Isanti, MN) to facilitate recovery of the animals on the rookery; and finally (4) an Argos satellite tag (model ST-6, Telonics, Mesa, AZ) attached to the back of the head to track movements and location (Fletcher *et al.*, 1996; Crocker *et al.*, 1997; Le Boeuf *et al.*, 2000a). The TDRs recorded depth every 10 s throughout the period at sea. The STDRs recorded relative swim speed every 10 s using a Logtron paddle wheel (Flash Electronic GmbH, Dachau, Germany) attached to the recorder; revolutions were counted and stored in memory until retrieved. The DATs and CAPs were programed to record continuous low-frequency sounds up to 8 h and 4 days after release, respectively.

The juveniles were captured at the Año Nuevo rookery and transported by truck to Long Marine Laboratory where they were weighed, measured, and instrumented. Animals were immobilized and instruments attached as previously described (Le Boeuf *et al.*, 1988, 2000a). The following day the seals were transported by ship and released approximately 3 km due west of the PIONEER SEAMOUNT 1 h prior to activation of the ATOC source. The release site required that seals returning directly to Año Nuevo would travel over the source. Seals were released from this site because, given that pilot studies suggest a horizontal transit rate of approximately 3 km/h, it maximized the likelihood that seals would cross over the sound source during its 20–25-min period of operation. This protocol was designed to place the animal in close proximity to the ATOC source at activation. In the worst-case scenario, where the seal (unexpectedly) travels west, directly away from the sound source, the protocol would result in animals being approximately 6 km from the source at activation. A direct line from the drop-off site to the rookery is 105 km. Upon return to the Año Nuevo rookery (3–5 days), the instruments were recovered and data downloaded for analysis.

E. Behavioral analysis

On recovery, each instrument was calibrated for swimming speed by plotting the rate of depth change against the

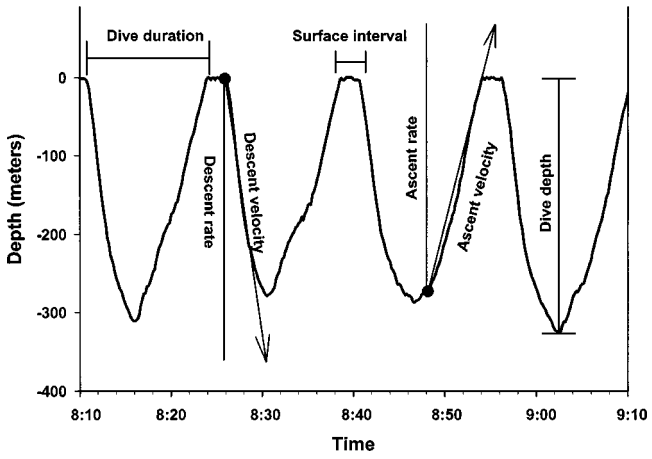


FIG. 2. Section from a typical elephant seal dive record along with the behavioral measurements used to describe the diving pattern.

number of impeller rotations over the same interval (Fletcher *et al.*, 1996; Blackwell *et al.*, 1999). A dive was defined as an excursion below 6 m. The times within dives when descent ended and ascent began were set manually by inspection (Crocker *et al.*, 1997). The following variables were calculated for each recorded dive: maximum depth, duration (start of descent to end of ascent), surface time between dives, descent rate, descent velocity, ascent rate, and ascent velocity (Fig. 2). Ascent and descent rate are measures of the rate that an animal moves vertically through the water column, without respect to any horizontal movement, and calculated as the rate of depth change per unit time. Ascent and descent velocities are measures of the animals' speed, in any direction, while ascending or descending; their velocities are functions of the fluid flow past the seal (Crocker *et al.*, 1997).

We compared the parameters of the dive occurring while the source was on to the mean parameters derived from a control period measured for each animal: an 18-h period of diving that started 18 h after the animal was first exposed to the sound source. Each animal was used as its own control to determine whether the behavior measured during the exposure dive fell within 1 or 2 standard deviations of the mean value for the 18 h of diving where the animal would be beyond the range of the ATOC sound source. Eighteen hours was chosen because all animals spent at least 36 h transiting over deep water after release from the boat. This allowed the animals 18 h to swim beyond the seamount and provided 18 h of data from which to derive a mean diving pattern. Comparisons were made only while the animals were in deep water, since their diving behavior changes when they move over the continental shelf (Le Boeuf and Crocker, 1996). In situations where more than one dive or portions of a dive occurred during exposure, we calculated the responses for the components of the dive that occurred during the exposure period. To account for the potential effect of translocation, a group of seven ''control'' animals was instrumented with time-depth recorders and satellite tags only and released in the same manner, except the sound source was not operating. The diving records of these animals were analyzed such that the ''exposure dive'' was assumed to be the dive occurring 1

TABLE I. Absolute and relative (%) change in behavior (during exposure dive as compared to the 18-h mean for each animal) for 7 dive variables and all 24 subjects, as a function of the maximum ATOC sound-pressure level (SPL) measured at each animal. A negative value would indicate a behavior with a magnitude less than control, for example the depth of dive in animal D9601 was 128 m shallower. Swim-speed data were not available for animals carrying the CAP tag. Control animals (ATOC source off) were assumed to have the mean measured ambient SPL of 100 dB *re*: 1 μPa. Numbers with an * and in bold are where the exposure dive was greater than ± 1 s.d. of each animal's 18-h mean. The sex of the animals is given.

	Depth		Duration		Surface interval		Descent velocity		Ascent velocity		Descent rate		Ascent rate		SPL
	Meters	%	Minutes	%	Minutes	%	m/s	%	m/s	%	m/s	%	m/s	%	dB *re*: 1 μPa
D9601 ♀	**−128***	−31.4	1.5	6.8	−0.2	−8.7	**0.24***	20.9	**−0.21***	−14.5	**0.38***	44.7	**−0.45***	−63.4	137
D96N2 ♂	57	15.5	2.4	13.8	−0.7	−17.5	0.06	4.3	0.07	4.7	**0.56***	76.7	0.19	29.2	135
B961 ♀	60	29.3	1.5	11.6	−0.1	−5.3	−0.14	−13.0	−0.21	−17.9	**0.31***	27.4	−0.02	−1.6	133
D96N1 ♀	**−63***	−19.5	**−3.9***	−20.3	0	0.0	0.15	11.6	0.1	8.2	0.1	14.7	0.04	5.3	132
SOX ♀	29	9.0	−2.1	−12.1	−0.1	−3.8	**0.53***	54.1	**−0.38***	−24.5	**0.51***	62.2	0.1	12.3	132
CARD ♂	81	35.2	−1.7	−11.3	−0.7	−23.3	**0.32***	26.9	0.12	9.6	**0.68***	63.6	−0.01	−1.0	131
MARLIN ♂	**−82***	−18.7	−2.6	−12.0	0	0.0	−0.03	−2.3	0.08	6.4	**0.42***	48.3	0.04	4.3	130
ASTRO ♀	3	0.8	−0.2	−0.9	0.4	16.7	0.11	8.9	−0.02	−1.5	0.09	9.4	−0.09	−8.7	126
CAP962 ♀	−13	−3.7	0.5	2.7	−0.1	−4.5	N/A	N/A	N/A	N/A	0.11	11.6	0.13	14.1	126
B962 ♀	53	26.1	1.8	13.6	−0.8	−26.7	−0.08	−7.1	−0.06	−5.0	**0.32***	34.4	0.04	4.2	124
EXPO ♂	31	10.3	0.9	5.9	0.2	11.8	0.02	2.0	0.32	24.6	0.15	15.0	0.31	46.3	124
RED ♂	21	5.3	−0.9	−4.3	0.3	12.5	N/A	N/A	N/A	N/A	**0.22***	25.9	**0.63***	70.8	123
TWIN ♂	−33	−11.0	−4.5	−20.2	0.1	4.5	0.13	11.2	0.12	12.8	0.13	12.6	0.06	7.7	121
MET ♀	−17	−3.9	−3.8	−18.0	−0.4	−16.7	N/A	N/A	N/A	N/A	0.12	14.0	0.01	1.0	118
CUB ♀	59	24.8	−1.4	−9.6	0.1	7.7	0.18	14.2	0	0.0	0.08	7.5	0.14	15.7	N/A
D96N3 ♀	−6	−2.2	−1.2	−7.4	0.3	15.8	0.11	8.7	−0.09	−7.1	**0.46***	51.7	−0.02	−1.8	N/A
D9603 ♀	16	8.6	−1.2	−8.9	0.2	11.1	0.16	12.7	0.02	1.5	0.02	2.0	0.06	5.8	N/A
JENNY ♀	57	22.8	2.6	15.4	0.3	17.6	−0.19	−12.4	0.02	1.4	0.12	13.3	0.18	15.1	100
SALLY ♀	62	27.1	4.3	33.3	−0.1	−5.3	−0.06	−4.9	−0.06	−4.8	−0.05	−4.2	−0.15	−12.8	100
PIO ♀	−2	−0.6	−2	−8.6	−0.3	−13.0	0.14	11.1	−0.1	−6.5	−0.12	−9.8	−0.05	−4.1	100
NEER ♀	−43	−11.9	−4.9	−24.0	−0.1	−3.8	0.2	20.0	0.08	6.9	0.01	0.9	0.09	9.3	100
SEA ♀	19	5.4	−1.1	−6.0	0.1	5.9	−0.23	−15.2	−0.1	−7.7	0.04	4.3	−0.11	−10.6	100
RIKKI ♀	78	24.0	0.8	5.0	−0.3	−17.0	−0.16	−14.0	−0.07	−6.0	0.15	19.0	0.15	21.0	100
OPRAH ♀	15	4.0	−1.4	−7.0	0.2	10.0	−0.08	−6.0	0.03	3.0	0.12	14.0	0.05	6.0	100

h after release from the vessel. As individual animals might respond differently to translocation, we examined the potential effect of translocation on the behavior of each individual animal by comparing the last pre-exposure dive to the mean of all the dives that occurred during the 18-h control period measured for each animal.

F. Acoustic analysis

Acoustic analysis was carried out using CANARY 1.2 (Cornell University, Ithaca, NY) on a Macintosh 8100 computer. Sound data were calibrated and then examined to exclude sections with high levels of flow noise due to a fast swimming seal. These sound files were bandpass filtered between 60 and 90 Hz (frequency range of interest) and then mean SPL measurements were made directly from the waveform. Maximum and minimum sound-pressure levels were averaged over a 5-s period. Depths associated with maximum and minimum levels were noted by comparing the time of the sound measurement with the dive record. Ambient levels before and after each ATOC transmission were determined by taking the lowest ambient sound levels received at the animal during the dive directly before and after ATOC transmissions. Lowest levels generally occurred during periods when the animal was not swimming (often during the glide phase of the dive). This provided estimates of ambient noise unaffected by the flow noise caused by swimming. In addition to measurements of SPL we were able to acoustically measure the breathing rate for eight of the animals before, during, and after exposure to the ATOC sound source (Le Boeuf *et al.*, 2000b).

TABLE II. The SPL recorded for each animal before, during and after an ATOC sound transmission is provided. Where available, the depth that corresponded to each SPL measurement is given in parentheses under the SPL measurement.

	Sound-pressure level (dB *re*: 1 μPa)			
		During ATOC transmission		
Seal	Prior to ATOC transmission	Low	High	After ATOC transmission
D96N2	102 (430 m)	123 (55 m)	135 (25 m)	104 (289 m)
D9601	104 (284 m)	126 (30 m)	137 (24 m)	103 (243 m)
D96N1	102 (216 m)	119 (75 m)	132 (20 m)	103 (268 m)
Card	105 (268 m)	124 (71 m)	131 (7 m)	103 (218 m)
Sox	102 (320 m)	122 (68 m)	132 (28 m)	104 (290 m)
Marlin	99 (355 m)	120 (44 m)	130 (8 m)	102 (347 m)
Met	87	110 (323m)	118 (67m)	89
Red	94	118 (251m)	123 (89m)	98
Expo	99	115 (112m)	124 (75m)	97
Twin	95	114 (2 m, 54 m)	121 (71 m)	93 (110 m)
Astro	98	120	126	96
B961	100	123	133	99
B962	99	115	124	98
CAP 962	98	118	126	99
Mean=	99	119	128	99
s.d.=	4.6	4.4	5.6	4.4

III. RESULTS

Data on diving behavior were obtained from 24 animals, 17 of which swam past the ATOC sound source when it was operating on a normal transmission schedule and 7 when the source was not operating at all while the animals were in the water (Table I). Received SPL measurements were obtained for 14 of the animals exposed to the ATOC sound source (Table II). An example of the data collected, including diving pattern and swim speed coupled with acoustic measurements of the ATOC transmission, is provided in Fig. 3. The highest SPLs measured during transmissions ranged from 118 to 137 dB *re*: 1 μPa for 60–90 Hz compared to ambient levels of 87–107 dB (60–90 Hz) (Table II). The highest exposure level averaged 128 dB for all subjects, which gave a mean increase of 29 dB over ambient SPLs. In no case did an animal end its dive or show any other obvious change in behavior (Fig. 4) in conjunction with a transmission. Other sounds recorded from the seals included a singing humpback whale with a broadband received level of 126 dB *re*: 1 μPa.

For control animals, the exposure dive was the dive that occurred 1 h after release from the boat. During the pre-exposure period, one control animal, NEER, and four experimental animals, D96N1, D9601, RED, and MET had at least one dive parameter that differed by more than one s.d. from the 18-h control period (Table I). Specifically, the control animal NEER exhibited a pre-exposure dive that had a shorter duration and surface interval, with a greater descent velocity, ascent velocity, and descent rate. Among the experimental animals, D96N1 exhibited shorter dive durations, D9601 exhibited a deeper dive and faster descent rate and velocity but a slower ascent velocity and rate, RED exhibited a faster descent rate, and MET exhibited a deeper, shorter dive with a faster descent rate. This implies that the animals had not assumed a normal diving pattern prior to exposure (Fig. 4). Two of these seals, D96N1 and D9601, were only in the water for 31 and 38 min prior to exposure, respectively (Table I, Fig. 4). All other animals exhibited normal diving behavior prior to exposure, and for these animals it was assumed that differences in the exposure dive were due to the operation of the ATOC source. Changes in diving behavior of translocated animals during exposure are provided for each behavioral measure for each animal in Table I. The behavior during exposure dives was within 2 standard deviations of the 18-h mean diving pattern for all animals.

The most sensitive component of diving behavior to ATOC sound transmissions was descent rate. Descent rate changed by more than ± 1 s.d. from each animal's 18-h mean in 9 of 14 seals. Dive depth and descent velocity increased by more than ± 1 s.d. in three animals, ascent velocity decreased in two animals, ascent rate increased in one animal and decreased in another, and dive duration decreased in only one animal. None of the animals showed a deviation by more than ± 1 s.d. in surface interval. For comparison, 31.7% of the observations (i.e., about 4 out of 14 animals for

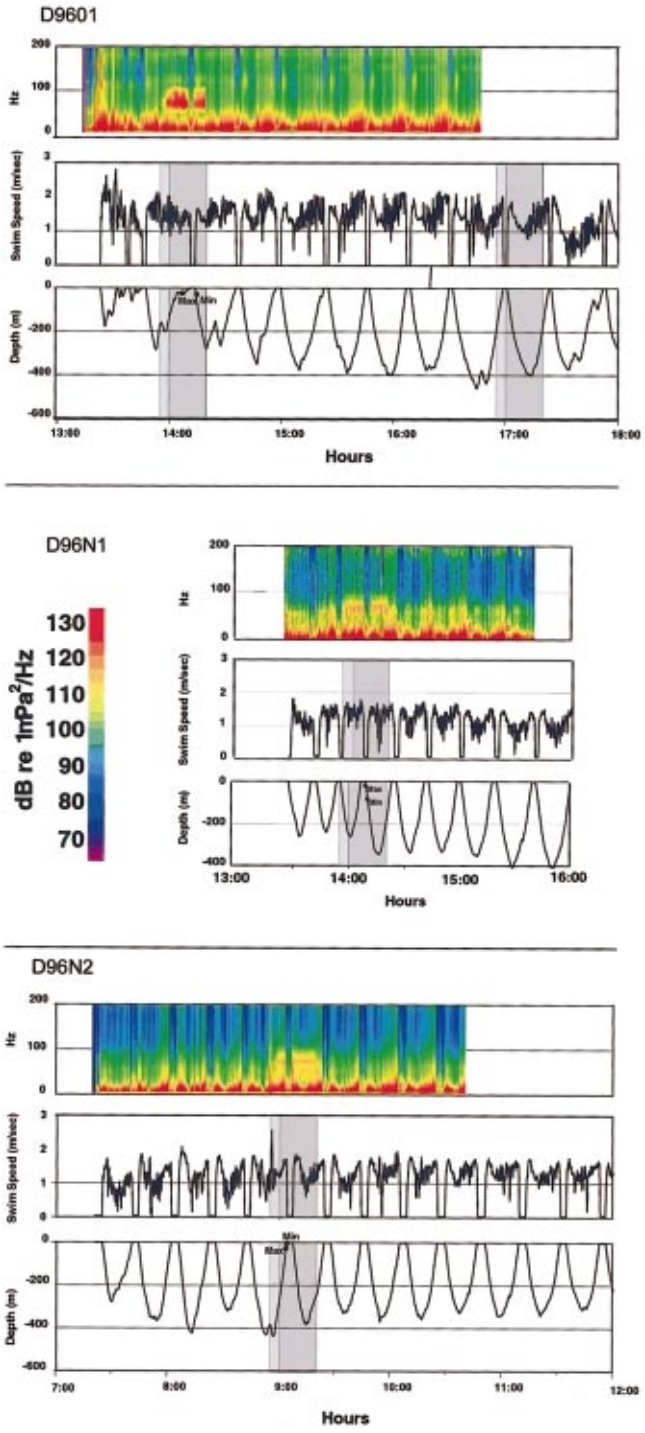


FIG. 3. Three examples of the data collected on juvenile elephant seals that passed near the ATOC sound sources. For each seal the top chart is a sound spectrogram, the middle chart is the swim-speed record, and the bottom chart the time-depth record. The 5-min ramp-up period is shaded in light gray and the 20-min full power operational period in darker gray. The 5-min ramp started 5 min prior to the hour and full power occurred on the hour. Points on the time-depth record delineate where the various SPL measurements were taken from the spectrogram. For each animal we measured the integrated SPL (while underwater) over the entire period of exposure, along with the maximum and minimum levels (5-s periods). These levels were: integrated 133, max 137, min 126 dB re: 1 μPa for D9601; integrated 127, max 132, min 119 dB re: 1 μPa for D96N1; and integrated 126, max 135, min 123 dB re: 1 μPa for D96N2.

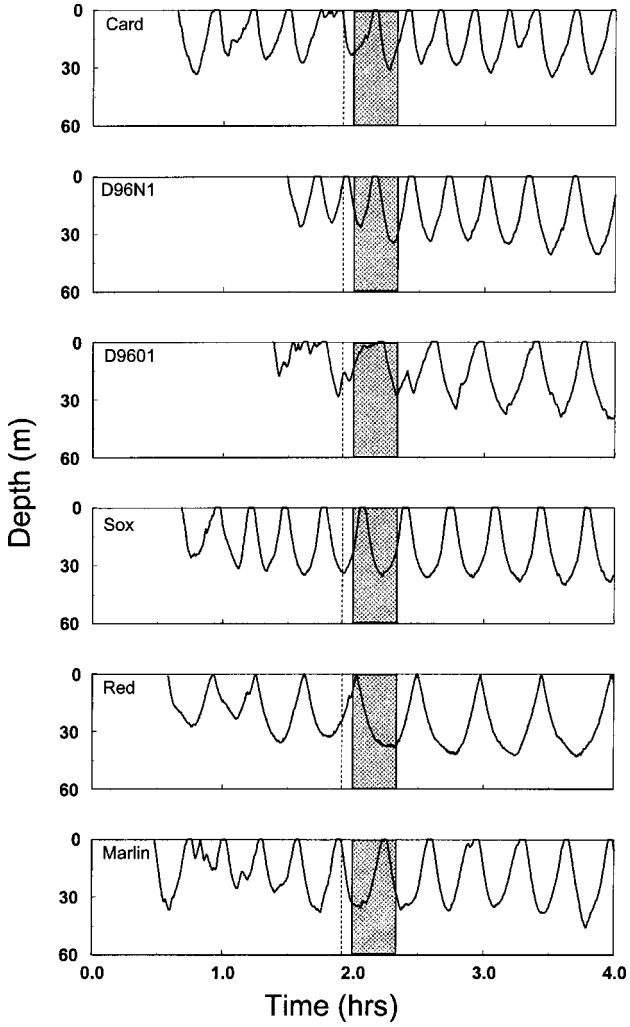


FIG. 4. Dive patterns of the six animals that showed the greatest changes in behavior. The dotted line shows the beginning of the 5-min ramp up and the shaded area shows the period of full power transmission. The first dive corresponds to the animal leaving the boat.

each behavioral measure) would be expected to exceed 1 s.d. for a Gaussian distribution. In only one animal (D96N1) was there a deviation by more than ± 1 s.d. in depth or duration, without a comparable deviation in descent speed (Table I). There was a highly significant positive correlation between descent rate and SPL (Fig. 5). This correlation was evident whether the animals showing a pre-exposure effect were excluded ($r^2=0.61$, $p=0.007$) or included ($r^2=0.42$, $p=0.012$) in the analysis. None of the other variables exhibited a significant correlation between SPL and the deviation from control. There was no significant difference in surface breathing rate for the eight animals for which we were able to obtain these data. The surface-breathing rate was 15.2 ± 2.0 breaths per minute (BPM) prior to exposure, 17.1 ± 3.8 BPM during exposure, and 13.4± 1.8 BPM after exposure.

Ten of the 17 seals used in the experiment had exposure dives with at least one dive parameter which differed by

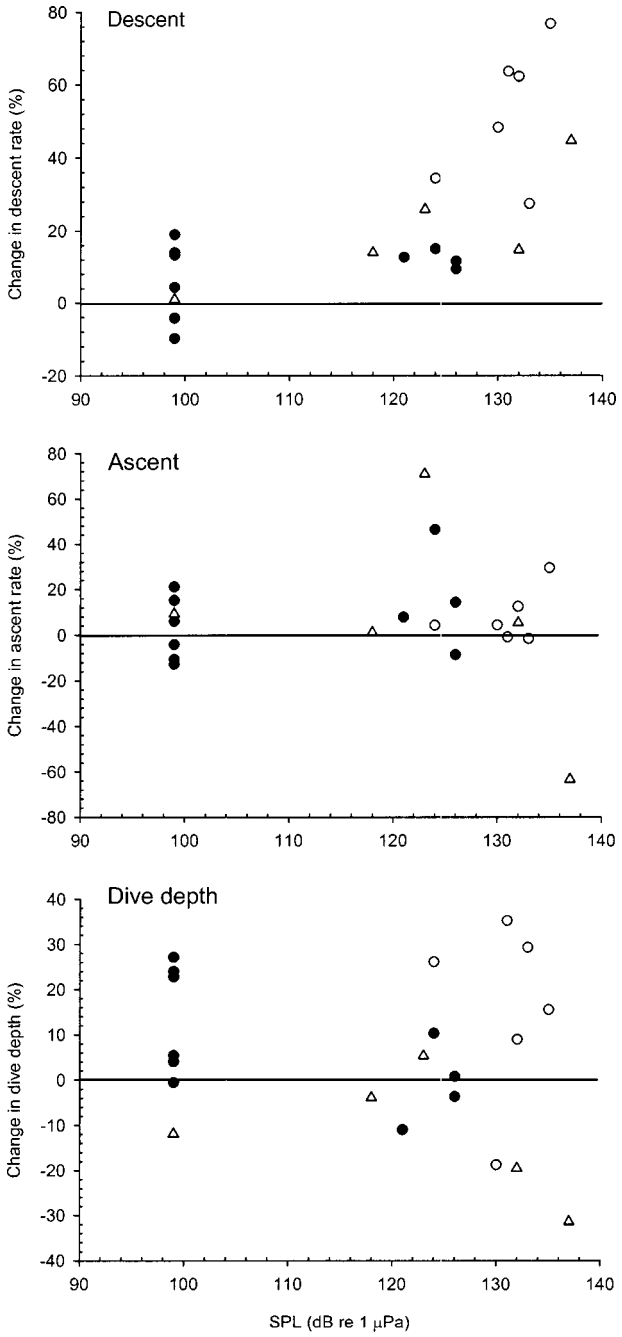


FIG. 5. Relative change in descent rate, ascent rate, and dive depth as a function of the SPL received by the animal, for all subjects used in the translocation studies. Solid symbols represent animals for which the plotted variable during the exposure dive was within ±1 s.d. of the 18-h mean. Open symbols represent animals whose change in ascent rate, descent rate or dive depth was greater than ±1 s.d. of the 18-h mean. Circles designate animals that had a normal pre-exposure dive. Triangles are for animals whose pre-exposure dive varied by more than ±1 s.d. from the 18-h mean. The correlation between change in descent rate and SPL was highly significant regardless of whether animals with a pre-exposure effect were excluded ($r^2=0.61$, $p=0.007$) or included ($r^2=0.42$, $p=0.012$) in the analysis.

more than ±1 s.d. from the 18-h mean (Table I). The relative number of behavioral parameters that exhibited a deviation greater than ±1 s.d. from the 18-h mean increased as a function of the SPL ($r^2=0.4$, $p=0.015$; Fig. 6).

Two approaches were used to look for statistically significant responses using combined data from all of the animals. To test for an effect on the exposure dive, repeated measures ANOVA was performed. The dive variables during the exposure dive were compared to the mean values for the 18-h postexposure period for each animal. This approach revealed significant differences between the exposure dives and nonexposure dives in descent rate ($F_{1,13}=32.5$, $p<0.01$) and descent velocity ($F_{1,13}=7.7$, $p<0.01$). For the control animals (ATOC source off) the dive 1 h after entry into the water was used as if it were the ''exposure dive.'' This approach yielded no significant differences for the control animals. To examine the effects of the short interval between release and exposure, a similar comparison was made between the exposure dive and the mean values of the pre-exposure period. This analysis revealed no significant differences between pre-exposure and exposure using all of the exposed or control animals. Note that some individual animals did show differences (see Fig. 5).

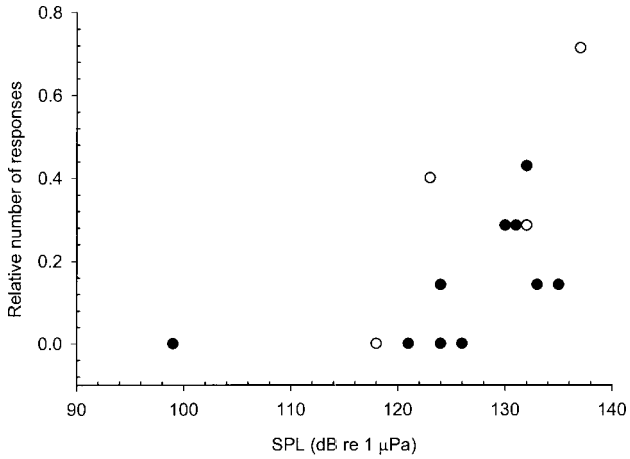


FIG. 6. The relative number of dive variables that varied by more than ±1 s.d. from their respective 18-h means are plotted as a function of the ATOC source SPL as received at the animal. The number of responses was normalized relative to whether there were five or seven variables measured. This was necessary because swim speed was not measured on all animals (those carrying CAP tags). Solid symbols designate animals that had a normal pre-exposure dive. Open circles are for animals that exhibited a pre-exposure dive with variables that differed by more than ±1 s.d. from the 18-h mean. There was a highly significant relationship between these variables ($r^2=0.4$, $p=0.015$).

To test for effects in the hour of diving after exposure using all of the exposure subjects, an ANOVA comparing the hour of postexposure diving to an hour of diving 18 h after exposure was performed. This ANOVA was run as a general linear model with exposure status (status) and individual (indiv) as effects terms. An interaction term was included in the model (status*indiv) to represent the individual differences in exposure levels. Pseudoreplication was avoided by nesting replicate dives during the hour sampled within the individual animals. The effect of status within this model was tested. This approach revealed significant differences in three of the measured response variables. Descent rate ($F_{1,15}=26.0$, $p<0.01$) and descent velocity ($F_{1,13}=13.8$, $p<0.01$) were significantly greater in the hour after exposure. Ascent rate was significantly lower during the hour after exposure ($F_{1,15}=10.6$, $p<0.01$). In the control animals (ATOC source off) the second hour of diving after water entry was consid-

TABLE III. Power analysis $(1-\beta)$ for multiway anova of response variables from a model with exposure status, individual, and (individual $\times$exposure status) as factors.

Variable	Exposure	Control
Depth (m)	0.28	0.27
Duration (min)	1.00	0.98
Surface interval (min)	0.07	0.53
Descent speed (m/s)	0.96	0.87
Ascent speed (m/s)	0.86	0.67
Total speed (m/s)	0.97	0.84
Descent rate (m/s)	0.93	0.87
Ascent rate (ms)	0.84	0.81

ered the postexposure period. No significant differences were evident among the control animals.

To assess the ability to detect significant differences in both exposure and control animals, power analysis was performed for the multiway ANOVA (Table III). This analysis revealed that our sample lacked sufficient power to detect differences in maximum depth and surface interval. For all variables that showed a significant effect in the exposure animals there was sufficient power to detect a difference in the control animals.

IV. DISCUSSION

There were three major findings with respect to the response of elephant seals to the ATOC sound source. The first was that the highest received level at the animal did not exceed 137 dB *re*: 1 μPa. This is important, considering that the seals were released so that they would have the highest probability of swimming directly over the ATOC sound source. Second, we did not observe cessation of diving in response to ATOC in any seal studied; and third, we were able to measure only subtle changes in the diving behavior of animals swimming past the ATOC source. With the exception of sperm and beaked whales, elephant seals are the only marine mammal known to be capable of routinely diving to the depth of the ATOC sound projector. This is important considering that SPLs at the seal were always below 137 dB *re*: 1 μPa and that only 7 out of 14 animals reached maximum exposures at or above 130 dB *re*: 1 μPa. Given that the experimental manipulation was designed to achieve the highest possible exposure, and that most other species cannot dive like elephant seals, it is likely that they would be exposed to SPLs lower than or at least no greater than those experienced by the elephant seals in this study. Unfortunately, ARGOS satellite uplinks were too infrequent (1–4 per day) and imprecise (1–10 km) to provide useful information on the return track of the animal. However, some estimate of the proximity of the elephant seals to the sound source can be made by comparison with empirical measurements of the ATOC sound field made with calibrated hydrophones (High Tech, Inc. HT1-SSQ-41b; 10 Hz–30 kHz $\pm$3 dB) at depths between 50–200 m (Gedamke and Costa, 1997). These data indicate that exposure to 120 dB *re*: 1 μPa or greater required the animals to be within 20 km of the source, while exposure to 130 dB *re*: 1 μPa or greater required the animals to be within 4 km of the source.

In order to put the observed behavioral response in perspective we must consider how an elephant seal might respond to a perceived threat or how it might attempt to ''escape'' exposure to the ATOC sound source. If an elephant seal perceived the ATOC sound source as a threat, we might find that the seal dove deeper or remained underwater longer. Such a response was recorded for an elephant seal swimming in shallow water as a boat passed overhead (Burgess *et al.*, 1998). This animal stopped swimming and stayed near the bottom as the boat passed overhead. Both northern and southern elephant seals have been observed to undergo prolonged dives that last more than 90 min. These dives are extremely rare, occurring only once or twice over an entire 3–9-month diving record and are not observed in all individuals. It has been suggested that these extremely long dives are a response to the presence of a potential predator like a white shark or killer whale (Le Boeuf and Crocker, 1996). These putative ''escape'' dives are very distinctive and are several times longer than the typical dive pattern. Similarly, if elephant seals perceived the ATOC sound source as dangerous they might have stopped swimming and extended their dive until the signal ceased. Even juvenile elephant seals are capable of dives lasting longer than the 20-min ATOC broadcast. At no time during exposure or at any other time during any dive record did we observe a diving pattern that was even remotely similar to these escape dives. It is important to note that none of these responses was observed in any of the 17 animals exposed to the ATOC sound source.

Although we did not observe obvious changes in the diving behavior of elephant seals, we were able to measure subtler changes. The measure of diving behavior that exhibited the greatest sensitivity to the ATOC sound source was the rate of descent. Rate of descent changed in nine animals and was significantly correlated with SPL (Fig. 5). The corresponding metric ascent rate only changed in two animals, and there was no significant relationship between the change in ascent rate and SPL. However, when the more powerful statistical analysis using the general linear model was carried out, differences in a number of response variables during exposures were detected. Descent rate, descent velocity, and ascent rate all changed in response to exposure or possibly due to the short interval after release. This is consistent with an escape response where the animal prefers to avoid the surface or seeks refuge at depth. Alternatively, if the elephant seal was trying to avoid the ATOC sound source, it could cease diving and remain at the surface and effectively reduce its exposure to the ATOC signal. Within $\frac{1}{4}$ wavelength (5 m for a 75-Hz signal) of the ocean surface (pressure-relief surface) the sound level is often 10–30 dB lower (Jensen, 1981). At no time did any seal cease diving or exhibit any other dramatic change in behavior during exposure to the ATOC signal.

It might seem odd that an animal would increase descent rate or decrease its ascent rate when this would prolong its exposure to the sound source. It is necessary to remember

that this is the predicted antipredator response for this animal. Given the relatively small interaural distance of an elephant seal relative to a 75-Hz signal with a 20-m wavelength, it is unlikely that an elephant seal has the capability of localizing a sound of this frequency and wavelength underwater. Given that the animal may lack information on the location or direction of the sound source, its response may be simply a ''startle response,'' which for elephant seals is to dive.

An alternative explanation of the minimal response observed in northern elephant seals is that they are not capable of hearing the ATOC signal. Recent data on the hearing sensitivity of a captive northern elephant seal indicates a hearing threshold of 98.3 dB *re*: 1 μPa for a 75-Hz tone (Kastak and Shusterman, 1998). The peak sensitivity for this animal was 59 dB *re*: 1 μPa for a 6300-Hz tone (Kastak and Shusterman, 1998). The mean fundamental frequencies of airborne calls of northern elephant seals are in the range 147–334 Hz for adult males (Le Boeuf and Petrinovich, 1974) and 500–1000 Hz for adult females (Bartholomew and Collias, 1962). Furthermore, audiometric studies suggest that pinnipeds in general and elephant seals in particular are relatively good at detecting tonal signals over masking noise (Southall *et al.*, 2000). Given that ambient sound was on average 99 dB *re*: 1 μPa and that the lowest SPL of the ATOC signal as received at the seals' location was 110 dB *re*: 1 μPa with an average low of 119 dB *re*: 1 μPa, we are confident that these animals were capable of hearing the ATOC signal.

To our knowledge this is the first study to quantify behavioral responses of an animal underwater with simultaneous measurements of SPL at the animal. Past studies of the effects of sound on animals have relied on acoustic propagation models and/or measurements of the sound field in the general vicinity of the animal to estimate the level of exposure (Richardson *et al.*, 1990, 1995; Richardson and Würsig, 1997; Würsig *et al.*, 1998; Frankel and Clark, 1998, 2000; Erbe and Farmer, 2000). The problem with this approach is that sound fields are uneven and one can only approximate the sound frequency and intensity to which the animal is exposed. These studies were also limited to measurements such as avoidance of the sound source, changes in respiration rate, fluke pattern, and dive duration. In our study no animal changed its surface interval or respiration rate, and only one animal decreased its dive duration. All of the other behavioral modifications observed here could not have been measured in these previous studies. However, observations of vocal behavior can be carried out while the animal is underwater (Miller *et al.*, 2000), but are still limited by the imprecision of estimates of SPL and animal location.

The Marine Mammal Protection Act makes it illegal to harass or otherwise harm marine mammals in the United States unless a specific authorization has been given. The problem has been in defining what constitutes harassment, and more importantly, defining when that harassment (or the change in behavior as a result of the harassment) impacts sufficient numbers of individuals that it becomes detrimental to the population. As a result of this gray area, there is often a poor link between the laws governing how marine mammals are managed and the actual biological issues underlying those laws. Although we were able to record small changes in the diving behavior of the juvenile elephant seals, it is unlikely that these changes would have any lasting impacts on the animals as they migrate past the ATOC sound source. First, changes in behavior were observed only in animals that were relatively close to the sound source. Second, all elephant seals passing through this area are migrating to or from their distant feeding grounds in the North Pacific Ocean, and thus would not be spending much time near the PIONEER SEAMOUNT (Le Boeuf *et al.*, 2000a). It is important to recognize that although we measured a small change in behavior of migrating animals, the significance of these changes relative to such critical activities as feeding or reproduction is likely to be minimal.

This study has shown the potential for using acoustic data loggers coupled with time-depth recorders to record the response of marine mammals to acoustic stimuli. This approach proved quite effective for use with elephant seals. Further development of the CAP tag is underway (Burgess and Tyack, 2000) and smaller DAT recorders are now available. Smaller devices coupled with recent advances in attachment techniques (Croll *et al.*, 1998; Hooker and Baird, 1999; Nowacek *et al.*, 2001) will allow these techniques to be applied to cetaceans as well as to smaller pinnipeds.

ACKNOWLEDGMENTS

We thank Chris Clark, Dawn Goley, Don Croll, and Luisa Williams for assistance with the ATOC Marine Mammal Research Program. We especially thank Walter Munk and Peter Worcester for their support throughout the ATOC project. James Ganong and Carl Haverl assisted with data analysis and graphics. John Richardson and an anonymous reviewer provided insightful comments on the manuscript. Clairol Inc supplied bleach or hair dye for marking the seals. The California Department of Parks and Recreation allowed access to Año Nuevo State Reserve. This research was conducted under permit No. 836 from the National Marine Fisheries Service. This work was funded by the Office of Naval Research (ONR N00014-94-1-0455) and the Scripps Institution of Oceanography Acoustic Thermometry of Ocean Climate program via subcontracts from grants ARPA MDA 972-93-1-003 and ONR N00014-94-1-0692.

The ATOC Consortium (**1998**). ''Ocean climate change: Comparison of acoustic tomography, satellite altimetry, and modeling,'' Science **28**, 1327–1332.

Au, W. W. L., Floyd, R. W., Penner, R. H., and Murchison, A. E. (**1974**). ''Measurement of echolocation signals of the Atlantic bottlenose dolphin, *Tursiops truncatus* Montagu, in open waters,'' J. Acoust. Soc. Am. **56**, 1280–1290.

Au, W. W. L., Nachtigal, P. E., and Pawloski, J. L. (**1997**). ''Acoustic effects of the ATOC signal (75 Hz, 195 dB) on dolphins and whales,'' J. Acoust. Soc. Am. **101**, 2973–2977.

Bartholomew, G. A., and Collias, N. E. (**1962**). ''The role of vocalization in the social behavior of the northern elephant seal,'' Anim. Behav. **10**, 7–14.

Blackwell, S. B., Haverl, C. A., Le Boeuf, B. J., and Costa, D. P. (**1999**). ''A method for calibrating swim speed recorders,'' Marine Mammal Sci. **15**, 894–905.

Burgess, W., Tyack, P., Le Boeuf, B. J., and Costa, D. P. (**1998**). ''An intelligent acoustic recording tag: First results from free-ranging northern elephant seals,'' Deep-Sea Res., Part II **45**, 1327–1351.

Burgess, W., and Tyack, P. (**2000**). Personal communication.
Costa, D. P. (**1993**). ''The secret life of marine mammals: New tools for the study of their biology and ecology,'' Oceanography **6**, 120–128.
Crocker, D. E., Le Boeuf, B. J., and Costa, D. P. (**1997**). ''Drift diving in female northern elephant seals: Implications for food processing,'' Can. J. Zool. **75**, 27–39.
Croll, D. A., Tershy, B., Hewitt, R., Demer, D., Fiedler, P., Smith, S., Armstrong, W., Popp, J., Kiekhefer, T., Lopez, V., and Urban, J. (**1998**). ''An integrated approach to the foraging ecology of marine birds and mammals,'' Deep-Sea Res., Part II **45**, 1353–1371.
DeLong, R. L., Stewart, B. S., and Hill, R. D. (**1992**). ''Documenting migrations of northern elephant seals using day length,'' Marine Mammal Sci. **8**, 155–159.
Erbe, C., and Farmer, D. M. (**2000**). ''Zones of impact around icebreakers affecting beluga whales in the Beaufort Sea,'' J. Acoust. Soc. Am. **108**, 1332–1340.
Fletcher, S., Le Boeuf, B. J., Costa, D. P., and Tyack, P. L. (**1996**). ''Onboard acoustic recording from diving elephant seals,'' J. Acoust. Soc. Am. **100**, 2531–2539.
Frankel, A. S., and Clark, C. W. (**1998**). ''Results of low-frequency playback of M-sequence noise to humpback whales, *Megaptera novaeangliae*, in Hawaii,'' Can. J. Zool. **76**, 521–535.
Frankel, A. S., and Clark, C. W. (**2000**). ''Behavioral responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals,'' J. Acoust. Soc. Am. **108**, 1930–1937.
Green, D. M., DeFerrari, H. A., McFadden, D., Pearse, J. S., Popper, A. N., Richardson, W. J., Ridgway, S. H., and Tyack, P. L. (**1994**). *Low-frequency Sound and Marine Mammals: Current Knowledge and Research Needs* (National Research Council, Washington, D.C.).
Gordon, J., and Moscrop, A. (**1996**). ''Underwater noise pollution and its significance for whales and dolphins,'' in *The Conservation of Whales and Dolphins: Science and Practice*, edited by M. P. Simmonds and J. D. Hutchinson (Wiley, New York), pp. 281–320.
Hindell, M. A., Slip, D. J., and Burton, H. R. (**1991**). ''The diving behaviour of adult male and female southern elephant seals, *Mirounga leonina* (Pinnipedia, Phocidae),'' Aust. J. Zool. **39**, 595–619.
Hooker, S. K., and Baird, R. W. (**1999**). ''Deep-diving behavior of the northern bottlenose whale, *Hyperoodon ampullatus* (Cetacea: Ziphiidae),'' Proc. R. Soc. London, Ser. B **266**, 671–676.
Howe, B. M. (**1996**). ''Acoustic Thermometry of Ocean Climate (ATOC): Pioneer Seamount Source Installation,'' Tech. Memo. Applied Physics Laboratory-University of Washington, TM 3-96.
Jensen, F. B. (**1981**). ''Sound propagation in shallow water: A detailed description of the acoustic field close to surface and bottom,'' J. Acoust. Soc. Am. **70**, 1397–1406.
Kastak, D., and Schusterman, R. J. (**1998**). ''Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology,'' J. Acoust. Soc. Am. **103**(4), 2216–2228.
Le Boeuf, B. J., and Petrinovich, L. F. (**1974**). ''Dialects in northern elephant seals, *Mirounga angustirostris*: Origin and reliability,'' Anim. Behav. **22**, 656–663.
Le Boeuf, B. J., Costa, D. P., Huntley, A. C., and Feldkamp, S. D. (**1988**). ''Continuous, deep diving in female northern elephant seals, *Mirounga angustirostris*,'' Can. J. Zool. **66**, 446–458.
Le Boeuf, B. J., and R. M. Laws. (**1994**). *Elephant Seals: Population Ecology, Behavior and Physiology* (University of California, Berkeley), pp. 193–205.
Le Boeuf, B. J., Morris, P. A., Blackwell, S. B., Crocker, D. E., and Costa, D. P. (**1996**). ''Diving behavior of juvenile northern elephant seals,'' Can. J. Zool. **74**, 1632–1644.
Le Boeuf, B. J., and Crocker, D. E. (**1996**). ''Diving behavior of elephant seals: Implications for predator avoidance,'' in *Great White Shark: The Biology of Carcharodon Charcharias*, edited by P. Klimley (Academic, San Diego), pp. 193–205.
Le Boeuf, B. J., Crocker, D. E., Costa, D. P., Blackwell, S. B., Webb, P. M., and Houser, D. S. (**2000a**). ''Foraging ecology of northern elephant seals,'' Ecol. Monogr. **70**, 353–382.
Le Boeuf, B. J., Crocker, D. E., Grayson, J., Gedamke, J., Webb, J. M., Blackwell, S. B., and Costa, D. P. (**2000b**). ''Respiration and heart rate at the surface between dives in northern elephant seals,'' J. Exp. Biol. **203**, 3265–3274.
Mate, B. R., Lagerquist, B. A., and Calambokidis, J. (**1999**). ''Movements of north Pacific blue whales during the feeding season off southern California and their southern fall migration,'' Marine Mammal Sci. **15**, 1246–1257.
Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P. L. (**2000**). ''Whales songs lengthen in response to sonar,'' Nature (London) **405**, 903.
Munk, W. H., and Forbes, A. M. G. (**1989**). ''Global ocean warming: An acoustic measure?'' J. Phys. Oceanogr. **10**, 1765–1778.
Munk, W. H., Spindel, R. C., Beggeroer, A., and Birdsall, T. G. (**1994**). ''The Heard Island feasibility test,'' J. Acoust. Soc. Am. **96**, 2330–2342.
Nowacek, D. P., Johnson, M. P., Tyack, P. L., Shorter, K. A., McLellan, W. A., and Pabst, D. A. (**2001**). ''Buoyant balanenids: The ups and downs of buoyancy in right whales,'' Proc. R. Soc. London, Ser. B **268**, 1811–1816.
Oliver, G. W., Morris, P. A., Thorson, P. H., and Le Boeuf, B. J. (**1998**). ''Homing behavior of juvenile northern elephant seals,'' Marine Mammal Sci. **14**, 245–256.
Potter, J. R. (**1994**). ''ATOC: Sound policy or enviro-vandalism? Aspects of a modern mediafueled policy issue,'' J. Environ. Develop. **3**, 47–76.
Richardson, W. J., Greene, C. R. Jr., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic Press, San Diego).
Richardson, W. J., and Würsig, B. (**1997**). ''Influences of man-made noise and other human actions on cetacean behavior,'' Mar. Freshwater Behav. Physiol. **29**, 183–209.
Richardson, W. J., Wursig, B., and Greene, C. R. (**1990**). ''Reactions of bowhead whales, *Balaena mysticetus*, to drilling and dredging noise in the Canadian Beaufort Sea,'' Mar. Environ. Res. **29**, 135–160.
Southall, B. L., Schusterman, R. J., and Kastak, D. (**2000**). ''Masking in three pinnipeds: Underwater, low-frequency critical ratios,'' J. Acoust. Soc. Am. **108**, 1322–1326.
Stewart, B. S., and DeLong, R. L. (**1991**). ''Diving patterns of northern elephant seal bulls,'' Marine Mammal Sci. **7**, 369–384.
Thompson, T. J., Winn, H. E., and Perkins, P. J. (**1979**). ''Mysticete sounds,'' in *Behavior of Marine Animals, Vol. 3: Cetaceans*, edited by H. E. Winn and B. L. Olla (Plenum, New York), pp. 403–431.
Tyack, P. L., and Whitehead, H. (**1983**). ''Male competition in large groups of wintering humpback whales,'' Behaviour **83**, 132–154.
Tyack, P. L., and Clark, C. W. (**2000**). ''Communication and acoustic behavior of dolphins and whales,'' in *Springer Handbook of Auditory Research; Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 156–224.
Van Parijs, S. M., Hastie, G. D., and Thompson, P. M. (**1999**). ''Geographical variation in temporal and spatial vocalization patterns of male harbour seals in the mating season,'' Anim. Behav. **58**, 1231–1239.
Watkins, W. A., Daher, M. A., Reppucci, G. M., George, J. E., Martin, D. L., DiMarzio, N. A., and Gannon, D. P. (**2000**). ''Seasonality and distribution of whale calls in the North Pacific,'' Oceanography **13**, 62–67.
Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). ''The 20-Hz signals of finback whales (*Balaenoptera physalus*),'' J. Acoust. Soc. Am. **82**, 1901–1912.
Worcester, P. F., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Howe, B. M., Mercer, J. A., Spindel, R. C., Colosi, J. A., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (**1999**). ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–201.
Würsig, B., Lynn, S. K., Jefferson, T. A., and Mullin, K. D. (**1998**). ''Behavior of cetaceans in the northern Gulf of Mexico relative to survey ships and aircraft,'' Aquat. Mam. **24**, 41–50.

Simulation of ultrasonic focus aberration and correction through human tissue

Makoto Tabei
Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

T. Douglas Mast[a)]
Applied Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16801

Robert C. Waag
Departments of Electrical and Computer Engineering and Radiology, University of Rochester, Rochester, New York 14627

(Received 29 January 2002; revised 26 October 2002; accepted 4 November 2002)

Ultrasonic focusing in two dimensions has been investigated by calculating the propagation of ultrasonic pulses through cross-sectional models of human abdominal wall and breast. Propagation calculations used a full-wave *k*-space method that accounts for spatial variations in density, sound speed, and frequency-dependent absorption and includes perfectly matched layer absorbing boundary conditions. To obtain a distorted receive wavefront, propagation from a point source through the tissue path was computed. Receive focusing used an angular spectrum method. Transmit focusing was accomplished by propagating a pressure wavefront from a virtual array through the tissue path. As well as uncompensated focusing, focusing that employed time-shift compensation and time-shift compensation after backpropagation was investigated in both transmit and receive and time reversal was investigated for transmit focusing in addition. The results indicate, consistent with measurements, that breast causes greater focus degradation than abdominal wall. The investigated compensation methods corrected the receive focus better than the transmit focus. Time-shift compensation after backpropagation improved the focus from that obtained using time-shift compensation alone but the improvement was less in transmit focusing than in receive focusing. Transmit focusing by time reversal resulted in lower sidelobes but larger mainlobes than the other investigated transmit focus compensation methods. [DOI: 10.1121/1.1531986]

PACS numbers: 43.80.Qf, 43.20.Fn [FD]

I. INTRODUCTION

Simulation of large-scale ultrasonic propagation through realistic tissue structures has recently become feasible.[1–3] Computations of wavefront distortion produced by human abdominal wall[1,4] and breast tissue[5] models have shown the tissue models produce propagation effects similar to those measured[6–8] and has provided insight about the way various structures in tissue produce aberration.[4,9]

Simulations of ultrasonic focus aberration by tissue have previously employed models such as random phase screens[10–12] or homogeneous layers.[13–15] These models, however, do not incorporate detailed anatomic structure. Although wavefront distortion is known to limit focus and image quality,[4,10,16] further investigation of focusing is needed with more realistic models that explicitly include anatomic structure of tissue to extend current understanding.

Direct simulation of ultrasonic focusing through realistic tissue structures can elucidate the physical processes involved in ultrasonic image aberration. Of special interest is the effect of morphology on focus correction for synthetic (receive) focusing of the aberrated wavefronts in image formation[17–19] and for physical (transmit) focusing through tissue.[20–22] Although tissue inhomogeneities are known to cause wavefront distortion and, thus, focus degradation, the relationship between specific tissue structures and focus quality has not received much attention. Transmit focus correction, in particular, has received limited previous attention in the literature, so that realistic computations of aberration-corrected transmit focusing are needed to improve understanding of transmit focus correction.

This paper presents simulations of transmit and receive focusing through two-dimensional models of abdominal wall[1,4] and breast tissue. Cylindrical wavefronts aberrated by propagation through each tissue cross section are refocused with and without correction. Focus quality is described by metrics that quantify the focal width and the relative amount of energy outside the focal region. For receive focus correction, time-shift compensation in the receiving aperture[17–19] and after backpropagation from the aperture[23,24] were employed. For transmit focus correction, these methods as well as time reversal[20] were used.

The results indicate that the quality of corrected and uncorrected focus depends on tissue type as well as the method of correction. The relative performance of correction methods for transmit and receive focusing has been shown under directly comparable conditions. In general, breast tissue in the simulations caused greater focus aberration than

[a)]Present address: Ethicon Endo-Surgery, 4545 Creek Rd. ML 40, Cincinnati, OH 45242.

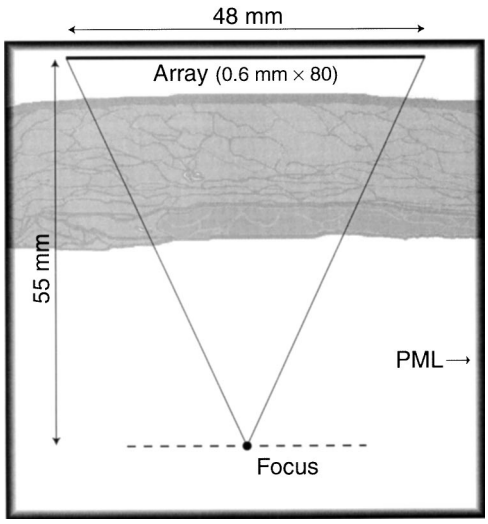


FIG. 1. Configuration for calculations of propagation. A virtual array, tissue model (abdominal wall A04a), and focal point are situated as shown in a 65.03×65.03 mm^2 region around which is a perfectly matched layer (PML) absorbing boundary.

abdominal wall tissue both before and after compensation. Examination of time-domain received, corrected, and focused wavefronts as well as pressure fields within tissue has provided insight into the physical basis, strengths, and limitations of each correction method. Time-shift compensation was more effective after backpropagation for both receive and transmit focusing, although this difference was smaller in the case of transmit focusing. Time reversal produced a good transmit focus due to the coherent point source employed, but was limited by spatial and temporal windowing of scattered signals as well as by frequency-dependent absorption.

II. METHODS

A. Wavefield simulations

All computations performed here used two-dimensional maps of human abdominal wall and breast tissue cross sections. The abdominal wall maps were those used in Ref. 1. The breast tissue maps were created using methods described in Ref. 1. These breast tissue maps were derived from regions of the breast where little parenchymal tissue was present so that the sections consisted primarily of fat, skin, and connective tissue. The maps represented the tissue as regions of a single tissue type: fat, muscle, or connective tissue.

The computations were performed using a virtual array, tissue map, and focal point configured as in Fig. 1. To obtain a distorted receive wavefront, a cylindrical pressure wave pulse with a sinusoidal time variation and a Gaussian envelope was propagated from a point 55 mm away to an 80-element virtual array of receivers that had a pitch of 0.60 mm and spanned 48 mm. At the start of the computations, the particle velocity was defined to be zero everywhere and the pressure distribution was defined to be

$$p(x,y;0) = e^{-[\pi b r/(2c_0)]^2/\ln(2)} J_0(2\pi f_0 r/c_0), \tag{1}$$

TABLE I. Physical properties at 37 °C for each tissue type employed in the simulations. "CT" denotes connective tissue including septa as well as connective structures within muscle layers.

	Physical parameter				
Tissue	Sound speed mm/μs	Density g/cc	Absorption dB/(cm×MHz)	$\tau_1 = 33$ ns $(\kappa_1/\kappa_\infty) \times 10^3$	$\tau_2 = 200$ ns $(\kappa_2/\kappa_\infty) \times 10^3$
Water	1.524	0.993	⋯	⋯	⋯
Reference	1.524	0.993	0.50	4.32	3.64
Fat	1.478	0.950	0.52	4.36	3.67
Muscle	1.547	1.050	0.91	7.98	6.72
Skin/CT	1.613	1.120	1.61	14.72	12.40

where $r = [(x-x_0)^2 + (y-y_0)^2]^{1/2}$, the source position is (x_0, y_0), f_0 and b are the pulse center frequency and -6 dB bandwidth, respectively, and $J_0(\cdot)$ is a zeroth-order Bessel function of the first kind. This specification of the initial values corresponds to the sum of an outgoing wave and an incoming wave each band limited in temporal frequency and centered at (x_0, y_0). The starting pressure is continuous because the singularity in the temporal-frequency domain Green's function associated with the outgoing wave and used in the superposition of temporal frequencies is canceled by a corresponding singularity associated with the incoming wave. In all the simulations, f_0 was 3 MHz and b was 1.8 MHz. The array of receivers was located about 8 mm from the skin surface in the tissue map. The straight-line tissue path length from the focal or source point to the array averaged 30 mm. The source was centered in the lateral span of the array and was at least 10 mm from the other surface of the tissue map in each case. Element directivity and cross talk were emulated by integrating the pressure field over the span of each element using a trapezoidal weighting that consisted of a 0.4 mm flat region at the center and 0.2 mm linear transition that overlapped with adjacent elements on both sides.

The maps of tissue were derived from cross-sectional images that were sampled on a uniform x-y grid at 0.084 67 mm intervals (300 pixels per in.). The same grid interval was used for all the computations. Each map was used for multiple simulations by choosing sections that filled the whole computational window with a uniform thickness central portion and included no more than 25% overlap between apertures.

The tissue properties employed in the simulations are summarized in Table I. Sound speed and density values are those employed in Ref. 1. Relaxation-process absorption was implemented using two relaxation processes with compressibility parameters κ_1, κ_2 and time constants τ_1 and τ_2. The compressibility parameters and the time constants were chosen to approximate a linear dependence of absorption on frequency over the pulse bandwidth by using the formula for frequency-dependent absorption given in Ref. 25. The relaxation time constants were defined by the relations

$$\tau_1 = 1/(6 f_{\max}) \tag{2a}$$

and

$$\tau_2 = 1/f_{\max}, \tag{2b}$$

where $f_{\max}$ is the nominal maximum frequency of interest. For a maximum nominal frequency that was 5.0 MHz in the simulations described here, the relaxation time constants are $\tau_1 = 33$ ns and $\tau_2 = 200$ ns. Given this choice of relaxation time constants, a reference frequency-dependent absorption of 0.50 dB/(cm·MHz) is best approximated (in a least-squares sense with an rms error of 0.02 dB/cm) for a reference density of 0.993 g/cc and a reference sound speed of 1524 mm/μs in the frequency range $1.0 < f < 5.0$ MHz by the compressibility parameters $\kappa_1 = 4.32 \times 10^{-3} \kappa_\infty$ and $\kappa_2 = 3.64 \times 10^{-3} \kappa_\infty$, where $\kappa_\infty = 1/(\rho_0 c_0^2)$ and is the compressibility of water that was 4.336×10^{-10} m·s^2/kg in the simulations. To obtain relaxation parameters for each tissue component, the coefficients were scaled using the ratio of absorption in the tissue component and the reference value and the corresponding ratio of sound speeds.

Computations were performed using a full-wave *k*-space method based on coupled first-order differential equations for linear acoustic propagation.[26] The method accounts for spatially-varying sound speed, density, and relaxation absorption processes, and includes perfectly matched layer (PML) absorbing boundary conditions. This method is temporally exact for homogeneous media and is also accurate for general inhomogeneous media.[26,27] The low numerical dispersion inherent to the *k*-space method allows the effects of frequency-dependent absorption and physical dispersion associated with relaxation-process absorption to be accurately modeled over long paths.

A grid of 768×768 points that spanned 65.03×65.03 mm^2 was used in each computation. To avoid artifactual scattering caused by boundaries between tissue types in the models,[27,28] the tissue maps were lowpass filtered using a Gaussian shaped filter with a 1/*e* response at 67% of the spatial Nyquist frequency. Density maps were shifted one-half sample in the *x* and *y* directions by shifting the phase of their spatial spectra to obtain spatial values for the staggered grid employed in the first-order *k*-space method.[26] The time step was 30 ns in all cases so that the Courant-Friedrichs-Lewy (CFL) number,[29] defined as $c_0 \Delta t / \Delta x$, is 0.53 for water. This choice of CFL number was sufficiently small to maintain high accuracy for the soft-tissue propagation paths considered here.[26,27]

B. Receive focusing

Waveforms received at the simulated array were corrected for geometric delay by using the known positions of the point source and aperture and the assumed sound speed in water. This removed the curvature produced by the propagation geometry and facilitated subsequent analysis of wavefront distortion as well as focusing. After geometric correction, waveforms were temporally windowed before further processing. The window was 2 μs long and had 0.5 μs cosine tapers at each end to be comparable with the window in previous simulations[1,4] and measurements.[7,8]

Wavefront distortion statistics were computed using methods analogous to those described previously.[23] For time-shift estimation, a one-dimensional version of the reference waveform method was employed to calculate the arrival time of the wavefront at each element in the simulated array. The arrival time fluctuations and their correlation length in the receiving aperture were calculated after subtracting a linear fit from the geometrically corrected wavefront, to be comparable with the fit in previous studies[1,4,7,8] that employed specimens or cross sections also with nonparallel surfaces. Wavefronts were then aligned using the computed arrival time fluctuations, and the aligned wavefronts were employed to compute energy level fluctuations, waveform similarity factor, and correlation length of the energy level fluctuations. Energy level fluctuations were defined as the sum of the squared amplitude of the waveform within the processing window, in dB units, also after removal of a linear fit. Correlation lengths were defined by the -6 dB width of the corresponding autocorrelation function. The waveform similarity factor[23] is a kind of generalized cross-correlation bounded by 0 and 1 and equal to unity when all the waveforms are identical.

The received wavefront was synthetically focused for each tissue path to obtain an image of the point source, i.e., the point-spread function for the imaging configuration. The focus was obtained using a Fourier transform implementation of the Rayleigh-Sommerfeld diffraction formula[30] in two dimensions. The implementation for a wave traveling from y_0 to y may be expressed[31]

$$p(x,y;t) = \mathrm{FT}^{-1}\left\{ e^{i(2\pi f|y-y_0|/c_0)} \int_{-\infty}^{\infty} \mathrm{FT}[p(x',y_0;t)] \times \frac{|y-y_0|}{r} \pi \sqrt{f/(c_0 r)}\, e^{-i(2\pi f r/c_0 - \pi/4)}\, dx' \right\}, \tag{3}$$

where

$$r = \sqrt{(x-x')^2 + (y-y_0)^2},$$

FT $[\cdot]$ is the temporal Fourier transform, and $|y\text{-}y_0|$ is the distance of propagation. The first exponential term corresponds to a time delay that centers the focused wavefront in the same 2 μs time window as the received wavefront. To ensure an acyclic temporal convolution, time sequences were zero-padded to double their size before the FFT. Use of the real-space Green's function rather than its spatial Fourier transform ensured that spatial wraparound artifacts were not a problem.[32,33] Prior to focusing, wavefronts were spatially interpolated from the step size of the element pitch (0.6 mm) to the spatial step in the simulations (0.084 67 mm). The same interpolation procedure was used for backpropagation in receive and transmit focusing along with appropriate integration to obtain waveforms at the element positions after backpropagation.

The receive focus was computed for waveforms that were uncompensated, time-shift compensated in the receive aperture, and time-shift compensated after backpropagation a distance of 40 mm. The same distance was used for all the backpropagations because trials showed the waveform similarity maximum was broad, the maximum typically occurred at a distance in the neighborhood of 40 mm, and the performance of the compensation method was not strongly dependent on the precise value of the distance.[23] In each case, a

Hamming window[34] was used to apodize the aperture before focusing. Time-shift compensation employed a sinc function multiplied by a 10-point Kaiser window[35] to interpolate amplitudes at shifts not equal to the original sampling interval. The backpropagation was performed after geometric correction, i.e., using a planar geometry.[23] Use of essentially planar wavefronts simplified the computation by eliminating the need to resample a converging wavefront on a finer spatial grid during backpropagation. Although this may seem different from physical backpropagation, a geometric acoustics argument shows that the effect of geometric correction can be represented by a scaling.[22] Propagation of a converging wave in a cylindrical geometry a distance $r_1 - r_0$ from a cylinder of radius r_1 to a cylinder of radius r_0 ($r_1 > r_0$) is equivalent to propagation in a planar geometry a distance $(r_1 - r_0) r_1 / r_0$ followed by a reduction of size of the result by a factor of r_0 / r_1.

C. Transmit focusing

To examine the effect of propagation through a tissue path on the transmit focus, transmit focusing was also simulated both with and without aberration correction. These simulations started with the specification of a waveform on the elements of the emulated array. The waveforms were spatially apodized with a Hamming window and geometric delays were included in the wavefront to produce a focus at a distance of 55 mm from the array. The *k*-space method was used to propagate the wavefront.

For transmit focusing without compensation for distortion, the unapodized waveform at the elements of the array was

$$a(t) = e^{-(\pi b t/2)^2/\ln(2)} \sin(2\pi f_0 t). \tag{4}$$

The temporal spectrum of this waveform is the same as that of the source given by Eq. (1) in the homogeneous water path region around the source. This provides a basis for comparison of transmit focuses computed using the wavefront defined by Eq. (4) with those computed using time reversal of received wavefronts.

To represent a line source extending in the x direction at the y coordinate of the array, the source term

$$s(x, t + \Delta t/2) = 2\left(\frac{c_0 \Delta t}{\Delta y}\right)\left(\frac{p(x,t) + p(x, t+\Delta t)}{2}\right) \tag{5}$$

was defined. In this expression, $p(x,t)$ is the temporal waveform $a(t)$ after apodization and inclusion of focusing delays. Equation (5) prescribes a pressure wavefront between temporal time steps as required by the *k*-space method using coupled first-order equations.[26] Part of this wave propagates upward (in the $-y$ direction) and part propagates downward (in the $+y$ direction) from the line source. The up-going wavefront is absorbed by the PML boundary. This specification of the source is convenient for the processing described below because the wavefield is observed as a pressure.

To compensate for changes in the apparent source amplitude for the off-axis portion of the wavefront, $p(x,t)$ was multiplied by the following obliquity factor $I(f, k_x, c_0)$ in the spatial-temporal frequency domain:

$$I(f, k_x, c_0) = \begin{cases} [1 - (k_x c_0/(2\pi f))^2]^{1/2}, & \text{if } |k_x c_0/(2\pi f)| < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

where f is temporal frequency and k_x is the spatial frequency in the x direction. The obliquity factor allows the pressure to be used instead of the particle velocity in the Rayleigh integral.[36]

For transmit focusing that employed time-shift compensation, time shifts the same as those used for receive focus correction were applied to the waveform $a(t)$. For transmit focusing that used backpropagation, the transmit wavefront to be apodized and have focusing delays included was obtained by backpropagating a distance of 40 mm the geometrically corrected received wavefront in a 2 μs temporal window, estimating time shifts, applying these shifts to the waveform $a(t)$, and backpropagating the wavefront from the plane of the time-shift estimation to the aperture. For transmit focusing that employed time reversal, waveforms were obtained by reversing in time the waveform in a temporal 2 μs window at each element.

D. Focus evaluation

The focus was described as in Ref. 19 by an effective width in the array (x) and time (y) directions as a function of level below the peak amplitude, by a peripheral energy ratio, and by an effective radius. Effective width, defined in a given direction as the width of the maximum amplitude projection in that direction, was determined using the envelope of the analytic signal as the amplitude in the projection. The temporal effective width was converted to a spatial width using the assumed sound speed for water. Peripheral energy ratio, defined as the ratio of the pulse energy outside a reference ellipse to the pulse energy inside the ellipse, was computed at an amplitude level 10 dB below the peak. Like the reference ellipsoid used in the three-dimensional focus evaluation described in Ref. 19, the reference ellipse was centered at the position of peak amplitude. The width of the ellipse along each axis was the -10 dB effective width in the corresponding direction. The effective radius was defined as one-half the geometric mean of the effective width in the x and in the y directions.

III. RESULTS

A wavefront at an instant of time during propagation through a representative breast tissue map (B03b) from the point source is shown in Fig. 2. Visible in the wavefield are the primary cylindrical wavefront, secondary wavefronts reflected from the water-tissue boundary, and complicated scattering caused by the network of septa around lobules of subcutaneous fat in the breast. Also apparent are local time shifts in the main wavefront caused by propagation along septa aligned with the direction of propagation. These time shifts lead to interference that causes amplitude fluctuations in the received wavefront.[1]

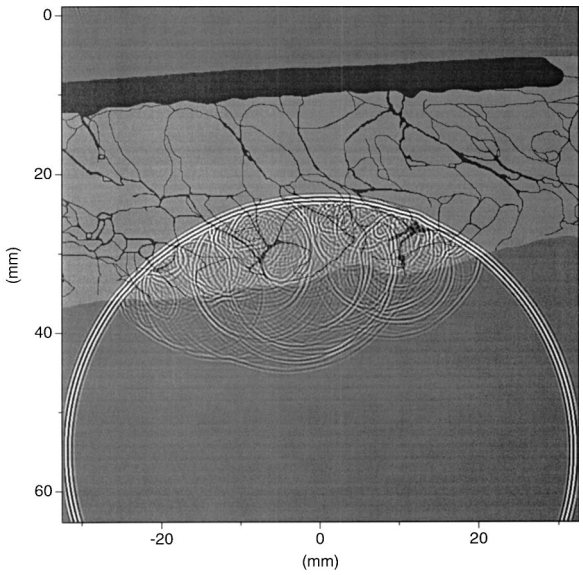


FIG. 2. A pulse wavefront at an instant of time during propagation from a point source through a representative breast tissue map (B03b). The wavefront is superimposed on the map and displayed on a 60 dB bipolar logarithmic gray scale. In the map, dark gray denotes connective tissue or skin and light gray denotes fat.

Received wavefronts are shown in Fig. 3 for a representative selection of abdominal wall and breast tissue maps. The wavefronts in the left column exhibit low arrival time and energy level fluctuations and low waveform distortion. The wavefronts in the center column have moderate arrival time and energy level fluctuations as well as moderate waveform shape distortion. The wavefronts in the right column show high arrival time and energy level fluctuations and also high waveform distortion. The wavefronts suggest that abdominal wall produces wavefront distortion with more small-scale spatial variation than breast, while breast produces distortion with more large-scale spatial variations than abdominal wall. The amplitude fluctuations apparent in the figure are the result of interference between distorted wavefronts as they propagate.

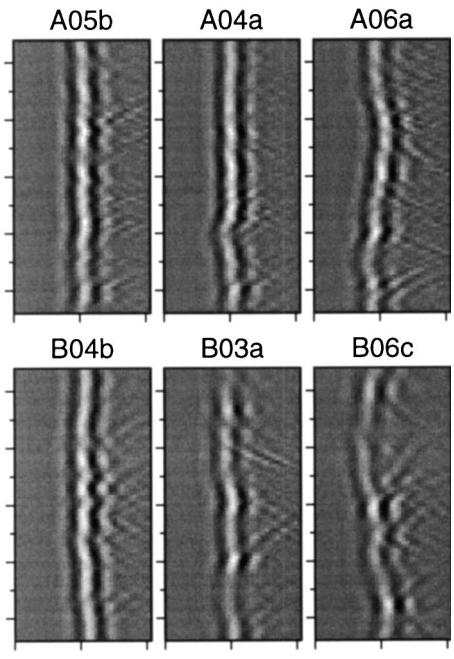


FIG. 3. Wavefronts after propagation through representative tissue maps and geometric correction. The upper row is for abdominal wall paths (from left to right: low, moderate, and high aberration), while the lower row is for breast tissue paths (from left to right: low, moderate, and high aberration). In each panel, the vertical axis spans 48 mm and the horizontal axis spans 2 μs. The amplitude gray scale is linear.

Statistics that describe the received wavefronts are presented in Fig. 4. The average root-mean-square (rms) arrival time fluctuation (ATF), energy level fluctuation (ELF), and ATF correlation length are about the same for both tissue paths (avg$\pm$std dev: 42.5$\pm$15.4 ns, 3.1$\pm$0.5 dB, and 4.4 $\pm$2.7 mm, respectively, for abdominal wall and 43.9$\pm$15.2 ns, 3.3$\pm$0.8 dB, and 3.8$\pm$1.2 mm, respectively, for breast). Also for both abdominal wall and breast tissue paths, backpropagation resulted in little change of the arrival time fluctuation, energy level fluctuation, and ATF correlation length. However, the geometric scale factor discussed above is 0.58 and indicates that the true correlation lengths are much smaller (2.5 mm and 2.2 mm for abdominal wall and breast, respectively, before backpropagation). The average apparent ELF correlation length was decreased by backpropagation (from 2.3$\pm$0.4 mm to 2.0$\pm$0.3 mm for abdominal wall and from 3.3$\pm$1.5 mm to 2.7$\pm$1.0 mm for breast) but the geometric scale factor as in the case of ATF correlation lengths reduces the true length (to 1.2 mm and 1.6 mm for abdominal wall and breast, respectively). The values of waveform similarity factor (WSF) were appreciably increased and their standard deviations were decreased by backpropagation (from 0.952$\pm$0.019 to 0.978$\pm$0.008 for abdominal wall and from 0.940$\pm$0.028 to 0.975$\pm$0.014 for breast).

Wavefronts in the aperture with and without compensation and the corresponding receive focuses are illustrated in Fig. 5 for a representative highly aberrating abdominal wall path and for a representative highly aberrating breast tissue path. The wavefronts are more alike after backpropagation, as quantified by the WSF statistics plotted in the previous figure, so that backpropagation processing improves receive focusing. The focus improvement that results from backpropagation followed by time-shift compensation is greater than the improvement from time-shift compensation alone but the focus still is not ideal.

The focus improvement visible in Fig. 5 is quantified by the effective radii and peripheral energy ratios in Fig. 6 and by other receive focus statistics in Table II. The statistics for the uncompensated focus are substantially improved by time-shift compensation (TSC) and improved still further by the use of backpropagation followed by time-shift compensation (BP+TSC). For example, the mean -20 dB effective radius for breast tissue improves from 2.5$\pm$0.8 mm before compensation to 1.4$\pm$0.4 mm after time-shift compensation and to 1.2$\pm$0.4 mm after backpropagation followed by time-shift compensation. The corresponding -20 dB radius for the water path (ideal) case is 0.9 mm. In general, the statistics indicate breast causes somewhat greater receive focus degradation than abdominal wall. The breast-tissue wavefronts also focus more poorly after compensation; nevertheless, use of backpropagation processing appreciably improves the focus for the breast paths.

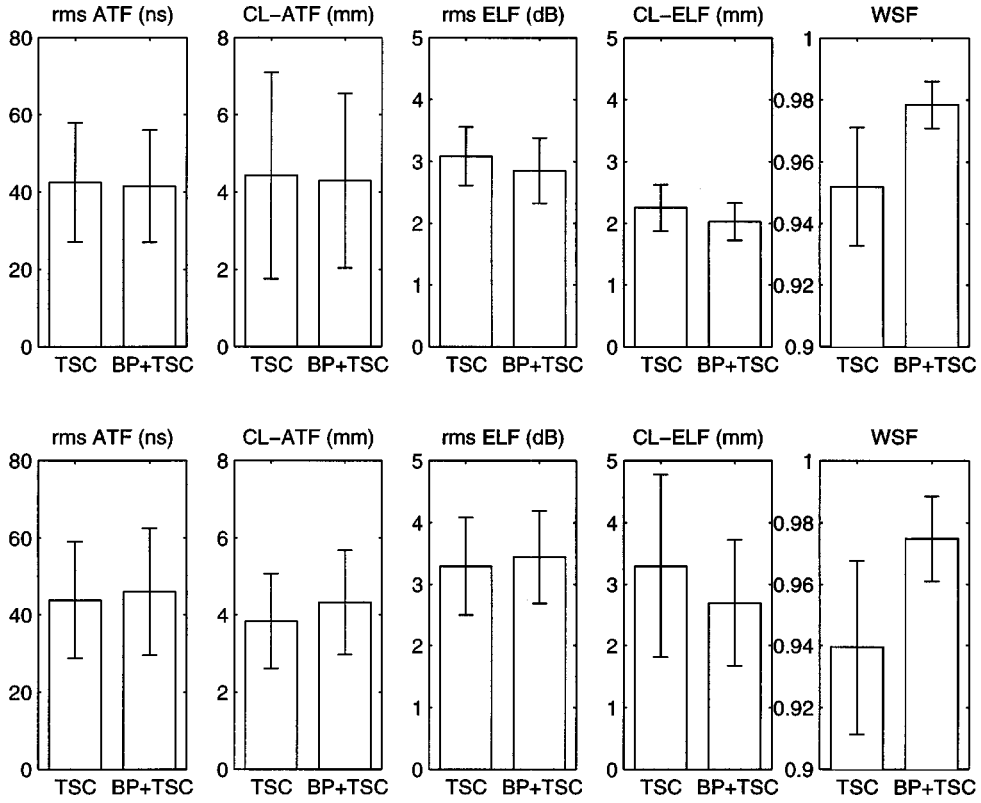


FIG. 4. Statistics of wavefronts received by a virtual array after propagation from a point source through a tissue path. The upper row shows the average and standard deviation for 12 abdominal wall paths and lower row shows the average and standard deviation for 14 breast paths. From left to right: root-mean-square (rms) arrival time fluctuation (ATF), correlation length (CL) of ATF, rms energy level fluctuation (ELF), correlation length (CL) of ELF, and waveform similarity factor (WSF). Each panel shows the results after time-shift compensation (TSC) and after backpropagation followed by time-shift compensation (BP+TSC).

The mainlobe widths of the receive focus using TSC and using BP+TSC are broader than those for the water path case. This difference arises from frequency-dependent attenuation through the tissue path. The attenuation decreases the center frequency of the received waveforms, particularly at the ends of the array and results in a smaller effective aperture. (When propagation was simulated without frequency-dependent absorption, main-lobe widths for compensated receive focusing were close to those for the ideal case.)

A converging uncompensated wavefront and secondary scattered wavefronts at an instant of time during propagation

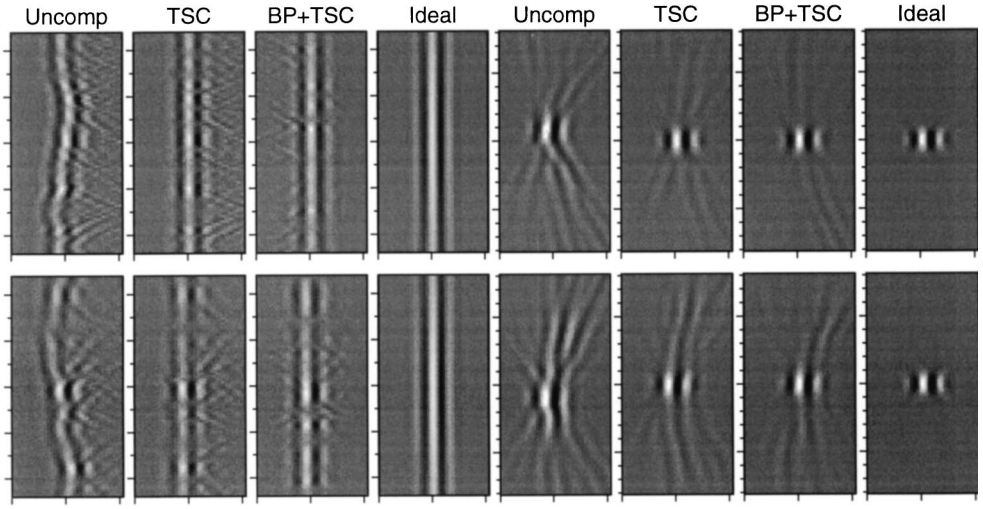


FIG. 5. Uncompensated and compensated receive wavefronts and corresponding focuses. The top row is for a representative highly aberrating abdominal wall path (A06a), while the bottom row is for a representative highly aberrating breast tissue path (B06c). The wavefronts (from left to right) are: uncompensated (Uncomp), time-shift compensated (TSC), backpropagated and time-shift compensated (BP+TSC), and water path (Ideal), respectively, and the corresponding focused wavefronts. The vertical axis spans 48 mm for wavefronts in the aperture and 16 mm for the focused wavefronts, while the horizontal axis spans 2 μs in each case. The amplitude gray scale is linear.

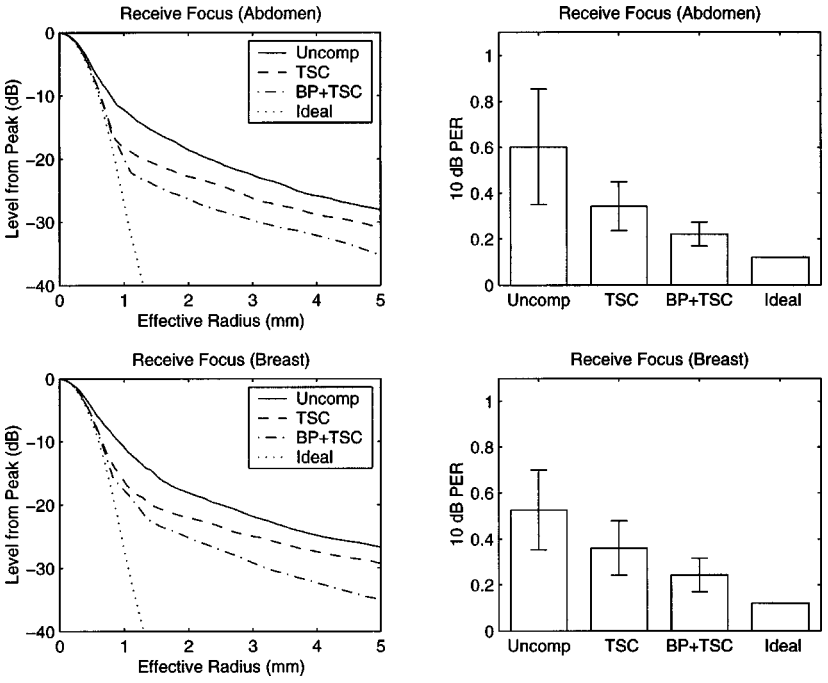


FIG. 6. Effective width and peripheral energy ratio (PER) for receive focusing. The upper panels show the average for 12 abdominal wall paths and the lower panels show the average for 14 breast tissue paths. Uncomp =uncompensated, TSC=time-shift compensation, BP+TSC=backpropagation followed by time-shift compensation, Ideal=water path.

through a representative breast tissue map from the virtual array is shown in Fig. 7. The relatively large element size (0.6 mm, 1.2 times the wavelength at the 3 MHz center frequency) results in grating-lobe wavefronts that appear as curved lines at the sides of the converging wavefront. However, these wavefronts dissipate as the main wavefront approaches the focus. As in the case of propagation from a point source, scattering occurs as the wavefront strikes boundaries between different tissue components.

Transmit wavefronts without and with compensation (before the addition of geometric delay in each case) in the emulated aperture and the corresponding transmit focuses after propagation through a representative highly aberrating abdominal wall and a representative highly aberrating breast tissue path are shown in Fig. 8. The uncompensated transmit focuses show aberration similar to the uncompensated receive focuses even though the focusing processes are different. Compared to time-shift compensation with and without backpropagation, the focus resulting from time-reversal has a broader mainlobe as a consequence of the apodization and has lower sidelobes as a consequence of the relatively weak inhomogeneity of the attenuation and the invariance of lossless propagation to the direction of time.

In contrast to receive focus correction that removes distortion from the wavefront, transmit focus correction predistorts or modifies the wavefront to include a kind of distortion in the transmitted wavefront. Although the focus resulting from time-shift compensation of the backpropagation is improved over the focus resulting from time-shift compensation alone, the improvement is smaller than in receive focusing. As in receive focusing, breast caused greater degradation than abdominal wall. This is expected from the greater amplitude fluctuations in the BP+TSC and time reversed wavefronts.

TABLE II. −10 dB and −20 dB effective radii (ER) in mm and −10 dB peripheral energy ratios (PER) for simulated focusing. Effective widths and peripheral energy ratios are shown for breast and abdominal wall tissue simulations of transmit (TX) and receive (RX) focusing with wavefronts that were uncompensated (Uncomp), time-shift compensated (TSC), time-shift compensated after backpropagation (BP+TSC), or time reversed (TR). Each statistic is shown using the format mean±standard deviation.

Statistic	Focus	Tissue	Uncomp	TSC	BP+TSC	TR
−10 dB	RX	Abdomen	0.80±0.18	0.64±0.02	0.64±0.01	···
ER		Breast	0.92±0.30	0.66±0.03	0.64±0.02	···
	TX	Abdomen	0.78±0.15	0.62±0.01	0.63±0.01	0.65±0.02
		Breast	0.93±0.35	0.65±0.07	0.64±0.04	0.65±0.03
−20 dB	RX	Abdomen	2.33±0.80	1.29±0.33	1.01±0.22	···
ER		Breast	2.51±0.80	1.43±0.43	1.20±0.36	···
	TX	Abdomen	2.34±0.79	1.32±0.44	1.24±0.40	0.98±0.08
		Breast	4.32±2.75	2.40±1.51	2.09±1.33	1.57±0.93
−10 dB	RX	Abdomen	0.60±0.25	0.34±0.11	0.22±0.05	···
PER		Breast	0.53±0.17	0.36±0.12	0.24±0.07	···
	TX	Abdomen	0.70±0.22	0.48±0.16	0.41±0.11	0.27±0.06
		Breast	0.78±0.29	0.63±0.27	0.53±0.22	0.36±0.14

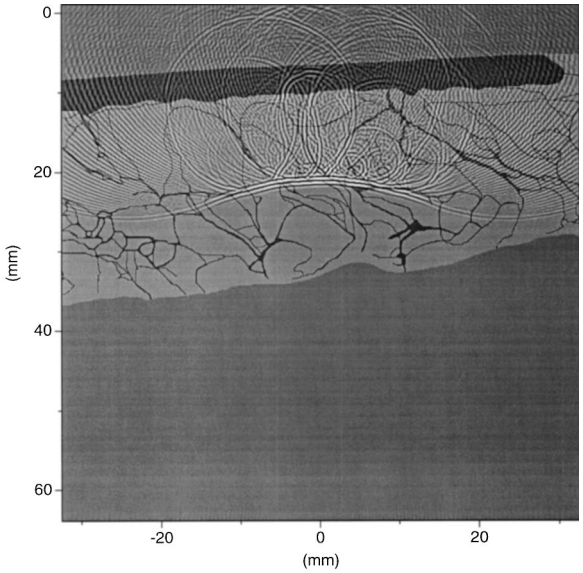


FIG. 7. A converging uncompensated pulse wavefront and secondary scattered wavefronts at an instant of time during propagation from a virtual array through a representative breast tissue map (B03b). The wavefront is superimposed on the map and displayed on a 60 dB bipolar logarithmic gray scale. In the map, dark gray denotes connective tissue or skin and light gray denotes fat.

Statistics of the transmit focuses are shown in Fig. 9 as well as in Table II. The transmit focus quality is generally similar to that for receive focusing in the case of abdominal wall. However, the size of the focus as quantified by the −20 dB effective radii shown in Table II is much larger for transmit focusing through breast tissue than for receive focusing (e.g., 4.3±2.8 mm vs 2.5±0.8 mm uncompensated and 2.4 ±1.5 mm vs 1.4±0.4 mm using time-shift compensation). As expected from the representative focuses shown in Fig. 8, BP+TSC produces a smaller improvement over TSC, in contrast to receive focusing where significant improvement is evident. Also, as seen from comparison of the −10 dB effective radii in Table II, the mainlobe for time-reversal focusing is wider than for TSC and BP+TSC correction. This occurs for two main reasons. First, the received wavefront has undergone frequency-dependent absorption during propagation through the tissue and, consequently, the waveform is lengthened and the wavefront amplitude is further reduced by a longer attenuating path to the edges of the aperture. Second, the spatial weighting of the time-reversed wavefront by the product of the cylindrical spreading factor $1/\sqrt{r}$ and the Hamming window reduces the effective size of the aperture. Nevertheless, the time reversal procedure yields a good focus, at least for the case in which a wavefront from a point source is available. Once again, the degradation caused by the breast tissue is larger than that from abdominal wall tissue, both before and after compensation. These differences are larger for transmit focusing than for receive focusing.

IV. DISCUSSION

The ATF values in the current simulations are about 20% smaller than those in previous simulations[1,4] that employed plane wave propagation. This decrease is attributed mainly to the difference in geometry between point-source and plane-wave propagation. During propagation from the point source, the wavefronts pass through a region that is narrow near the source and widens toward the array so the wavefront encounters fewer different inhomogeneities while passing through the tissue. The values of ATF in simulations of propagation through tissue are much smaller than the corresponding values in measurements because, as previously discussed,[1,4] the tissue maps are less complicated than tissue and propagation is physically different in two and three dimensions.

The ELF values are about the same as those in previous simulations[1,4] and measurements[7] using abdominal wall and about 1 dB less than measurements[8] using breast. The similarity of values is attributed mostly to the local nature of the energy level fluctuations. The short-range correlation is not greatly affected by the fitting process that employed a linear

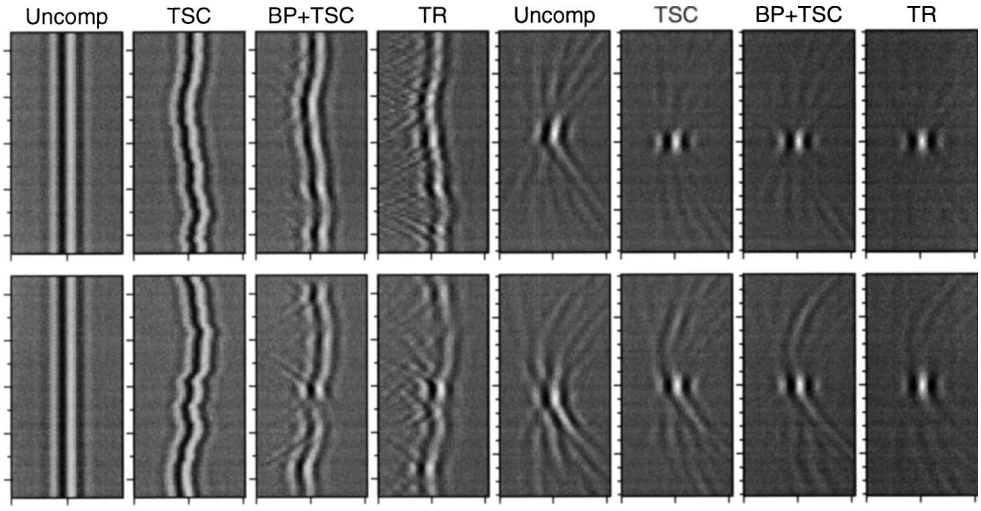


FIG. 8. Uncompensated and compensated transmit wavefronts and corresponding focuses. The top row is for a representative highly aberrating abdominal wall path (A06a), while the bottom row is for a representative highly aberrating breast tissue path (B06c). The wavefronts (from left to right) are: uncompensated (Uncomp), time-shift compensated (TSC), backpropagated and time-shift compensated (BP+TSC), and time reversed (TR), and the corresponding focused wavefronts. The vertical axis spans 48 mm for wavefronts in the aperture and 16 mm for the focused wavefronts, while the horizontal axis spans 2 μs in each case. The amplitude gray scale is linear.

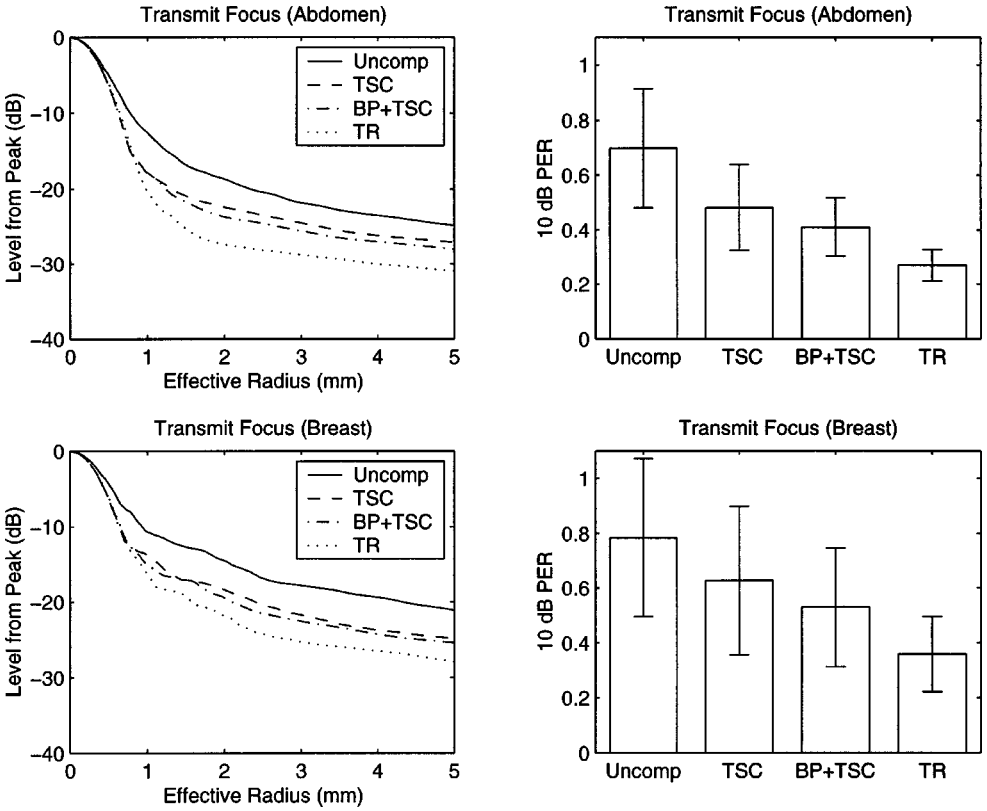


FIG. 9. Effective beam width and peripheral energy ratio (PER) for transmit focusing. The upper panels show the average for 12 model abdominal wall paths and lower panels show the average for 14 model breast tissue paths. The line convention is the same as in Fig. 6 except that TR=time reversal.

fit in the referenced (two-dimensional) simulations of plane-wave propagation through tissue cross sections and a fourth-order, two-dimensional polynomial fit in the referenced measurements of point-source (three-dimensional) propagation through tissue specimens.

The statistics of receive focus metrics agree, however, qualitatively with those obtained from measurements of breast[8] and abdominal wall tissue,[23] although the skirts of the receive focus effective-radius curves are higher in the present study. A major reason for this is that sidelobes decay more slowly in a two-dimensional focus compared to those in a three-dimensional focus. However, waveform similarity factors and energy level fluctuations have similar values to those in Ref. 23 and improvement before and after backpropagation processing is similar to that in Ref. 23 as well. These observations indicate that the distortion simulated here is similar to that measured in human tissue Refs. 8 and 23 even though the arrival-time and energy-level fluctuations differ due to differences in geometry.

The investigated compensation methods performed better for receive focusing than for transmit focusing. Although effective radii for uncompensated focusing are almost identical down to -17 dB for abdominal wall paths and -12 dB for breast tissue paths, effective radii at lower levels are wider for the transmit case. Peripheral energy ratios are also substantially higher for uncompensated transmit focusing. The values of these two metrics indicate higher sidelobe levels compared to receive focusing. Furthermore, backpropagation followed by time-shift compensation in transmit focus correction showed less improvement over time-shift compensation alone than in receive focus correction. This observation agrees with a corresponding observation in an experimental study[22] that employed a two-dimensional array.

The difference between receive and transmit focus improvement is attributed primarily to the difference of the estimation process using homogeneous medium and physical propagation through distributed inhomogeneities. In compensation during reception, the homogeneous path assumed in estimation is also used in focusing. In compensation during transmission, however, the path used for estimation is different from the physical path used for transmission. Although this process works well for a phase screen or a very thin layer of aberration, the performance is degraded as the thickness of aberrating medium increases.

Focus compensation employing backpropagation produces edge waves that merit comment. These edge waves affect the estimation of time-shifts at the boundary of the aperture. In receive focusing, apodization reduces the influence of any time-shift errors and the receive focus improvement that results from inclusion of the backpropagation step before time-shift estimation is appreciable. In transmit focusing, the edge waves depend on the way the compensation is implemented. For example, instead of time-shifting a specific pulse waveform to obtain the wavefront that is backpropagated from the position of the time-shift screen to the position of the aperture and then apodizing the resulting wavefront in the aperture as in the studies described here, the wavefront that is time-shifted can be obtained by backpropagation of an already apodized wavefront from the transmit aperture as in Ref. 22. This latter procedure reduces the edge

waves in the backpropagation from the position of the phase screen but residual edge effects remain along the inner border of the transmit aperture. Since trials showed the transmit focus is about the same using each implementation and since shifting a given pulse at the position of the phase screen is straightforward and requires less computation, that method was used here.

A comparison of time reversal and backpropagation followed by time-shift compensation provides insight about the limited improvement in transmit focusing provided by backpropagation followed by time-shift compensation. While time reversal implicitly incorporates distortion effects along the entire propagation path, the backpropagation processing employed here effectively concentrates distributed aberration in a single phase screen at the backpropagation distance. Although wavefronts compensated using backpropagation processing appear similar to time-reversed wavefronts as seen in Fig. 8, the time-reversed wavefronts contain larger amplitude fluctuations. The smaller amplitude fluctuations produced by backpropagation are insufficient to compensate fully for amplitude variations caused by the distributed inhomogeneities.

The time-reversal method for transmit focusing through the realistic tissue paths used in this study did not approach an ideal diffraction-limited focus. The performance of time-reversal compensation is affected by irreversible processes that exist in practice. Important among these are absorption, element directivity, finite time window, and scattering loss. Also, propagation of a prefocused wavefront serves better than the point source because less energy is lost outside the receive aperture. Although all effects other than scattering loss may be removed in simulations to improve the performance of time-reverse compensation significantly, this has not been done because changes were not used with other methods and a fair comparison of methods under similar practical circumstances is the goal here.

One may ask how well time-reversal processing can perform in practice with finite apertures and temporal windows. The answer is apparently that as long as a point source is available at the focus, time-reversal compensation is superior to the investigated compensation methods in sidelobe suppression though the difference may be reduced depending on physical constraints. This good performance is likely because the time-reversed wavefront still contains some information associated with secondarily scattered and refracted wavefronts. However, in most practical cases, point sources are not available.

The backpropagation method has the advantage of not requiring a point source. The time-reversal method, however, requires at least a pointlike source. Although the difficulty of time-shift estimation from diffuse scattering may slightly increase due to reduced coherence, time-shift estimation in this setting is not a significant problem.[37]

In the realistic situations, further improvement of transmit focus quality is also possible using a spatio-temporal inverse filter.[38] This has been shown to provide a better focus than time reversal in an inhomogeneous medium with attenuation.[39] However, the inverse filtering in this method requires received wavefronts for multiple, well-defined source locations, a configuration not possible in most practical pulse-echo medical imaging applications. Nevertheless, if accurate tissue models are available, simulations like those presented here may be useful in the development of new methods for transmit focus correction.

Notable is that the present simulations, which employed two-dimensional cross-sectional models of breast and abdominal wall tissue, may somewhat underestimate the focus degradation caused by human tissue. As generally understood[40] and recently shown for human breast-tissue simulations,[5] propagation through three-dimensional inhomogeneous media causes greater distortion than propagation through two-dimensional models of the same media. Thus, two-dimensional simulations are expected to result in less focus degradation. However, as discussed above, a two-dimensional focus has higher sidelobe levels than a three-dimensional focus so that this difference may be diminished.

In general, breast tissue causes larger aberration than abdominal tissue. However, the aberration appears at relatively lower spatial frequencies. This may explain why commercial scanners operating at a higher frequency and using a higher *f*-number than employed in this study are more effective in breast imaging than in abdominal imaging. In addition, the adverse effects of aberration on high-frequency breast imaging may be partially compensated by use of tissue compression and selection of transducer position for optimal image quality.

Overall, despite simplifications associated with the tissue model and with propagation in two dimensions, the qualitative agreement of results in this study with results from measurements and the realistic appearance of the wavefront distortion and focus aberration in this study show the employed combination of the tissue maps and *k*-space method of calculation is useful for investigation of focusing and aberration correction in ultrasonic applications through a tissue path.

V. CONCLUSIONS

Simulations employing cross-sectional models of human abdominal wall and breast tissue have provided new information about the effects of tissue structure and compensation on receive and transmit focusing in medical ultrasound. The receive and transmit focus degradation caused by breast tissue were greater than those caused by abdominal wall before and after compensation. The quality of the transmit focus obtained through breast tissue was particularly low. This is attributed to the large amplitude fluctuations caused by large-scale connective tissue structures in breast tissue.

Aberration correction for receive focusing was effective both for abdominal wall and breast tissue paths, although focus quality remained lower for breast after compensation. Time-shift compensation significantly improved focusing but time-shift compensation after backpropagation was substantially better. Since the receive focus can be considered a point-spread function of an analogous imaging system, these results indicate that aberration correction can, in principle, greatly improve the quality of images obtained through tissue paths with large-scale distributed inhomogeneities, provided a satisfactory estimate of the aberration can be obtained.

The improvement of transmit focusing by the methods studied was not as great as obtained in receive focusing. In particular, time-shift compensation after backpropagation provided only small improvement over time-shift compensation alone. Time-reversal produced a transmit focus with low sidelobe levels but a wider mainlobe relative to the other compensation methods investigated and was apparently limited by a smaller effective aperture and spatially varying, frequency-dependent attenuation.

ACKNOWLEDGMENTS

Jeffrey P. Astheimer, James C. Lacefield, and Adrian I. Nachman are thanked for helpful discussions, suggestions, and comments about material in this paper. This research was funded by NIH Grants Nos. HI 50855, CA 74050, and CA 81688, US Army Grant No. DAMD-17-98-1-8141, DARPA Grant No. N00014-96-0749, and the University of Rochester Diagnostic Ultrasound Research Laboratory Industrial Associates.

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

[2] G. Wojcik, B. Fornberg, R. Waag, L. Carcione, J. Mould, L. Nikodym, and T. Driscoll, ''Pseudospectral methods for large-scale bioacoustic models,'' 1997 IEEE Ultrason. Symp. Proc. **2**, 1501–1506 (1997).

[3] J. L. Aroyan, ''Three-dimensional modeling of hearing in *Delphinus delphis*,'' J. Acoust. Soc. Am. **110**, 3305–3318 (2001).

[4] T. D. Mast, L. M. Hinkelman, M. J. Orr, and R. C. Waag, ''The effect of abdominal wall morphology on ultrasonic pulse distortion. Part II. Simulations,'' J. Acoust. Soc. Am. **104**, 3651–3664 (1998).

[5] T. D. Mast, ''Two- and three-dimensional simulations of ultrasonic propagation through human breast tissue,'' Acoust. Res. Lett. Online **3**, 53–58 (2002).

[6] Q. Zhu and B. D. Steinberg, ''Large-transducer measurements of wavefront distortion in the female breast,'' Ultrason. Imaging **14**, 276–299 (1992).

[7] L. M. Hinkelman, D-L. Liu, L. A. Metlay, and R. C. Waag, ''Measurements of ultrasonic pulse arrival time and energy level variations produced by propagation through abdominal wall,'' J. Acoust. Soc. Am. **95**, 530–541 (1994).

[8] L. M. Hinkelman, D.-L. Liu, R. C. Waag, Q. Zhu, and B. D. Steinberg, ''Measurement and correction of ultrasonic pulse distortion produced by the human breast,'' J. Acoust. Soc. Am. **97**, 1958–1969 (1995).

[9] T. D. Mast, L. M. Hinkelman, L. A. Metlay, and R. C. Waag, ''The effect of abdominal wall morphology on ultrasonic pulse distortion. Part I. Measurements'' J. Acoust. Soc. Am. **104**, 3635–3649 (1998).

[10] G. E. Trahey, P. D. Freiburger, and D. C. Sullivan, ''The impact of acoustic velocity variations on target detectability in ultrasonic images of the breast,'' Invest. Radiol. **26**, 782–791 (1991).

[11] A. P. Berkhoff and J. M. Thijssen, ''Correction of concentrated and distributed aberrations in medical ultrasound imaging,'' 1996 IEEE Ultrason, Symp. Proc. **2**, 1405–1410 (1996).

[12] T. Christopher, ''Finite amplitude distortion-based inhomogeneous pulse echo ultrasonic imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 125–139 (1997).

[13] L. Ødegaard, E. Halvorsen, B. Ystad, H. G. Torp, and B. A. Angelsen, ''Delay and amplitude focusing through the body wall; A simulation study,'' 1996 IEEE Ultrason. Symp. Proc. **2**, 1411–1414 (1996).

[14] I. M. Hallaj and R. O. Cleveland, ''FDTD simulation of finite-amplitude pressure and temperature fields for biomedical ultrasound,'' J. Acoust. Soc. Am. **105**, L7–L12 (1999).

[15] P. Roux, H. C. Song, M. B. Porter, and W. A. Kuperman, ''Application of the parabolic equation method to medical ultrasonics,'' Wave Motion **31**, 181–196 (2000).

[16] Q. Zhu, B. D. Steinberg, and R. L. Arenson, ''Wavefront amplitude distortion and image sidelobe levels: Part II—*in vivo* experiments,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 754–762 (1993).

[17] M. O'Donnell and S. W. Flax, ''Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 768–774 (1988).

[18] L. Nock, G. E. Trahey, and S. W. Smith, ''Phase aberration correction in medical ultrasound using speckle brightness as a quality factor,'' J. Acoust. Soc. Am. **85**, 1819–1833 (1989).

[19] D.-L. Liu and R. C. Waag, ''Time-shift compensation of ultrasonic pulse focus degradation using least-mean-square error estimates of arrival time,'' J. Acoust. Soc. Am. **95**, 542–555 (1994).

[20] M. Fink, ''Time-reversal of ultrasonic fields—Part I: Basic principles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–567 (1992).

[21] H. Wang, E. S. Ebbini, M. O'Donnell, and C. A. Cain, ''Phase aberration correction and motion correction for ultrasonic hyperthermia phased arrays: Experimental results,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 34–43 (1994).

[22] J. C. Lacefield and R. C. Waag, ''Evaluation of backpropagation methods for transmit focus compensation,'' 2001 IEEE Ultrason. Symp. Proc., paper 2D-4 (2001).

[23] D.-L. Liu and R. C. Waag, ''Correction of ultrasonic wavefront distortion using backpropagation and a reference waveform method for time-shift compensation,'' J. Acoust. Soc. Am. **96**, 649–660 (1994).

[24] C. Dorme and M. Fink, ''Ultrasonic beam steering through inhomogeneous layers with a time reversal mirror,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 167–175 (1996).

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

[26] M. Tabei, T. D. Mast, and R. C. Waag, ''A *k*-space method for coupled first-order acoustic propagation equations,'' J. Acoust. Soc. Am. **111**, 53–63 (2002).

[27] T. D. Mast, L. P. Souriau, D.-L. D. Liu, M. Tabei, A. I. Nachman, and R. C. Waag, ''A *k*-space method for large-scale models of wave propagation in tissue,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 341–354 (2001).

[28] J. C. Mould, G. L. Wojcik, L. M. Carcione, M. Tabei, T. D. Mast, and R. C. Waag, ''Validation of FFT-based algorithms for large-scale modeling of wave propagation in tissue,'' 1999 IEEE Ultrason. Symp. Proc. **2**, 1551–1556 (1999).

[29] E. Turkel, ''On the practical use of high-order methods for hyperbolic systems,'' J. Comput. Phys. **35**, 319–340 (1980).

[30] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968), Sec. 3.4.

[31] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Sec. 7.3.

[32] M. Tabei and M. Ueda, ''On the sampling conditions for reconstruction of an acoustic field from a finite sound source,'' J. Acoust. Soc. Am. **111**, 940–946 (2002).

[33] R. C. Waag, J. A. Campbell, J. Ridder, and P. R. Mesdag, ''Cross-sectional measurements and extrapolations of ultrasonic fields,'' IEEE Trans. Sonics Ultrason. **SU-32**, 26–35 (1985).

[34] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra* (Dover, New York, 1958), Sec. II B 5.

[35] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1989), Sec. 7.4.

[36] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Sec. 5.2.

[37] D.-L. D. Liu and R. C. Waag, ''Estimation and correction of ultrasonic wavefront distortion using pulse-echo data received in a two-dimensional aperture,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 473–490 (1998).

[38] M. Tanter, J.-F. Aubry, J. Gerber, J.-L. Thomas, and M. Fink, ''Optimal focusing by spatio-temporal inverse filter. I. Basic principles,'' J. Acoust. Soc. Am. **110**, 37–47 (2001).

[39] J.-F. Aubry, M. Tanter, J. Gerber, J.-L. Thomas, and M. Fink, ''Optimal focusing by spatio-temporal inverse filter. II. Experiments. Application to focusing through absorbing and reverberating media,'' J. Acoust. Soc. Am. **110**, 48–58 (2001).

[40] J. M. Martin and S. M. Flatté, ''Simulation of point-source scintillation through three-dimensional random media,'' J. Opt. Soc. Am. A **7**, 838–847 (1990).

Erratum: "Geoacoustic inversion for fine-grained sediments" [J. Acoust. Soc. Am. 111, 1560–1564 (2002)]

Charles W. Holland
SACLANT Undersea Research Centre, Viale San Bartolomeo 400, 19138 La Spezia, Italy

(Received 22 November 2002; accepted for publication 22 November 2002)

[DOI: 10.1121/1.1537711]

PACS numbers: 43.20.El, 43.30.Pc, 43.30.Ma, 43.10.Vx

There was a typographical error in the density of the seawater on page 1562. The correct density is 1.029 g/cm^3 (not 1.092 g/cm^3).

 0001-4966/2003/113(2)/1177/1/$19.00