CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 115, No. 4 April 2004

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 115, NO. 4, APRIL 2004

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 115, NO. 4, APRIL 2004

CONTENTS—*Continued from preceding page*

ACOUSTIC SIGNAL PROCESSING [60]

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

SPEECH PERCEPTION [71]

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 115, NO. 4, APRIL 2004

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

ERRATA

CUMULATIVE AUTHOR INDEX

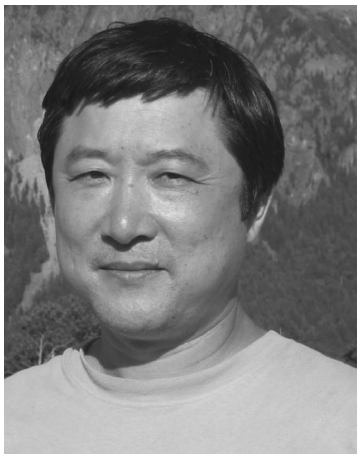
ACOUSTICAL NEWS—USA

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

Abeer Alwan—For contributions to research in speech production and perception and applications to speech technology.

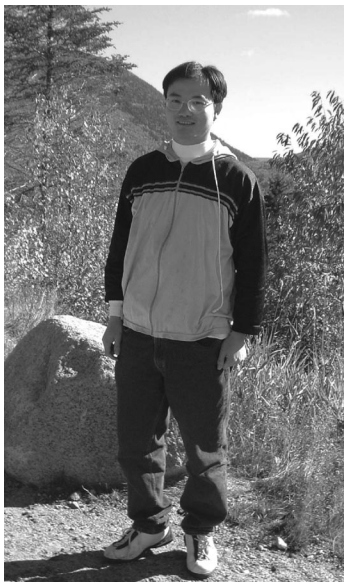


Li Deng—For contributions ot speech processing and recognition, especially robust recognition of noisy and conversational speech.

Andrew J. Oxenham—For contributions to understanding peripheral nonlinearity, temporal processing, and perceptual grouping.

F. V. Hunt Postdoctoral Research Fellowship awarded to Xuedong Zhang

The 2004–06 F. V. Hunt Postdoctoral Research Fellowship in Acoustics was awarded to Xuedong Zhang. Under the fellowship, Mr. Zhang will undertake a research program at The Massachusetts Institute of Technology's Research Laboratory of Electronics. The subject of his research is on the understanding of low-frequency human cochlear mechanics. Mr. Zhang received the B.E. and M.E. degrees from Southeast University in the People's Republic of China. His Ph.D. thesis is titled ''Cross-frequency coincidence detection in processing of complex sounds.''



Xuedong Zhang

The Hunt Fellowship is granted each year to an ASA member who has recently received his or her doctorate or will be receiving the degree in the year in which the fellowship is to be granted. The recipient of the fellowship is that individual who, through personal qualifications and a proposed research topic, is judged to exhibit the highest potential for benefiting any aspect of the science of sound and promoting its usefulness to society. Further information about the fellowship is available from the Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Phone: 516-576-2360; fax: 516-576-2377; E-mail: asa@aip.org; Web: asa.aip.org/fellowships.html

The 146th meeting of the Acoustical Society of America held in Austin, Texas

The 146th meeting of the Acoustical Society of America was held 10–14 November at the Renaissance Austin Hotel in Austin, Texas. This is the sixth time the Society has met in this city, the previous times being 1954, 1964, 1975, 1985, and 1994.

The meeting drew a total of 985 registrants, including 164 nonmembers and 226 students. Attesting to the international ties of our organization, 131 of the registrants, (that is, 13%) were from outside North America. There were 28 registrants from Japan, 11 from Germany, 9 from the United Kingdom, 7 from Korea, 6 each from Italy and Norway, 5 each from Australia and France, 3 each from the Netherlands, Republic of China and Russia, 2 each from Argentina, New Zealand and Switzerland, and 1 each from Belgium, Chile, Czech Republic, Denmark, Ecuador, Finland, India, Ireland, Poland, Singapore, Spain, and Sweden. North American countries, the United States, Canada, and Mexico, accounted for 782, 27, and 4, respectively.

A total of 737 papers, organized into 83 sessions, covered the areas of interest of all 13 Technical Committees. The meeting also included seven meetings dealing with standards. The Monday evening tutorial lecture series was continued by Peggy B. Nelson of the University of Minnesota. Her

 0001-4966/2004/115(4)/1369/5/$20.00

FIG. 1. Attendees at the 25th Anniversary Hunt Fellowship luncheon (l–r): Anthony Atchley, Thomas Hunt, Lily Wang, Gregory Sandell, Tyrone Porter, T. Douglas Mast, Logan Hargrove, Carr Everbach, David Blackstock, Kenneth Cunefare, Steven Garrett, Ian Lindevald, Constantin Coussios, Mark Hasegawa-Johnson, Mark Hamilton, Constantine Trahiotis, Chao-Yang Lee, and Wayne Wright.

tutorial ''Classroom acoustics'' was presented to an audience of about 100.

The Short Course on Time Reversal was held on Sunday and Monday and the course instructors were William A. Kuperman of the Scripps Institution of Oceanography and Mathias Fink of the University of Paris. Nineteen registrants attended the short course.

A seminar on Forensic Acoustics was held on Friday, 14 November. A group of trial attorneys and specialists were present to share their knowledge in support of expert witness activity. The seminar drew 29 participants.

A Distinguished Lecture titled ''Analysis of community response to transportation noise a quarter century after Schultz (1978)'' was presented by Sanford Fidell of BBN Technologies.

A one-day colloquium titled ''Topical meeting on shock waves in medicine,'' organized by the Technical Committee on Biomedical Ultrasound/Bioresponse to Vibration, was held on Friday.

The Society's 13 Technical Committees held open meetings during the Austin meeting, where they made plans for special sessions at upcoming ASA meetings, discussed topics of interest to the attendees, and held informal socials after the end of the official business. These are working, collegial meetings, and all people attending Society meetings are encouraged to attend and to participate in the discussions.

The equipment exhibit drew 14 exhibitors, and an exhibit opening reception was held on Monday evening.

The Society held a Job Fair at the meeting, which was organized and operated by the Career Services Division of the American Institute of Physics. The job fair drew 20 employers posting 33 jobs, 2 of which conducted interviews at the meeting. There were 81 job seekers who participated in the Fair.

The local committee organized two Technical Tours. The tour to National Instruments (NI) was held on Tuesday, 11 November, and featured a visit to the NI's manufacturing facilities and a demonstration of a multichannel-networked audio measurement system. The tour to the Bates Recital Hall and Visser-Rowland Tracker Organ was held on Thursday, 13 November. The tour included a presentation by Dr. Pieter Visser of Visser and Associates, which built the organ and a demonstration of the organ's capabilities. Mr. Charles Boner gave a presentation about the acoustics of the Bates Recital Hall at the University of Texas.

Social events included the two social hours held on Tuesday and Thursday, a reception for students, a Fellows Lounge and a Fellows Luncheon and the morning coffee breaks. A special program for students to meet one-on-one with members of the ASA over lunch, which is held at each meeting, was organized by the Committee on Education in Acoustics.

A luncheon in celebration of the 25th Anniversary of the F. V. Hunt Postdoctoral Research Fellowship in Acoustics was held at the meeting with 13 of the recipients attending (see Fig. 1). The luncheon was also attended by several guests who have been instrumental in this fellowship program. A special guest was Thomas Hunt, son of Frederick Hunt, who attended the luncheon and presented a talk about his father.

The local committee arranged for a special speaker at the Fellows Luncheon, Professor Steven Weinberg of the University of Texas at Austin. Dr. Weinberg is the recipient of the 1979 Nobel Prize in Physics (see Fig. 2).

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

FIG. 2. Professor Steven Weinberg delivers talk at the Fellows Luncheon.

FIG. 3. ASA President Ilene Busch-Vishniac (l) congratulates Emily Thompson (r), recipient of the 2003 Science Writing Award in Acoustics for Journalists.

FIG. 4. ASA President Ilene Busch-Vishniac presents the Silver Medal in Physical Acoustics to Philip L. Marston.

The Women in Acoustics Luncheon was held on Wednesday afternoon and was attended by over 75 people.

Other special events included a book signing and reception to announce the newly published book, *Concert Halls and Opera Houses: Music, Acoustics and Architecture* by Leo L. Beranek. Meeting attendees were invited to attend the reception to celebrate the publication of the book. Dr. Beranek was present to autograph copies of his book.

ASA President Ilene J. Busch-Vishniac presented three Society awards and presented certificates to newly-elected Fellows at the plenary session.

The 2003 Science Writing Award in Acoustics for a Journalist was presented to Emily Thompson for her book *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America*, 1900–1933 (see Fig. 3). The President announced the Science Writing Award for Professionals in Acoustics that was awarded to Joe Wolfe for his website on clarinet acoustics.

The Silver Medal in Physical Acoustics was presented to Philip L. Marston of Washington State University ''for contributions to generalized ray theories for acoustical scattering, and the acoustical manipulation of fluids to study fundamental phenomena in fluid mechanics and optics'' (see Fig. 4).

The Trent-Crede Medal was presented to Sabih I. Hayek of Pennsylvania State University ''for contributions to the understanding of sound interaction with submerged structures'' (see Fig. 5).

President Busch Vishniac introduced Leo Beranek, who had received the

FIG. 5. ASA President Ilene Busch-Vishniac congratulates Sabih Hayek, recipient of the Trent-Crede Medal.

FIG. 6. ASA President Ilene Busch-Vishniac congratulates Leo L. Beranek, who received the 2003 President's National Medal of Science in Washington, DC, on 9 November.

2003 President's National Medal of Science at the White House the previous week (see Fig. 6).

The election of 21 members to the Fellow grade was announced and fellowship certificates were presented. New fellows are George L. Augspurger, D. Murray Campbell, Dezhang Chu, Kenneth A. Cunefare, Grant B. Deane, David R. Dowling, Timothy J. Foulkes, Ronald R. Freiheit, Peter Gerstoft, John J. Guinan, Jr., John M. Harrison, Steven G. Kargl, Joseph F. Lingevitch, Thomas J. Matula, James H. Miller, Jeffrey A. Nystuen, Michael D. Richardson, Brigitte Schulte-Fortkamp, Julius O. Smith, III, Hee-Chun Song, and Evan K. Westwood (see Fig. 7).

The President expressed the Society's thanks to the Local Committee for the excellent execution of the meeting, which clearly evidenced meticulous planning. He introduced the Chair of the Meeting, Clark S. Penrod (see Fig. 8), who acknowledged the contributions of the members of his committee, including Evan K. Westwood, Technical Program Chair; James E. Stockton, Deputy to the General Chair; Lorri Polvado, Hotel/Facilities; B. J. Gatlin, Food Service/Social Events; Leonard Hebert, Audio-Visual; Jan Chambers, Registration; Carlie Tilly, Accompanying Persons Program, Leonard Hebert and Richard Garcia, Signs; Joy Whitney, Treasurer. The meeting Chair also expressed thanks to the members of the Technical Program Organizing Committee: Evan K. Westwood, Chair; James H. Miller and Peter F. Worcester, Acoustical Oceanography; Robert H. Benson, Animal Bioacoustics; Jack E. Randorff, Architectural Acoustics; R. Glynn Holt, Biomedical Ultrasound/Bioresponse to Vibration; Courtney B. Burroughs, Education in Acoustics; Gary W. Elko, Engineering Acoustics; James M. Cottingham and Wilson Nolle, Musical Acoustics; Brandon D. Tinianov, Noise; Mark F. Hamilton and Evgenia A. Zabolotskaya, Physical Acoustics; Emily A. To-

FIG. 7. New Fellows of the Acoustical Society of America.

FIG. 8. Clark Penrod, Chair of the Austin meeting, with several members of the local committee.

bey, Psychological and Physiological Acoustics; Paul J. Gendron, Gary W. Elko, and Preston S. Wilson, Signal Processing in Acoustics; Peter MacNeilage and Emily A. Tobey, Speech Communication; Courtney B. Burroughs, Structural Acoustics and Vibration; David P. Knobles and Christopher D. Tiemann, Underwater Acoustics.

The full technical program and award encomiums can be found in the printed meeting program for readers who wish to obtain further information about the Austin meeting. We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information on future meetings can be found in the *Journal* and on the ASA Home Page at ⟨http://asa.aip.org⟩.

ILENE J. BUSCH-VISHNIAC
President 2003–2004

USA Meetings Calendar

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

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
12–14 July	Noise-Con 2004, Baltimore, MD [Institute of Noise Control Engineering of the USA, Inc., INCE/USA Business Office, 212 Marston Hall, Iowa State Univ., Ames, IA 50011-2153; Tel.: 515-294-6142; Fax: 515-294-3528; E-mail: ibo@inceusa.org; WWW: http://www.inceusa.org/NoiseCon04call.pdf]
3–7 August	8th International Conference of Music Perception and Cognition, Evanston, IL [School of Music, Northwestern Univ., Evanston, IL 60201; E-mail: WWW: www.icmpc.org/conferences.html].
20–24 September	ACTIVE 2004—The 2004 International Symposium on Active Control of Sound and Vibration, Williamsburg, VA [INCE Business Office, Iowa State Univ., 212 Marston Hall, IA 50011-2153; Fax: 515 294 3528; E-mail: ibo@ince.org; WWW: inceusa.org].
15–19 November	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: http://asa.aip.org].

2005

16–19 May	Society of Automotive Engineering Noise & Vibration Conference, Traverse City, MI [Patti Kreh, SAE International, 755 W. Big Beaver Rd., Ste. 1600, Troy, MI 48084; Tel.: 248-273-2474; E-mail: pkreh@sae.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. Price: ASA members $40 (paperbound); nonmembers $90 (clothbound).

Revision List

New Associates

Allen, Donald R., 10001 Anaheim Ave., NE, Albuquerque, NM 87122
Bergman, David R., 46 Green Village Rd., A21, Madison, NJ 07940
Bosch, Lauren A., ChEARS, 6645 Alvarado Rd., 4th Fl., San Diego, CA 92120
Burton, Thomas E., P.O. Box 232, Hakalau, HI 96710-0232
de Heer, Raymond, C., Animal Sciences Group—Ethology, Wageningen UR, P.O. Box 338, Wageningen, Gelderland, 6700 AH, The Netherlands
Early, III, Thomas A., 2449 Brussels Court, Reston, VA 20191
Firszt, Jill B., Otolaryngology & Communication Sciences, Medical College of Wisconsin, 9200 West, Wisconsin Ave., Milwaukee, WI 53226
Friedman, Adam D., 18407 Charity Ln., Accokeek, MD 20607
Gibbons, Joel D., BBN Technologies, 1300 North 17th St., Ste. 400, Arlington, VA 22209

Gilmour, Sean, Luashbane, Pallasgreen, Limerick, Ireland
Higgs, Dennis M., Biology Dept., Univ. of Windsor, P.O. Box 33830, Detroit, MI 48232
Hino, Junichi, Mechanical Engineering, Univ. of Tokushima, 2-1 Minami josan jima, Tokushima, 770-8506, Japan
Horowitz, Seth S., SUNY Stony Brook, Psychiatry, Health Sciences Ctr., T-10, Room 086, Stony Brook, NY 11794
Jarzabkowski, Michael J., 955 Mix Ave., #3A, Hamden, CT 06514
Lalime, Aimee L., Arup Acoustics, 2440 South Sepulveda Boulevard, Ste. 180, Los Angeles, CA 90064
Leiderman, Ricardo, Rua Marques de Sao Vicente 96, Bl. B apto 802 Gavea, Rio de Janeiro, 22451-040, Brazil
Leske, Lawrence A., 2619 Hastings Drive, Belmont, CA 94002-3323
Mellacheruvu, Venkata S., Dept. of Mathematics, DRBCCC Hindu College, Pattabiram, Chennai, Tamil Nadu 600 072, India
Mikhin, Dmitry, Acacia Research Pty Ltd., 5 Butler Drive, Hendon SA 5014, Australia
Mitchell, Kelton D., Miller, Bean & Paganelli, Inc., 6723 Whittier Ave., Ste. 101, McLean, VA 22101
Moore, Robert E., Speech Pathology and Audiology, Univ. of South Alabama, 2000 UCOM, Mobile, AL 36688
Noble, Kevin M., 4020B Clinton Ave., Klamath Falls, OR 97603
Peters, Gustav, Mammal Dept., Forschungsinstitut und Museum Alexander Koenig, Adenauerallee, 160, Bonn D-53113, Germany
Prevedini, Val J., ASG Telecom, 447 BRd. St., Meriden, CT 06450
Ryodo, Kawasaki, Defense Systems Division, Oki Electric Industry Co., Ltd., 4-10-3 Shibaura, Minato-ku, 108-8551 Tokyo, Japan
Sarkar, Kausik, Mechanical Engineering, Univ. of Delaware, 126 Spencer Laboratory, Newark, DE 19716
Schairer, Kim S., Boys Town National Research Hospital, 555 North 30th St., Omaha, NE 68131
Schallock, Ryan D., 143 North 8th St., Philomath, OR 97370
Tang, Hsin Min, Pro-Wave Electronics, Corp., 3rd Fl., #4, Ln. 348, Sec. 2, Chung Shan Rd., Chung Ho City, Taipei Hsien 235, Taiwan
Tufts, Jennifer B., 7057 Carroll Ave., #2, Takoma Park, MD 20912
Watry, Derek L., Wilson, Ihrig & Associates, 5776 Broadway, Oakland, CA 94618
Zia, Jalal H., Praxair Inc., 175 East Park Drive, Tonawanda, NY 14150

New Students

Altinsoy, Mehmet E., Universitaetsstr. 91, Bochum NRW 44789, Germany
Bauer, Dominik, Inst. of Phonetics, Saarland Univ., Im Stadtwald, Saarbruecken, Saar 66041, Germany
Byron, Bethany M., 3571 Folly Quarter Rd., Ellicott City, MD 21042
Chen, Xi, Physics and Astronomy, Brigham Young Univ., N283 ESC, Provo, UT 84602
Coker, Ian C., 226 Cheddington, Katy, TX 77450
Culley, Tonya L., North Carolina State Univ., Architecture, P.O. Box 7710, Raleigh, NC 27695
Davis, Brian A., Dept. of Mechanical Engineering, North Carolina State Univ., 3211 Broughton Hall, Raleigh, NC 27695
Ericsson, Johan A., Nygatan 39, 1tr, Norrkoeping 602-34, Sweden
Gamah, Ieish, 16 Tonbridge Rd., West Molesey, Surrey, KT8 2EL, England
Hegewisch, Katherine C., 1225 NW Hall, #2, Pullman, WA 99163
Jang, Kyunga, Dept. of Information and Telecommunication Engineering, Soongsil Univ., 1-1 Sangdo5dong, dongjakgu, Seoul 156743, Republic of Korea
Knight, Gary D., Physical Electronics, Ottawa-Carelton Inst. of Physics, 1125 Col. By Dr., Ottawa ON K1S 5B6, Canada
Liu, Siyun, 700 East Loop Rd., L2178C Chapini Apartment, Stony Brook, NY 11790
Massot, Olivier, 326 Graham Ave., 4F, Brooklyn, NY 11211
Morgans, Richard C., Mechanical Engineering, Active Noise and Vibration Control Group, Univ. of Adelaide, Adelaide SA 5005, Australia
Murray, Rupert A. G., 1370 Wolcott Ave., Apt. Bsmt., Chicago, IL 60622
Nosal, Eva-Marie, Geology and Geophysics, Univ. of Hawaii at Manoa, 1680 East–West Rd., Post 813, Honolulu, HI 96822
Oyewole, Olanrewaju, Omega, Inc., P.O. Box 913 Ikeja-Lagos, Lagos 23401, Nigeria
Parchment, J., P.O. Box 41471, Tucson, AZ 85717-1471
Rampersad, Joanne V., 300 Farmstead Ln., #3, State College, PA 16803
Roxberg, Jillyn, Communicative Disorders, Univ. of Wisconsin—Madison, 1975 Willow Dr., Madison, WI 53706
Saweikis, Meghan G., Psychological Sciences, Purdue Univ., 703 3rd St., West Lafayette, IN 47907
Smitthakorn, Pattra, 1241 SW 4th Ave., Apt. 34, Gainesville, FL 32601
Tucker, Benjamin V., Dept. of Linguistics, Univ. of Arizona, 1100 East Bldg. Boulevard, Tucson, AZ 85721
Turner, Andrew D., 14858 Gilomore St., Van Nuys, CA 91411
Usher, John, Faculty of Music, McGill Univ., Strathcona Music Bldg., Room E203, 555 Sherbrooke St., West, Montreal PQ H3A 1E3, Canada
Vigeant, Michelle C., Architectural Engineering, Univ. of Nebraska, 1110 South 67th St., Omaha, NE 68182-0176
Wei, Wei, 410 SE Pioneer St., Apt. 1, Pullman, WA 99163
Wilson, Ian L., 5638 Crown St., Vancouver BC V6N 2B5, Canada
Yoo, Junehee, Joogong Apt. 302-104, Kwacheon, Gyeonggi-do 427-040, Republic of Korea
Zhang, Xuedong, Biomedical Engineering, Boston Univ., 44 Cummington St., Boston, MA 02215

New Electronic Associates

Baumgarte, Frank, 395 Ano Nuevo Ave., Apt. 102, Sunnyvale, CA 94085
Biwa, Tetsushi, Crystalline Materials Science, Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8603, Japan
Botbol, Meir, Deepbreeze Ltd., Hilan 2, Or-Akiva 30600, Israel
Buske, Stefan, Inst. of Geological Sciences, Univ. of Berlin, Malteserstr. 74-100, Bldg. D, Berlin 12249, Germany
Davis, Jon B., VACC, Inc., 490 Post St., #1441, San Francisco, CA 94102
Fischer, Brynn A., Fabric Care Technology, Whirlpool Corporation, 303 Upton Dr., 0212, St. Joseph, MI 49085
Gerhard-Multhaupt, Reimund, Dept. of Physics, Univ. of Potsdam, Am Neuen Palais 10, Brandenburg D014469, Germany
Gyurko, Harrison D., 1055 Beechmont St., Dearborn, MI 48124
Piliavin, Michael A., Physical Optics Corp., 20600 Gramercy Place, Torrance, CA 90501
Ribeiro, Maria, FEUP DEC, Rua Dr. Roberto Frias s/n, Porto 4200-465, Portugal
Sa Ribeiro, Rui Miguel, Rua da Mianca 15002, S. Mamede de Infesta, Porto 4465-197, Portugal
Sharma, Devdutt, 192 Pandan Loop, #02–09, 128381, Singapore
Shore, Steven N., Dept. of Physics, Univ. of Pisa, via Buonarroti 2, Pisa I-56127, Italy
Stultz, Carlton D., Commander Patrol Recon Force Atlantic, Tactics and New Technology (N8), 7927 Ingersol St., Ste. 250, Norfolk, VA 23511

Associates Elected Members

J. P. Cottingham, S. E. Demain, J. S. Dembowski, B. J. Forbes, B. Kostek, E. S. Livingston, M. Ohki, C. H. Overweg, J. H. Page, P. Pernod, V. L. Preobrazhnesky, M. E. Ravicz, P. Roux, S. Shimada, V. J. Soudek, E. Q. Wang

Student to Associate

L. Wu

Member to Electronic Associate

M. A. Garces

Reinstated

J. D. Miller—*Fellow*
L. G. Hopkins—*Associate*
S. R. Hahn—*Electronic Associate*

Deceased

A. P. G. Peterson—*Fellow*
P. W. Jusczyk, R. A. Rhodes, II, J. L. Wildermuth—*Members*

Fellows	906 Members	2491 Associates	2724
Students			947
Electronic Associates			215
			7283

ACOUSTICAL NEWS—INTERNATIONAL

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

International Meetings Calendar

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

April 2004

5–9 **18th International Congress on Acoustics (ICA2004)**, Kyoto, Japan. (Fax: +81 66 879 8025; 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)

25–30 **9th Meeting of the European Society of Sonochemistry**, Badajoz, Spain. (European Society of Sonochemistry, Avenida de Europa 3, 06004 Badajoz, Spain; Web: www.ess9.com)

27–28 **Advanced Metrology for Ultrasound in Medicine**, Teddington, UK. (Fax: +44 20 8943 6735; Web: www.amum2004.npl.co.uk)

May 2004

8–10 **116th AES Convention**, Berlin, Germany. (Web: aes.org/events/116)

10–12 **Tenth AIAA/CEAS AeroAcoustics Conference**, Manchester, UK. (Web: www.aiaa.org)

17–21 **International Conference on Acoustics**, **Speech**, **and Signal Processing (ICASSP 2004)**, Montréal, Canada. (Web: www.icassp2004.com)

June 2004

6–9 **13th International Conference on Noise Control**, Gdynia, Poland. (Web: www.ciop.pl/noise_04)

8–10 **Joint Baltic–Nordic Acoustical Meeting**, Mariehamn, Åland, Finland. (Fax: +358 09 460 224; e-mail: asf@acoustics.hut.fi)

8–10 **Transport Noise and Vibration 2004**, St. Petersburg, Russia. (Web: webcente.ru/~eeaa/tn04)

July 2004

5–8 **7th European Conference on Underwater Acoustics (ECUA 2004)**, Delft, The Netherlands. (Fax: +31 70 322 9901; Web: www.ecua2004.tno.nl)

5–8 **Eleventh International Congress on Sound and Vibration (ICSV11)**, St. Petersburg, Russia. (Web: www.iiav.org)

11–16 **12th International Symposium on Acoustic Remote Sensing (ISARS)**, Cambridge, 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**, Montréal, Canada. (Fax: +1 978 927 4099; Web: www.ieee-uffc.org/index2-asp)

22–25 **Inter-noise 2004**, Prague, Czech Republic. (Web: www.internoise2004.cz)

30–1 **Low Frequency 2004**, Maastricht, The Netherlands. (G. Leventhall, 150 Craddlocks Avenue, Ashtead, Surrey KT 21 1NL, UK; Web: www.lowfrequency2004.org.uk)

September 2004

1–3 *__Subjective and Objective Assessment of Sound__, Poznań, Poland. (Institute of Acoustics, Adam Mankiewicz University, Poznań, Poland. Fax: +48 61 8295 123; Web: www.soas.amu.edu.pl/soas.html)

6–10 *__51st Open Seminar on Acoustics__; __9th School on Acousto-optics and Applications__; __Ultrasound in Biomeasurements__, Gdańsk, Poland. (University of Gdańsk, Institute of Experimental Physics, 80-952 Gdańsk, Poland; Fax: +48 58 341 31 75; Web: univ.gda.pl/~osa)

13–17 **4th Iberoamerican Congress on Acoustics**, **4th Iberian Congress on Acoustics**, **35th Spanish Congress on Acoustics**, Guimarães, Portugal. (Fax: +351 21 844 3028; Web: www.spacustica.pt/novidades.htm)

14–16 **International Conference on Sonar Signal Processing and Symposium on Bio-Sonar Systems and Bioacoustics**, Loughboro, UK. (Fax: +44 1509 22 7053 [c/o D. Gordon]; Web: ioa2004.lboro.ac.uk)

15–17 **26th European Conference on Acoustic Emission Testing**, Berlin, Germany. (DGZIP, Max-Planck-Str. 26, 12489 Berlin, Germany; Web: www.ewgae2004.de)

20–22 **International Conference on Noise and Vibration Engineering (ISMA2004)**, Leuven, Belgium. (Fax: +32 16 32 29 87; Web: www.isma-isaac.be/fut_conf/default_en.phtml)

20–22 **9th International Workshop "Speech and Computer" (SPECOM'2004)**, St. Petersburg, Russia. (Web: www.spiiras.nw.ru/speech)

28–30 **Autumn Meeting of the Acoustical Society of Japan**, Naha, Japan. (Fax: +81 3 5256 1022; Web: wwwsoc.nii.ac.jp/asj/index-e.html)

October 2004

4–8 **8th Conference on Spoken Language Processing (INTERSPEECH)**, Jeju Island, Korea. (Web: www.icslp2004.org)

6–8 **Acoustics Week in Canada**, Ottawa, ON, Canada. (J. Bradley, NRC Institute for Research on Construction [Acoustics Section], Ottawa, Ontario, K1A 0R6; Fax: +1 613 954 1495; Web: caa-aca.ca/ottawa-2004.html)

8–9 ***Reproduced Sound 20**, Oxford, UK. (Web: www.ioa.org.uk)

November 2004

3–5 ***Australian Acoustical Society Conference—Transportation Noise & Vibration**, Surfers Paradise, Queensland, Australia. (Fax: +61 7 6217 0066; Web: www.acoustics.asn.au/conference/index.htm)

4–5 **Autumn Meeting of the Swiss Acoustical Society**, Rapperswil, Switzerland. (Fax: +41 419 62 13; Web: www.sga-ssa.ch)

April 2005

18–21 ***International Conference on Emerging Technologies of Noise and Vibration Analysis and Control**, Saint Raphaël, France. (Fax: +33 4 72 43 87 12; e-mail: goran.pavic@insa-lyon.fr)

May 2005

16–20 **149th Meeting of the Acoustical Society of America**, Vancouver, B.C., Canada. (ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Fax: +1 516 576 2377; Web: asa.aip.org)

 0001-4966/2004/115(4)/1375/2/$20.00

August 2005

6–10 **Inter-Noise**, Rio de Janeiro, Brazil. (Web: www.internoise2005.ufsc.br)

28–2 **EAA Forum Acusticum Budapest 2005**, Budapest, Hungary. (I. Bába, OPAKFI, Fö u. 68, Budapest 1027, Hungary; Fax: +36 1 202 0452; Web: www.fa2005.org)

September 2005

4–8 **9th Eurospeech Conference (EUROSPEECH'2005)**, Lisbon, Portugal. (Fax: +351 213145843; Web: www.interspeech2005.org)

5–9 *__Boundary Influences in High Frequency, Shallow Water Acoustics__, Bath, UK. (Web: acoustics2005.ac.uk)

October 2005

19–21 *__36th Spanish Congress on Acoustics joint with 2005 Iberian Meeting on Acoustics__, Terrassa (Barcelona), Spain. (Sociedad Española de Acústica, Serrano 114, 28006 Madrid, Spain; Fax: +34 914 117 651; Web: www.ia.csic.es/sea/index.html)

June 2006

26–28 **9th Western Pacific Acoustics Conference (WESPAC 9)**, Seoul, Korea. (Web: www.wespac8.com/WespacIX.html)

September 2007

2–7 **19th International Congress on Acoustics (ICA2007)**, Madrid, Spain. (SEA, Serrano 144, 28006 Madrid, Spain; Web: www.ia.csic.es/sea/index.html)

Preliminary Announcement

June 2008

23–27 **Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**. Paris, France. (Details to be announced later)

ADVANCED-DEGREE DISSERTATIONS IN ACOUSTICS

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

Computation of exhaust mixing noise using large-eddy simulation turbulence modeling and Lighthill's acoustic analogy [43.28.Ra, 43.50.Nm]—David Brian Schein, *School of Engineering and Applied Science, University of California, Los Angeles, CA January 2003 (Ph.D.).* This research involves computational study of free, heated jet flow and resultant far-field sound performed using large-eddy simulation (LES) and Lighthill's acoustic analogy. A subgrid-scale model for small-scale compressible turbulence was developed using a combination of the popular Smagorinsky model and a deductive model. An existing software package for compressible flow field computation was substantially modified to perform temporal LES and aerosound simulations. Cases studied extend to large Reynolds number (Re), high subsonic (compressible) flow with realistic geometries more representative of aircraft engine exhausts than typically considered using direct numerical simulation (DNS). Flow-field fluctuations are stored over a period of time and used to calculate rms turbulence within the computational domain. The far-field sound and directivity is computed using the time-derivative form of Lighthill's source-integral result formulated in terms of quadrupole sources from the simulated flow field, which is integrated in time and contains the fluctuations set up by the time-varying stress tensor. A simulation for a WR19-4 turbofan engine exhaust ($Re>106$ based on exit velocity and diameter) is presented, and propagated jet noise results are compared with experimental acoustics data.

Advisor: William C. Meecham

Invariant patterns in articulatory movements [43.70.Bk, 43.70.Fq, 43.70.Jt]—Patrizia Bonaventura, *Department of Speech and Hearing, The Ohio State University, Columbus, OH, December 2003 (Ph.D.).* The purpose of the reported study is to discover an effective method of characterizing movement patterns of the crucial articulator as the function of an abstract syllable magnitude and the adjacent boundary, and at the same time to investigate effects of prosodic control on utterance organization. In particular, the speed of movement when a flesh point on the tongue blade or the lower lip crosses a selected position relative to the occlusion plane is examined. The time of such crossing provides an effective measure of syllable timing and syllable duration according to previous work. In the present work, using a very limited vocabulary with only a few consonants and one vowel as the key speech materials, effects of contrastive emphasis on demi-syllabic movement patterns were studied. The theoretical framework for this analysis is the C/D model of speech production in relation to the concept of an invariant part of selected articulatory movements. The results show evidence in favor of the existence of ''iceberg'' patterns, but a linear dependence of slope on the total excursion of the demisyllabic movement, instead of the approximate constancy of the threshold crossing speed as suggested in the original proposal of the iceberg, has been found. Accordingly, a revision of the original concept of iceberg seems necessary. This refinement is consistent with the C/D model assumption on ''prominence control'' that the syllable magnitude determines the movement amplitude, accompanying directly related syllable duration change. In this assumption, the movement of a consonantal component should also be proportional to syllable magnitude. The results suggest, however, systematic outliers deviating from the linear dependence of movement speed on excursion. This deviation may be caused by the effect of the immediately following boundary, often referred to as phrase-final elongation.

Advisor: Osamu Fujimura

An analytic model for acoustic scattering from an impedance cylinder placed normal to an impedance plane [43.20.Fn, 43.28.En]—Michelle E. Swearingen, *Graduate Program in Acoustics, The Pennsylvania State University, State College, PA 16804. May 2003 (Ph.D.).* An analytic model, developed in cylindrical coordinates, is described for the scattering of a spherical wave off a semi-infinite right cylinder placed normal to a ground surface. The motivation for the research is to have a model with which one can simulate scattering from a single tree and which can be used as a fundamental element in a model for estimating the attenuation in a forest comprised of multiple tree trunks. Comparisons are made to the plane wave case, the transparent cylinder case, and the rigid and soft ground cases as a method of theoretically verifying the model for the contemplated range of model parameters. Agreement is regarded as excellent for these benchmark cases. Model sensitivity to five parameters is also explored. An experiment was performed to study the scattering from a cylinder normal to a ground surface. The data from the experiment is analyzed with a transfer function method to yield frequency and impulse responses, and calculations based on the analytic model are compared to the experimental data.

[DOI: 10.1121/1.1648679]

Thesis advisor: David C. Swanson

BOOK REVIEWS

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

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

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

Cochlear Implants: Objective Measures

Helen E. Cullington (Editor)

Whurr Publishers/Taylor & Francis, London, 2003
241 pp. Price: $46.95 (softcover), ISBN: 1861563248.

This book contains an excellent collection of papers on objective measures for assessing and managing patients with cochlear implants. In recent years, cochlear implants have become increasingly more successful and have gained wide use around the world. Some of the challenges still facing the clinician include confirmation of candidacy, verification of the integrity of the device, and determining ideal programming parameters for individual patients. While some patients are able to provide subjective feedback, many, including children and those with disabilities, cannot. Determining appropriate stimulation levels for each patient can be challenging and time-consuming. A number of objective measures have therefore been developed or adapted that are aimed at facilitating and standardizing clinical practice. This book is intended as a guide to clinicians on objective measures that can be used to determine preimplantation candidacy as well as postimplantation device performance. Clinical availability of such tools is extremely important for achieving high success rates in implant patients. An introductory chapter by the editor provides a good anchor and framework for the chapters that follow, with a helpful table for summarizing the main functions provided by each objective measure.

The seven chapters that follow are written by expert clinicians and researchers. Each author pays close attention to the integrity of the measurements and their feasibility in a clinical setting. They include definitions of terms, evidence from clinical and basic research, and suggestions for clinical fitting strategies. A number of the authors specifically discuss which of the currently available implants can be evaluated using their measures. A particularly effective structure is the inclusion of a literature review, description of methods, required equipment, and procedures for administration of the measures. This book is not intended as a detailed research-oriented resource. The chapters generally refer to the research that ultimately provides guidelines for parameters that are most useful clinically. However, it is often made clear that a larger body of research exists on most topics, beyond the scope of the book. The book is therefore particularly helpful as a guide for one who is skilled in the art of cochlear implant mapping and evaluation using behavioral techniques. In addition, it can serve as a useful resource for researchers who may not be intimately familiar with objective methods. Having all the current objective methods presented in a cohesive package is extremely valuable. The manageable length of each chapter, and the care that the authors generally take in providing outlines and summaries, is effective for comparison of the various methods, their pros and cons.

To date, there are three cochlear implant devices with FDA approval in the USA: Clarion, Med-El, and Nucleus. Laura and MXM are two other devices discussed in a number of chapters as well. Not all devices can be evaluated with each objective method, and the applicability of each method to the various devices is apparent in each of the chapters. For example, different methods of telemetry allow for measurement of the link between the radio frequency output and link to the implant, electrode-tissue interface, and evoked activity. These measures depend on device-specific components supplied by the manufacturer; while the first two features are available for the three common devices, the latter one is not. The electrically evoked stapedial reflex measure does not depend on device-specific interfaces, and can be applied to any of the devices. Similarly, electrically evoked brainstem, middle latency, and cortical responses can serve as a powerful tool for showing evidence of cochlear-implant driven responses at various levels within the auditory system. Measures at the levels of the brainstem offer information regarding the presence of intact auditory neurons and candidacy, as well as postoperative assessment and management, and have been widely used in young children. Measures at the level of the cortex are important for understanding how the use of an implant affects higher brain functions such as language. By learning about which objective method may be readily available or implementable, clinicians can refer to this book as a guide for determining which device may be most appropriate for each patient.

RUTH LITOVSKY
Waisman Center
1500 Highland Avenue
University of Wisconsin—Madison
Madison, Wisconsin 53705-1103

Cochlear Implants for Young Children (2nd edition)

Barry McCormick and Sue Archbold, Editors

Whurr Publishers Ltd., London and Philadelphia, 2003.
425 pp. Price: $69.95 (softcover), ISBN: 1861562187.

Cochlear implantation of profoundly deaf children has become commonplace as the benefits of the implant have become more understood. Deaf children less than a year old increasingly receive cochlear implants, and in many cases are being mainstreamed successfully within a few years postimplantation. As the number of children receiving cochlear implants continues to increase, there is an increased urgency in establishing guidelines for the clinical management of cochlear-implanted children. Cochlear implants, in general, present challenges for all involved (the recipients and their families, the doctors, audiologists, research and medical professionals, educators, etc.); cochlear implants in children present unique challenges. Methods of pre- and postimplantation assessment, evaluation, and progress monitoring must all be modified considerably from the adult model to be usable for the infant or the growing child. In the second edition to their book, *Cochlear Implants for Young Children*, editors Barry McCormick and Sue Archbold, both of the Nottingham Paediatric Cochlear Implant Programme, provide a framework to address these challenges. Pre- and postimplant evaluation methods, surgical techniques, electrophysiological measures, and rehabilitation techniques are discussed in the book's 11 chapters, as well the emotional effects of hearing loss and implantation on children and their families. This volume is likely to become a definitive guide for clinicians working with cochlear-implanted children.

The Nottingham Paediatric Cochlear Implant Programme is the largest cochlear implant center for children in the U.K. Since its inception in 1989, the Nottingham center has implanted more than 400 children, mostly between 3 and 5 years old. In the first edition of this book (1994), the emerging techniques of an evolving program were collected as a guide for other hearing centers considering cochlear implantation of children. In the 9 years since the first edition, implant technology has improved dramatically, and information regarding surgical and rehabilitation outcomes, as well as de-

 0001-4966/2004/115(4)/1385/2/$20.00

velopmental data with implanted children, has greatly increased, thereby necessitating a second edition. Indeed, this second edition offers much to both hearing professionals and interested family members.

Assessing a child's preimplantation auditory status can pose significant difficulties, particularly for very young children with severe hearing impairment. Barry McCormick provides a comprehensive review of different methodologies used at Nottingham and elsewhere in his chapter, ''Assessing Audiological Suitability Of Cochlear Implants For Children Below The Age Of Five Years.''

Cochlear implant technology continues to evolve, and the latest implant and speech processor designs are adequately reviewed in the second chapter, ''Cochlear Implant Systems,'' by Sara Flynn. Sound via the cochlear implant may be distorted (relative to the normal-hearing experience), but in the case of the developing auditory system, neither the degree of perceived distortion nor its effects are well understood. In general, the same multichannel speech-processing strategies employed for adults are provided for children. There are additional difficulties, however, in ensuring appropriate stimulation levels and tuning the processor to achieve acceptable sound quality for a child. The chapter ''Fitting and Programming the External System,'' by Yvonne Cope and Catherine Totten, covers a wide range of relevant issues regarding speech processor fitting for children.

In addition to audiologic evaluations, electrophysiological measures can provide objective measures of hearing loss (before implantation), implant functionality (during surgery), and device tuning (after implantation). With young children, it is difficult to fit a speech processor, as they may not reliably communicate if a sound is inaudible or too loud. However, it may be harmful to the child's auditory development if the speech processor is not tuned properly. Electrophysiological measures such as neural response telemetry can guide the fitting of the speech processor. Steve Mason's chapter, ''Electrophysiological and Objective Measures,'' provides a good overview of the many measures used to assess implant recipients' responses to electrical stimulation.

The appropriateness of implant technology (originally developed for postlingually deafened adults) for profoundly deaf children is often approached by a medical model, as Sue Archbold describes in her chapter ''A Paediatric Cochlear Implant Programme.'' The risks, costs, and efficacy of implantation are better understood by many medical professionals than the attendant social, educational, and even emotional effects. This lack of sensitivity to deaf issues contributed to the polarization of the deaf community, many of whom ''do not see deafness as a medical condition requiring treatment, but as a linguistic and cultural identity.'' The Nottingham center's approach to cochlear implantation is decidedly broad-based and includes, as part of the implant team, medical professionals, speech therapists, educators, and most importantly, the child's family. This multidisciplinary approach is vital to the child's success with the implant. The family's role in this success cannot be underestimated.

Besides the family's responsibility in the seemingly endless logistics of the implant program (shepherding the child to the many appointments with audiologists, surgeons, therapists; making sure the child wears the device; ensuring that the device is working properly and obtaining repairs; etc.), the family plays a pivotal role in the auditory rehabilitation of implanted children. The Nottingham center has developed a number of age-appropriate rehabilitation methods in which the family participates. Actually, as authors Sue Archbold and Margaret Tait point out in their chapter ''Facilitating Progress after Cochlear Implantation: Rehabilitation—Rationale and Practice,'' *habilitation* may be the more appropriate description for a child's auditory development with a cochlear implant. To enhance this development, the Nottingham center advocates ''active learning'' by the child, rather than strictly employing training techniques that are effective in the *re*-acquisition of speech and hearing skills. For young implant recipients, ''incidental learning,'' in which games and activities include some sort of auditory reinforcement, may best promote auditory development.

The Nottingham group also makes use of video to monitor chidren's progress, as described in another chapter by Margaret Tait. By incorporating video analysis in longitudinal outcome studies, researchers found that implanted children were able to develop speech production that was similar to better hearing aid users; however, the implanted children were able to achieve these levels more quickly than hearing aid users. Video was also useful in predicting children's pre- and postimplant performance. The Nottingham group uses other measures to assess children's progress, as described in Dee Dyar's and Thomas P. Nikolopoulos' chapter, ''Monitoring Progress: The Role and Remit of a Speech and Language Therapist.'' Because cochlear implant outcomes remain variable and because of the added difficulty in measuring performance in children, standard assessment protocols are important in monitoring an implanted child's progress. The appropriateness, efficiency, and reliability of these protocols are important in determining each child's auditory, cognitive, and neurolinguistic level of performance, both before and after implantation.

In general, the book provides a strong overview of the many issues relating to cochlear implants and children. Of course, the impact of these issues will be most keenly felt by the children and their families. The book's final chapter, ''Family Perspectives,'' by Hazel Lloyd Richmond, is especially insightful about the effects of deafness and implantation on the family dynamic. Through case studies and accounts written by the parents of implanted children, a wide range of emotional and psychological responses is presented, including the enormous stress felt by parents at the time of diagnosis, surgery, and initial hook-up of the implant. The decision to implant a child is in itself stressful, and many parents agonize about whether the implant is the best solution to the communication and social challenges their child will face in life. Fortunately, the Nottingham program's integration of the medical, therapeutic, education, and family communities provides a strong foundation, so that families can receive the support needed for successful outcomes. The Nottingham group's experiences with implanted children, along with the many recent studies cited in the chapters, will be invaluable guides in this continually developing area of research and clinical practice.

JOHN J. GALVIN III and MONITA CHATTERJEE
Department of Auditory Implants and Perception
House Ear Institute
2100 W. Third St.
Los Angeles, CA 90057

 0001-4966/2004/115(4)/1386/2/$20.00

OBITUARIES

Arnold P. G. Peterson • 1914–2003

Arnold Per Gustaf Peterson was born in DeKalb, Illinois. His parents had immigrated from Sweden in 1912. His first language was Swedish, but on entering school English prevailed and he eventually lost his fluency in Swedish. When he was three years old his family moved to Detroit. In school, he was good at arithmetic, as well as reading and spelling, and skipped half a grade. He was small for his age and not athletically inclined. In junior high school he did very well in mathematics and was above average in civics, history, and Latin. As a hobby he built a four-tube receiver from a kit and built a power supply for it, which started him on his career as an electronics engineer. He graduated summa cum laude from Cass Tech in Detroit when he was 16 and he enrolled at the University of Toledo (Ohio) in the school of engineering. He soon got the reputation as a whiz in mathematics. He graduated summa cum laude in June 1934 with a Bachelor of Engineering, at less than 20 years of age.

Arnold's mathematics professor in Toledo suggested he visit the Massachusetts Institute of Technology (MIT). His father agreed and said that if he was accepted at MIT, the finances would be possible. He was immediately interviewed, accepted, and enrolled. He shifted his interests to communications (now electronics). In the Fall of 1935, he received a ''Research Assistant'' position that paid $25 a week. This MIT arrangement allowed him to work at the General Radio Company (GR), a few blocks from MIT, and continue taking courses, part time. His first project at GR was to develop an oscillator for measurement purposes, which led to a tunable ultra-high-frequency oscillator, that General Radio produced as a product. He used this work for the thesis that led to his Master of Science degree in 1937.

He immediately began work toward his doctorate at MIT. His research was done largely at General Radio, and concentrated on achieving accuracy in voltmeters over a wide frequency range. His thesis was completed in the Fall of 1940, but his degree was formally presented at Spring commencement in June 1941.

Arnold became an employee of General Radio in September 1940. His first assignment at GR was working with Eduard Karplus and they co-invented the ''butterfly circuit'' for a wide range oscillator. During World War II, he worked on a series of search radio receivers, which were mass produced in the thousands, and on other classified projects. Arnold met his wife Carolyn Bryant, a native of Seattle, in 1942 and they were married in Seattle in 1943. They rented an apartment a few blocks from General Radio in Cambridge.

In the first couple of years after the war, Arnold worked on various projects. In 1947, General Radio decided that they should develop a line of modern sound measuring equipment. It was then that I met Arnold and began what became a long period of work and association with him. General Radio hired me as a one-day-a-week consultant, because of the experience I had accumulated at Harvard University during the war as Director of the Electro-Acoustic Laboratory. I remember entering his office each Friday and first turning on the lights—Arnold liked a very low level of illumination. Initially, we planned and executed three instruments, the 1551-A sound level meter, the 1550-A octave band analyzer, and the 1552-A sound level calibrator. I was always amazed at his incisive mind and his ability to learn a new field in short order. Following that, we attempted to convince the management that they should develop a line of about ten instruments, but they said that they expected a depression in a few years, and did not want to be over committed. No depression developed and the restriction was later removed. By the time Arnold retired from GenRad in 1979, their catalog showed 53 instruments falling into four categories: (1) hearing conservation, (2) community noise measurements, (3) product noise reduction, and (4) other instruments, transducers, and accessories. Arnold's hand was in all of these. General Radio later moved from Cambridge to Concord. A further move for Arnold was to a facility they built in Bolton, Massachusetts.

Most acousticians remember the long series of ''Handbook of Noise Measurement'' books issued by General Radio. Arnold and I produced the first three in 1950, 1954, and 1958. After I resigned from MIT and moved to BBN's office in Harvard Square, Peterson and Ervin Gross produced the fourth through the eighth editions. Arnold produced the ninth edition alone after Ervin Gross retired in 1978. Around that time the company changed its name to GenRad, Inc.

Arnold joined the Acoustical Society of America in 1946 and was elected a Fellow in 1951. He served in many ASA voluntary positions through the years, Executive Council (1953–56), Vice President (1958–59), Standards Adviser, Chair of the Membership Committee, and other Society committees. Elaine Moran says, ''He was a member continuously for 57 years, volunteered his time and expertise, and helped make the Acoustical Society a strong and flourishing organization.''

When Arnold retired from GenRad, he and Carolyn moved to Bainbridge Island, Washington. They spent considerable time exploring Kitsap County and the state of Washington. Arnold designed their house on Bainbridge Island and had it built on a waterfront lot. He joined the Ham Radio group on the Island. Arnold and Carolyn enjoyed genealogy research. They published nine books on family history and contributed articles to various County Genealogy journals as well as the *New England Historical Genealogical Journal*.

Arnold is survived by his wife, Carolyn; three children, Duncan, Alan, and Janet, and five grandchildren.

LEO L. BERANEK

 0001-4966/2004/115(4)/1387/1/$20.00

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:

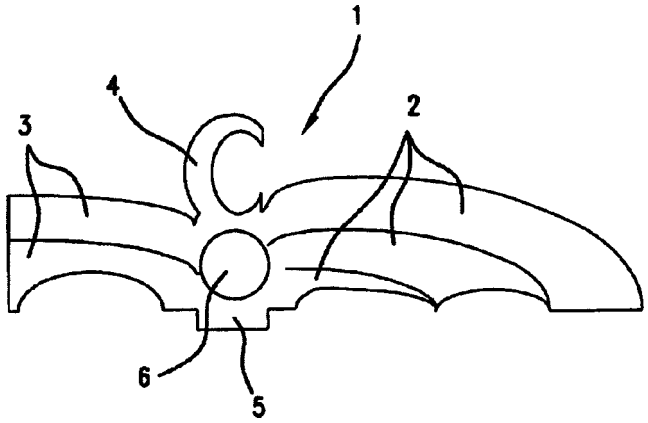
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
RICHARD S. MCGOWAN, *CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Univ. of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,627,805

43.20.Rz AUDIO DEVICE AND METHOD INCLUDING A MEMBRANE HAVING AT LEAST TWO TONGUES WITH DIFFERENT RESONANT FREQUENCIES

Marc Charbonneaux *et al.*, assignors to Pica–Sound International
30 September 2003 (Class 84/408); filed in France 24 August 1998

Did you realize that the circular radiators of cone loudspeakers and ribbons of electrostatic transducers ''don't create sound that is correct to the



human ear?'' The proposed solution is a multifinger diffuser with a hole **6**. The absence of theory and measurement makes this just so much wishful thinking.—MK

6,647,159

43.35.Sx TENSION-TUNED ACOUSTO-OPTIC BANDPASS FILTER

Duane Anthony Satorius, assignor to The United States of America as represented by The National Security Agency
11 November 2003 (Class 385/7); filed 25 January 2002

This is a chirped fiberoptic band-pass filter that does not use core block and that is said to use less electric power. It can accommodate multiple acoustic signals with individually controllable intensities and frequencies. One embodiment has a tension-tuned acousto-optic band-pass filter in which one or more bands of optical wavelengths may be selected for further transmission. All light within the optical bandwidth is first coupled from the core mode of an optical fiber to a specific cladding mode by a chirped broad-band cladding mode coupler. These light waves then enter a narrow-band core mode coupler where selected optical bands, tuned by the tension on the optical fiber, are recoupled back into the core of the optical fiber. An acoustic absorber limits the acoustic interference between the chirped broad-band cladding mode coupler and the narrow-band core mode coupler.—DRR

6,647,196

43.35.Sx OPTICAL DEVICE HAVING OPTICAL WAVE GUIDE PRODUCED IN THE PRESENCE OF ACOUSTIC STANDING WAVE

Hikaru Kouta, assignor to NEC Corporation
11 November 2003 (Class 385/129); filed in Japan 4 April 1996

This device incorporates an optical waveguide. An electro-acoustic transducer is attached to a block of optical material and generates an acoustic standing wave within the block. The standing wave changes the refractive index of the optical material in a periodic pattern, forming layered segments with alternating high and low refractive index. The highly refractive segments serve as optical waveguides, thus enabling the electro-acoustic transducer to produce an optical waveguide in the block.—DRR

6,641,534

43.35.Yb METHODS AND DEVICES FOR ULTRASOUND SCANNING BY MOVING SUB-APERTURES OF CYLINDRICAL ULTRASOUND TRANSDUCER ARRAYS IN TWO DIMENSIONS

Stephen W. Smith *et al.*, assignors to Duke University
4 November 2003 (Class 600/437); filed 25 January 2002

This two-dimensional transducer array includes at least one row of ultrasound transducer elements arranged along a curved surface and at least one column of ultrasound elements arranged in a straight line on that curved

 0001-4966/2004/115(4)/1391/18/$20.00

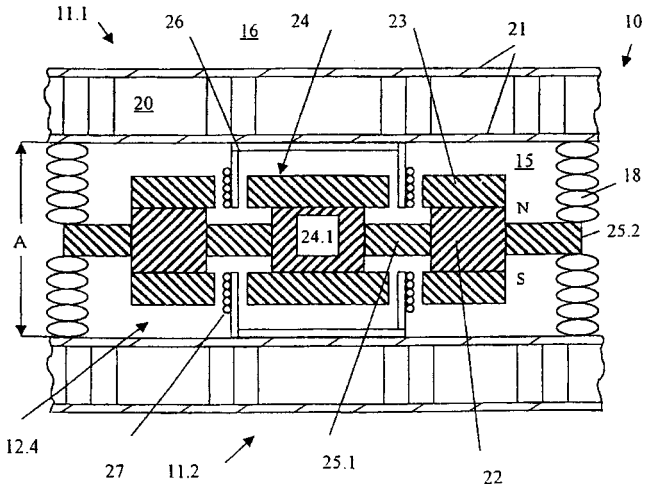
surface. Scanning is achieved by defining a subaperture of the 2-D ultrasound transducer array (which includes multiple transducer elements in each dimension of the array) and exciting the ultrasound elements included in the subaperture to generate a transmit–receive ultrasonic beam.—DRR

6,622,817

43.38.Ja SOUND REPRODUCTION DEVICE WORKING ACCORDING TO THE BENDING WAVE PRINCIPLE

Wolfgang Bachmann *et al.*, assignors to Harman Audio Electronic Systems GmbH
23 September 2003 (Class 181/150); filed in Germany 15 May 1998

At low frequencies, the panel of a bending wave loudspeaker behaves more like a conventional diaphragm, which means that some kind of baffle or enclosure must be provided to prevent a short circuit between front and



rear radiation. This patent describes several methods of driving two diaphragms **11.1**, **11.2** in opposition, with a confined air space between them. In the example shown, a common magnetic structure accomodates individual voice coils **26**, **27**. The enclosed air volume can be vented or coupled to a transmission line.—GLA

6,625,291

43.38.Ja DIFFUSED RESONANCE LOUDSPEAKER ENCLOSURE METHOD

David S. Mohler, assignor to Avaya Technology Corporation
23 September 2003 (Class 381/337); filed 6 April 1999

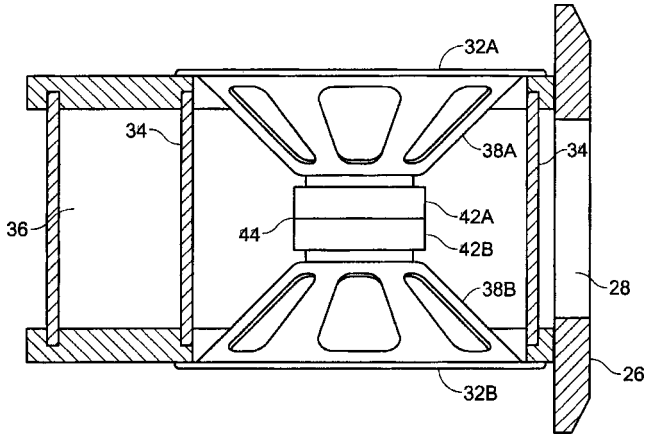
From time to time experimenters come up with the idea of mounting a loudspeaker on one end of a tapered pipe. Some inventors believe that this arrangement saves space without affecting pipe resonances. This patent takes the opposite stance, arguing that standing waves are suppressed because pipe length is indeterminate. In fact, neither statement is true.—GLA

6,628,792

43.38.Ja BACK TO BACK MOUNTED COMPOUND WOOFER WITH COMPRESSION/BANDPASS LOADING

Paul W. Paddock, McMinnville, Oregon
30 September 2003 (Class 381/182); filed 30 March 1999

Two woofers are mounted back-to-back in a self-contained module that can be mounted in a conventional loudspeaker enclosure. This is a practical, useful arrangement for cancelling low-frequency mechanical vibrations. The design also makes use of a constrained, spiral duct to add



dynamic air mass and thus lower the effective system resonance frequency. According to the patent, this somewhat complicated geometry ''... minimizes size, material, and electrical and custom components, [and] is inexpensive to manufacture.''—GLA

6,628,798

43.38.Ja VIBRATION ACTUATOR HAVING THREE VIBRATION MODES

Makoto Teshima *et al.*, assignors to NEC Tokin Corporation
30 September 2003 (Class 381/396); filed in Japan 13 April 1999

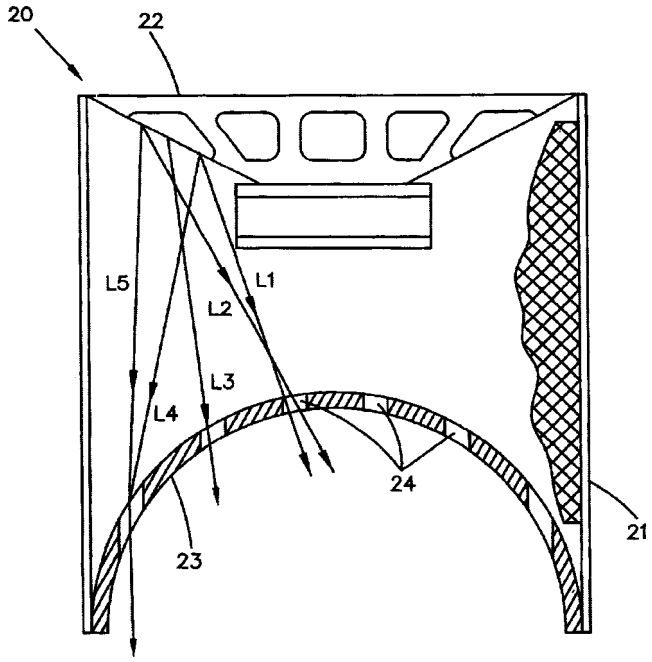
Cellular phones require miniature transducers to reproduce voice frequencies, generate tactile vibrations, and produce an audible beep tone. This 29-claim patent describes a single-coil transducer that can selectively or simultaneously provide all three functions.—GLA

6,636,610

43.38.Ja LOUDSPEAKER SYSTEMS

Robert W. Betts, assignor to Sonic Systems, Incorporated
21 October 2003 (Class 381/345); filed 19 October 2000

Enclosure **21** is cylindrical. Rear panel **23** is hemispherical. Distributed vents **24** are located at varying path lengths from woofer **22**. ''As a result, the internal volume of air within the enclosure of the present invention will not be resonant at any one specific frequency. This causes the



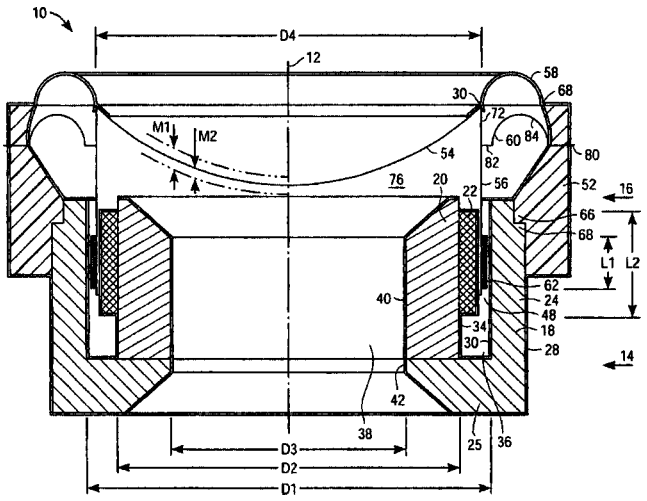
enclosure to be effectively or virtually acoustically infinite in volume.'' Oh, and it also eliminates harmonic, intermodulation, and phase distortion.—GLA

6,636,612

43.38.Ja SPEAKER FOR USE IN CONFINED SPACES

Donald J. North and Dai Z. Lee, assignors to Algo Sound, Incorporated
21 October 2003 (Class 381/412); filed 3 November 2000

This small, edge-driven loudspeaker features a self-shielded magnetic structure and relatively long cone travel. Dual suspensions **58** and **60** coun-



teract rocking motion. Unlike similar small speakers, the frame and magnetic assembly form a cylindrical enclosure and backwave sound pressure is vented exclusively through the central bore hole.—GLA

6,625,293

43.38.Kb MICROPHONE

Koji Nageno and Naoaki Matsumoto, assignors to Sony Corporation
23 September 2003 (Class 381/362); filed in Japan 27 July 1999

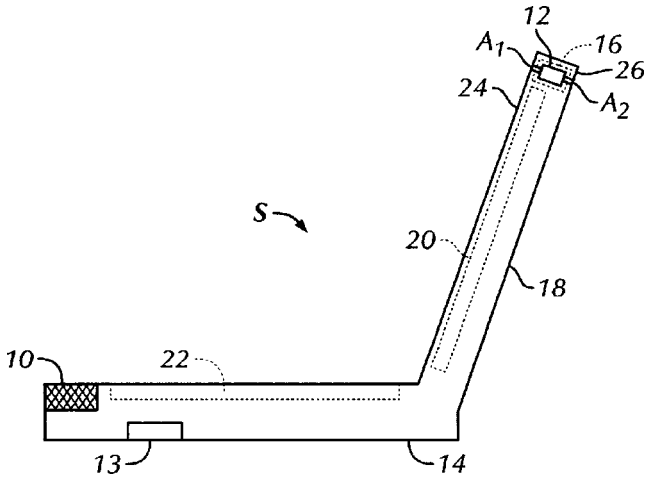
Headset boom microphones are sometimes designed like probe microphones, using a thin tube to conduct sound from a pickup location to a transducer at the other end of the tube. Internal damping can smooth out pipe resonances but also attenuates high-frequency transmission. Instead of internal stuffing, this invention substitutes a damped air leak near the microphone end of the tube. The result is shown to be an extended frequency response characteristic that actually accentuates the 5-kHz region.—GLA

6,633,647

43.38.Kb METHOD OF CUSTOM DESIGNING DIRECTIONAL RESPONSES FOR A MICROPHONE OF A PORTABLE COMPUTER

Mitchell A. Markow and David E. Gough, assignors to Hewlett–Packard Development Company, L.P.
14 October 2003 (Class 381/92); filed 30 June 1997

Almost all laptop computers include one or two built-in microphones, usually omnidirectional elements recessed into the computer base. But suppose that bidirectional elements were instead embedded near the top corners



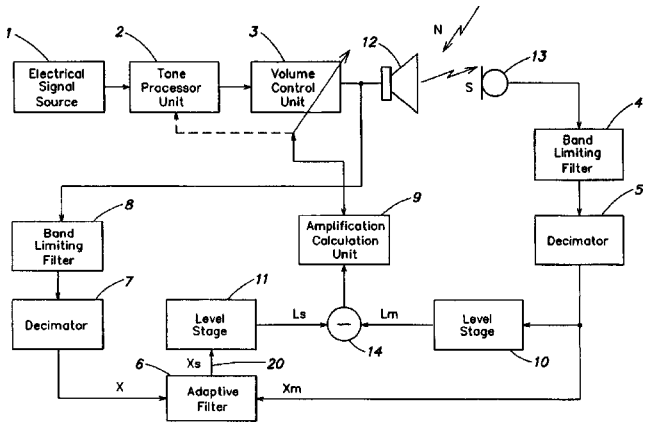
of the hinged lid and provided with front and rear sound openings. By varying the aperture areas, the location of the microphone elements, and internal damping, a desired directional response can be achieved.—GLA

6,628,788

43.38.Lc APPARATUS AND METHOD FOR NOISE-DEPENDENT ADAPTATION OF AN ACOUSTIC USEFUL SIGNAL

Seved Ali Azizi, assignor to Becker GmbH
30 September 2003 (Class 381/57); filed in Germany 27 April 2000

Over the past 50 years a number of devices have been designed to automatically adjust the level of a sound system to compensate for the presence of background noise. Applications include airport announcements, industrial paging, and automotive sound. The trick is to sense unwanted noise while ignoring the output of the sound system itself. The method described here compares the relative level of a band-limited ''useful'' signal



with that of the overall signal, including noise. According to the patent, the circuitry shown in the block diagram avoids the ''additional complexity'' of prior art.—GLA

6,627,807

43.38.Md COMMUNICATIONS APPARATUS FOR TONE GENERATOR SETTING INFORMATION

Satoru Motoyama, assignor to Yamaha Corporation
30 September 2003 (Class 84/615); filed in Japan 13 March 1997

Electronic musical instruments transmit key on and key off events via the MIDI standard. This patent concerns transmitting MIDI streams via the Internet. The strategy outlined includes the obvious time and sequence stamping along with a checksum. Strangely, for a patent that purports to be about an Internet protocol, there is no mention of the TCP/IP stack. For Yamaha, the patent is well written and reasonably complete.—MK

6,628,585

43.38.Md QUADRAPHONIC COMPACT DISC SYSTEM

Thomas Bamberg, North, South Carolina
30 September 2003 (Class 369/47.16); filed 13 October 2000

Even at this late date, some audiophiles don't want to give up quad. This one wants to change the well-established CD format to include four channels by using pairs of CD tracks. The issues of format compatability and the psychoacoustic imaging problems of quad are all conveniently neglected.—MK

6,636,170

43.38.Md SIGNAL CONVERTER WITH DIGITAL/ANALOG CONVERSION AND NOISE-DEPENDENT VARIABILITY OF OUTPUT SIGNAL BANDWIDTH

Eise C. Dijkmans, assignor to Koninklijke Philips Electronics N.V.
21 October 2003 (Class 341/144); filed in the European Patent Office 26 March 1999

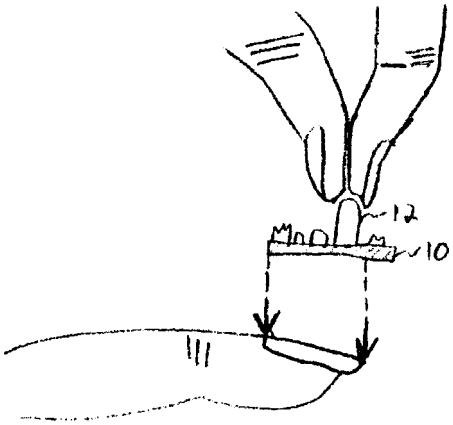
Setting a record for brevity, this obvious patent states that a suitable low-pass filter is needed on the output of a SACD DAC. No back references, no citations, and its obvious nature should have made it shameful.—MK

6,631,723

43.38.Ne ARTIFICIAL NAILS WITH THREE DIMENSIONAL FEATURES

Keith A. Mullin, assignor to Make Ideas, Incorporated
14 October 2003 (Class 132/73); filed 1 September 2000

You can attach almost anything to finger and toe nails. Why not a small soundtrack (among a list of possible 3-D reliefs)? This would be much shorter than the acoustic straw described in United States Patent 6,512,727



[reviewed in J. Acoust. Soc. Am. **114**(1), 28 (2003)]. The patent sets a new low standard in the quality of the illustrations.—MK

6,624,873

43.38.Vk MATRIX-ENCODED SURROUND-SOUND CHANNELS IN A DISCRETE DIGITAL SOUND FORMAT

Raymond E. Callahan, Jr. and Ioan R. Allen, assignors to Dolby Laboratories Licensing Corporation
23 September 2003 (Class 352/27); filed 5 May 1998

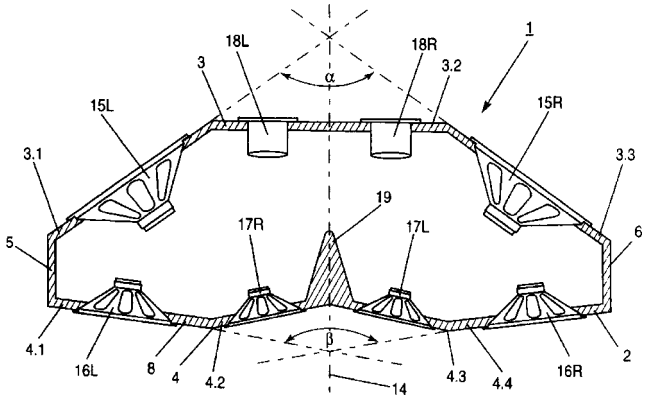
Dolby Laboratories pioneered the art of matrixing two discrete stereo recording channels into four playback channels—left, right, center, and surround. More than a dozen U.S. patents describe sophisticated encoding, steering, and noise reduction methods incorporated into the ''Dolby Stereo™'' system. It seems obvious that the same basic process could be applied to discrete surround sound channels. That is what is patented here; not any particular method, but simply the idea of matrixing two discrete surround channels into three, four, or five surround sound playback channels. In contrast to the patent title, the claims encompass both analog and digital recording.—GLA

6,625,289

43.38.Vk STEREO LOUDSPEAKER SYSTEM

Robert J. Oliemuller, assignor to Triple W Trading B.V.
23 September 2003 (Class 381/182); filed in the Netherlands 29 November 1995

A cross-section of this pedestal-mounted, multiple-driver speaker system looks more than a little like the original Bose ''direct-reflecting'' design. In this case, however, left and right channels are reproduced by left and right groups of loudspeakers housed in a single enclosure with a common back chamber. Sound from woofers **15L**, **15R** and tuned vents **18L**, **18R** bounces



off an adjacent room wall or attached reflector. Crossover frequencies are not mentioned in the patent, but one would expect the woofers to reproduce frequencies up to 500 Hz or higher for the arrangement to work very well.—GLA

6,631,193

43.38.Vk AUDIO SYSTEM ENHANCEMENT USING PSYCHO ACOUSTIC MATRIX

Francis Allen Miller, assignor to Kentech
7 October 2003 (Class 381/98); filed 7 January 1999

The main objective is to widen the perceived width of a two-channel stereo playback system without compromising the phantom center image. According to the patent, ''For all practical purposes, human hearing (in the audio frequency spectrum from about 200 Hz to approximately 7,000 Hz) functions much the same as a two-sensor phased interferometer.'' This leads to an analysis of perceived source direction in terms of relative power rather than sound pressure. More importantly, the generally ignored relationship between interaural delay and the perceived phantom center image is examined and then employed as a major element in psychoacoustic processing. The concept is interesting and the patent is clearly written.—GLA

6,633,648

43.38.Vk LOUDSPEAKER ARRAY FOR ENLARGED SWEET SPOT

Jerald L. Bauck, Tempe, Arizona
14 October 2003 (Class 381/303); filed 12 November 1999

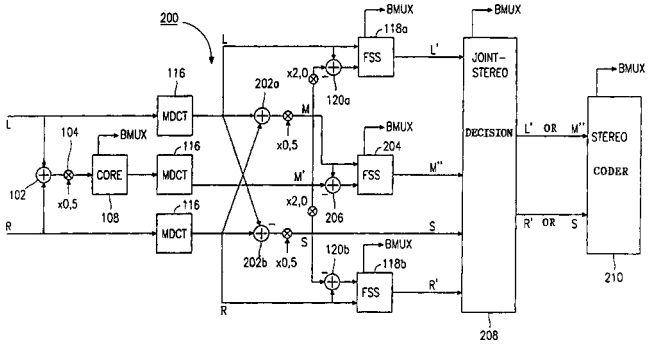
The patent document contains 12 pages of diagrams and four pages of exposition followed by a fairly succinct description of the invention. Left and right stereo loudspeakers are symmetrical two- or three-way systems with higher-frequency drivers on the inside and woofers on the outside. Suitable electronic crosstalk cancellation techniques are included. The arrangement is said to provide stable imaging over an enlarged listening area while minimizing low-frequency power demands.—GLA

6,629,078

43.38.Wl APPARATUS AND METHOD OF CODING A MONO SIGNAL AND STEREO INFORMATION

Bernhard Grill *et al.*, assignors to Fraunhofer–Gesellschaft zur Forderung der angewandten Forschung e.V.
30 September 2003 (Class 704/500); filed in Germany 26 September 1997

A scalable digital audio coder typically comprises a low bit-rate voice coder in cascade with a higher-grade audio coder. An input signal sampled at, say, 48 kHz is downsampled and fed to the voice coder, resulting in a modified signal of limited bandwidth, typically about 4 kHz. To transmit higher frequencies, a difference signal is formed, quantitized, and coded as



needed. Known scalable coders are restricted to monophonic signals. The inventors suggest that a coder designed specifically for joint stereo (M/S coding, for example) can achieve additional bit savings. Mono and stereo information can be derived and written into the bit stream in a way that results in a complete coded mono layer plus a layer with coded stereo information.—GLA

6,625,848

43.40.Kd STRIKING IMPLEMENT WITH IMPROVED ENERGY STORAGE AND VIBRATION DAMPENING PROPERTIES

Terry L. Schneider, Puyallup, Washington
30 September 2003 (Class 16/436); filed 5 October 2000

The reversible strain properties of shape memory alloys are used in such implements as golf clubs and hammers to store and release significant amounts of energy with less effort exerted by the implements' users. The shaft of an implement is divided into two or more portions, which are interconnected via elements that consist of a shape memory alloy in such a way that these elements flex when the implement strikes an object.—EEU

6,633,821

43.40.Le SYSTEM FOR SENSING FACTORY WORKSPACE

Warren B. Jackson *et al.*, assignors to Xerox Corporation
14 October 2003 (Class 702/56); filed 8 January 2001

This system for the collection of data for the purpose of workflow planning and fault diagnosis in a manufacturing process employs a sound or vibration sensor attached to each machine. A processor is used to monitor action of each machine and to follow each job via a tag that may consist of a rf, bar-code, or other suitable device.—EEU

6,634,457

43.40.Tm APPARATUS FOR DAMPING ACOUSTIC VIBRATIONS IN A COMBUSTOR

Christian Oliver Paschereit *et al.*, assignors to Alstom (Switzerland) Limited
21 October 2003 (Class 181/229); filed in Germany 26 May 2000

Acoustic vibrations in a combustor, such as that of a gas turbine, are attenuated by means of a Helmholtz resonator whose ''neck'' communicates with the combustor. The main volume of the resonator can be changed by adding or draining fluid via a supply line, so that the resonator's natural frequency can be adjusted without movable parts needing to pass through the combustor's envelope.—EEU

6,634,698

43.40.Tm VIBRATIONAL REDUCTION SYSTEM FOR AUTOMOTIVE VEHICLES

Thomas D. Kleino, assignor to L&L Products, Incorporated
21 October 2003 (Class 296/146.6); filed 16 May 2001

An expandable material, such as polymer-based foam, is placed into automotive closure panels (such as doors) prior to final vehicle assembly. As the vehicle undergoes final assembly, preferably during a paint operation, the material is activated to expand, bond, and fill the panel assembly.—EEU

6,628,457

43.40.Vn ANTIVIBRATION MICROSCOPE

Eiichi Ito, assignor to Asahi Kogaku Kogyo Kabushiki Kaisha
30 September 2003 (Class 359/368); filed in Japan 11 July 2000

Microscopes supported at the end of booms or pedestals, as for use in surgery, are provided with acceleration and rotation sensors. Signals from these sensors are used to move optical elements so as to steady the image seen by a viewer or displayed on a screen—EEU

6,631,641

43.40.Yq DEVICE AND METHOD FOR DETERMINING FREQUENCY AND AMPLITUDE OF AN OSCILLATING STRUCTURE, ESPECIALLY FOR MEASURING ACCELERATION OR ROTATIONAL RATES

Josef Schalk *et al.*, assignors to Eads Deutschland GmbH
14 October 2003 (Class 73/504.03); filed in Germany 8 June 1999

As indicated in the patent summary, a device according to this patent consists of a movable element that can be excited in vibration, a pair of position sensors to determine the deflection of the movable element (arranged so that during a half-wave of the vibration their signals exceed or fall short of each other), a circuit for comparing the signals from these sensors (for determining a threshold value of the half-wave of the vibration), and a device for determining the duration during which the signal from one of the two sensors exceeds or falls short of the threshold value.—EEU

6,633,822

43.40.Yq VIBRATION DATA PROCESSOR AND PROCESSING METHOD

Philip L. Maness and Johannes I. Boerhout, assignors to SKF Condition Monitoring
14 October 2003 (Class 702/56); filed 23 July 2001

Monitoring of rotating machinery is accomplished by feeding the output of a sensor via a fixed analog filter to an analog/digital converter, whose output is fed to a digital signal processor and from there to an event detector. In one embodiment a mask generator, which may be under user control, is coupled to the filter.—EEU

6,626,264

43.50.Gf RADIO FREQUENCY SHIELDED AND ACOUSTICALLY INSULATED ENCLOSURE

Walter J. Christen, assignor to IGT
30 September 2003 (Class 181/290); filed 30 October 2001

This enclosure for magnetic resonance imaging (MRI) equipment is intended to address both airborne and structure-borne noise transmission. The enclosure in essence consists of walls and a roof that are constructed from layers of sound isolating and rf shielding materials and of a floor that consists of an elastically supported nonmagnetic steel plate.—EEU

6,640,926

43.50.Gf ELBOW SILENCER

Jonathan Weinstein, assignor to Industrial Acoustics Company, Incorporated
4 November 2003 (Class 181/224); filed 29 December 2000

This elbow silencer carries the dual purpose of directing HVAC flow and providing silencing of the flow. It consists of a casing enclosing a set of baffles that provide a number of alternate flow channels thereby producing predictable noise reduction.—DRR

6,648,295

43.50.Gf VIBRATION AND SOUND DAMPENER FOR HEAVY MACHINERY

Andrew James Herren and Joel Barron, both of Anaheim, California
18 November 2003 (Class 248/636); filed 2 October 2001

The purpose of this device is to damp vibrations and sound generated by heavy machinery such as metal presses, metal stampers, and HVAC units. The device consists of a basically rectangular upper platform that supports the base of the machinery to be damped. A lower platform mirrors the shape of the upper platform. Multiple sound and vibration dampers are attached between these two platforms, each consisting of a male cylinder sliding within a female cylinder. The latter cylinder incorporates a fairly frictionless material on the inner wall to allow the male cylinder to slide freely in a vertical motion. An air spring is formed in the chamber created by the two cylinders, thereby acting as a damper between the upper and lower platforms. The patent argues that the dampers also constrain lateral movement of the machinery.—DRR

6,641,864

43.55.Ev MORE CONTROLLABLE ACOUSTIC SPRAY PATCH COMPOSITIONS

John R. Woods, assignor to Spraytex, Incorporated
4 November 2003 (Class 427/140); filed 6 December 2000

The patent describes a treatment of textured material that can be applied from a spray can to patch an acoustic ceiling. The material comprises a base, a filler, an adhesive binder, an anti-foaming agent, a suspension agent, and other stuff. The material can be sprayed on a prepared patch to blend in with the typical bumpy, pebbled finish of the surrounding acoustic ceiling.—CJR

6,640,507

43.55.Ti ACOUSTIC BUILDING STRUCTURE

Alain Leconte, assignor to Saint–Gobain Isover
4 November 2003 (Class 52/145); filed in France 23 September 1999

Metal brackets are applied to a wall or its framing so that panels of mineral fiber or wool can be held in place by the flanges of the brackets. Plasterboard can then be attached to the surface of the assembly.—CJR

6,644,435

43.55.Ti COMPOSITE SOUND INSULATION SYSTEM FOR ROOM BOUNDARY SURFACES

Manfred Elsasser, Patsch, Austria
11 November 2003 (Class 181/207); filed in the European Patent Office 29 December 1999

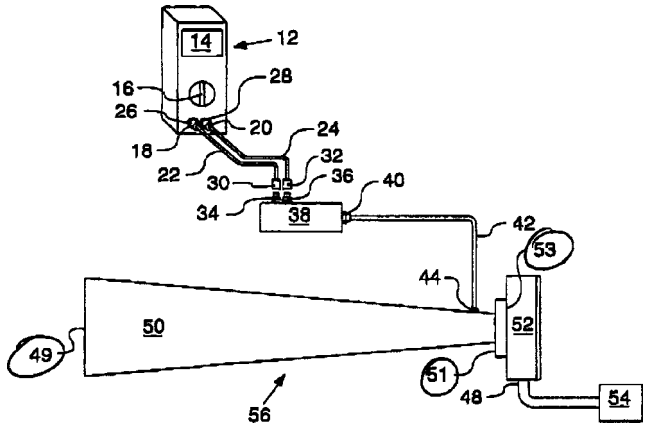
Separate air-blister sheets are combined with a thin sound-proofing interlayer to make a system for use as a floor, wall, or ceiling panel. By designating the inner loss factor and dynamic stiffness of the layers, the composite double-shell system enhances foot-fall sound insulation and airborne sound transmission loss and is thinner overall than other similar systems.—CJR

6,628,568

43.58.Vb SYSTEM AND METHOD FOR VERIFICATION OF ACOUSTIC HORN PERFORMANCE

Michael M. Mahler *et al.*, assignors to BHA Group Holdings, Incorporated
30 September 2003 (Class 367/13); filed 19 March 2002

This invention deals with a specialized kind of acoustic horn—a gas operated device that produces high-intensity energy in the 60 to 300-Hz range and is used in industrial cleaning applications. This patent suggests that the performance of such a device can be periodically verified by measuring relative sound pressure near the throat of the horn. Sound energy



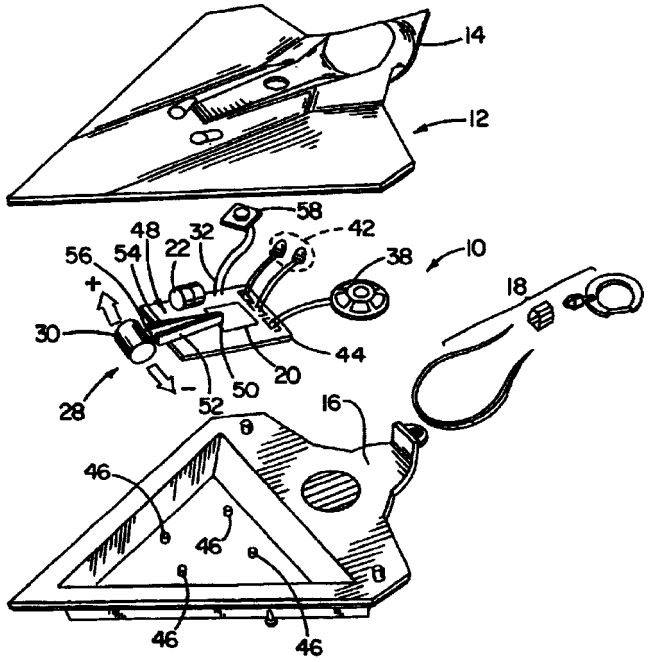
produced by one or more pulses is conducted from bore hole **44** through conduit **42** to pressure transducer **38** located in a more benign environment. The resulting electrical signal can be monitored by a handheld RMS voltmeter **12**.—GLA

6,623,326

43.58.Wc SOUND-EFFECTS GENERATING DEVICE WITH BIPOLAR MAGNETIC SWITCHING FOR ACTIVITY DEVICES

Daniel Judkins, assignor to Hasbro, Incorporated
23 September 2003 (Class 446/175); filed 19 December 2001

As shown, a circuit board with magnetic sensors **52** and **54** can detect another toy of the same or similar design with magnet **30**. The detection of



this proximity of attack means that the loudspeaker **38** can sound the alarm. And you can attach it to key ring **18** too.—MK

6,649,621

43.64.Gz PREVENTION OR REVERSAL OF SENSORINEURAL HEARING LOSS (SNHL) THROUGH BIOLOGIC MECHANISMS

Richard D. Kopke *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
18 November 2003 (Class 514/266.1); filed 23 January 2001

This is a pharmaceutical approach to preventing and/or reversing sensorineural or toxin-induced hearing loss. Agents are used that augment inner ear antioxidant defenses such as adenosine agonisis or up-regulating agents and/or agents which increase inner ear glutathione levels to prevent or reverse hearing loss induced by noise or toxin. Also covered in the patent are agents that curtail programmed cell death pathways and induce or enhance cell repair mechanisms in the inner ear.—DRR

6,639,987

43.64.Jb COMMUNICATION DEVICE WITH ACTIVE EQUALIZATION AND METHOD THEREFOR

Jason Darrell McIntosh, assignor to Motorola, Incorporated
28 October 2003 (Class 381/71.6); filed 11 December 2001

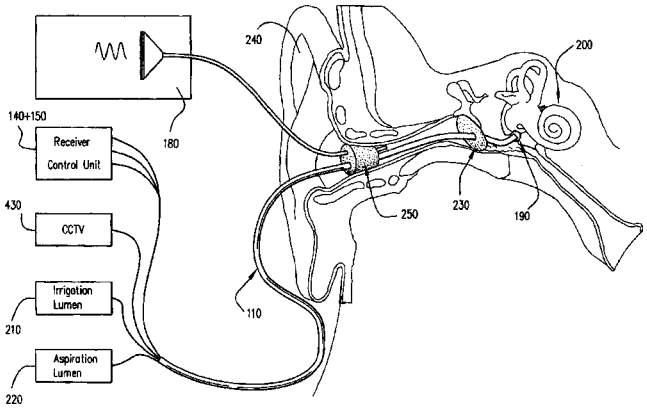
A drawback of the traditional cell phone is that the acoustic impedance varies with the position of the receiver with respect to the user's ear, affecting the frequency content of acoustic signals reaching the ear. A means for equalizing the acoustic input signal so as to counter this effect is provided by a sensor located near the earpiece. The sensor measures the acoustic pressure and generates a response signal. An active equalization circuit is coupled to the sensor and equalizes the acoustic input signal in reaction to the response signal.—DRR

6,640,121

43.66.Sr OTIC MICROPROBE FOR NEURO-COCHLEAR MONITORING

Fred Telischi *et al.*, assignors to The University of Miami
28 October 2003 (Class 600/379); filed 10 August 2000

This otic microprobe is said to be useful for assessing auditory function by enabling clinical and interpretive measurements of blood flow, particularly cochlear blood flow and neural compound action potentials, more specifically in the cochlea and vestibulocochlear nerve (i.e., the eighth cranial nerve). The microprobe consists of a fiberoptic laser Doppler flowmetry



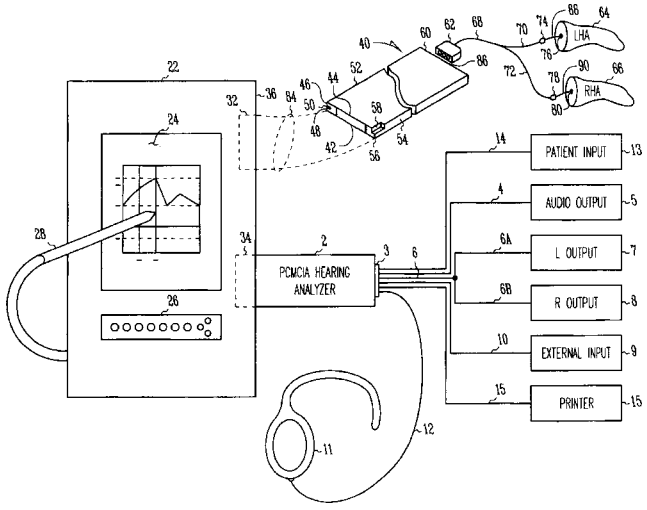
probe that measures blood flow and velocity, an electrocochleography electrode that measures neural compound action potentials, and a cap covering the tip of the probe. The microprobe features at least one irrigation lumen, at least one aspiration lumen, and a means for conducting an electrocochlear signal from the probe tip to the data monitoring unit.—DRR

6,647,345

43.66.Sr PORTABLE HEARING-RELATED ANALYSIS SYSTEM

Gordon J. Bye *et al.*, assignors to Micro Ear Technology, Incorporated
11 November 2003 (Class 702/57); filed 29 March 2002

We have here a portable hearing analyzer in the form of a PCMCIA card operating in a host computer. A microprocessor on the card executes the software for hearing analyses. A hearing aid interface for adjusting voltage levels and impedance levels is adapted for coupling signals to the hearing



aid being programmed. Two versions of systems for executing hearing-related analyses include a portable audiometer system to analyze a patient's hearing and a real-ear system to analyze output from a hearing aid in a patient's ear.—DRR

6,647,368

43.66.Sr SENSOR PAIR FOR DETECTING CHANGES WITHIN A HUMAN EAR AND PRODUCING A SIGNAL CORRESPONDING TO THOUGHT, MOVEMENT, BIOLOGICAL FUNCTION AND/OR SPEECH

Guerman G. Nemirovski, assignor to Think–A–Move, Limited
11 November 2003 (Class 704/270); filed 2 July 2001

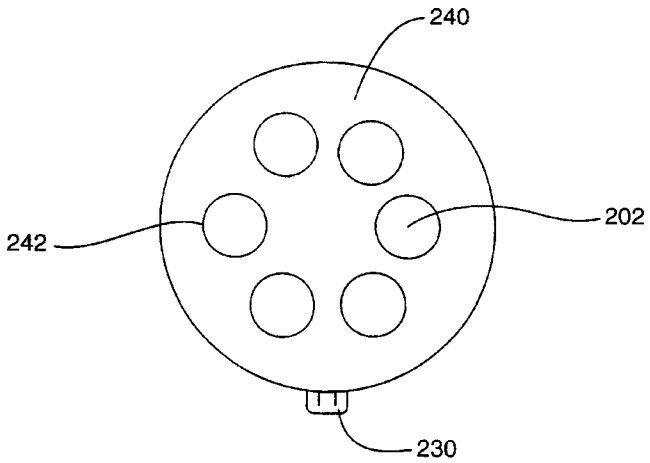
This sensor system is reportedly able to detect a person's physical or mental actions such as speech, thoughts, movements of the tongue with respect to the oral cavity, biological functions, etc. The basic idea is that it detects an air pressure change inside the ear. A miniature microphone placed at least partially within the ear detects changes in air pressure corresponding to an initiating action made by the person. An electrical signal is produced to yield an output corresponding to the pressure changes.—DRR

6,626,822

43.66.Ts IMPLANTABLE MICROPHONE HAVING IMPROVED SENSITIVITY AND FREQUENCY RESPONSE

Eric M. Jaeger *et al.*, assignors to Symphonix Devices, Incorporated
30 September 2003 (Class 600/25); filed 12 July 2000

An implantable microphone is implemented within the middle ear with either a linear-variable-differential transformer (LVDT) or a movable-plate differential capacitor (MPDC). A movable magnetic core of a LVDT or a movable plate of a MPDC attached to a moving component of the middle



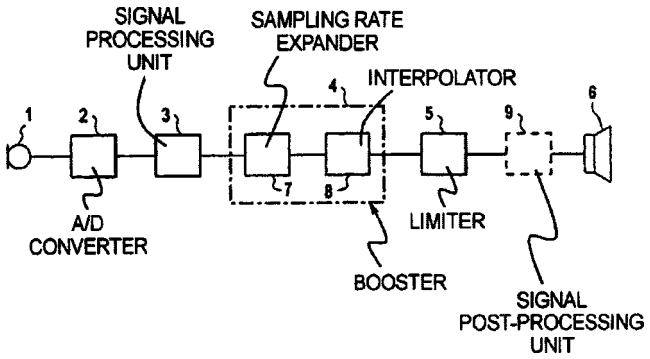
ear, e.g., the ossicles, measures relative motion and induces a modulating signal in the transformer windings or fixed plates, respectively. The resulting phase or voltage changes are detected as a representation of the audio signal incident at the eardrum.—DAP

6,628,794

43.66.Ts METHOD AND APPARATUS FOR LEVEL LIMITATION IN A DIGITAL HEARING AID

Thomas Dickel *et al.*, assignors to Siemens Audiologische Technik GmbH
30 September 2003 (Class 381/316); filed in Germany 26 November 1999

Signals in a digital hearing aid are normally limited to a maximum value to prevent intense sounds from exceeding the threshold of discomfort for a wearer. To prevent the spectra of the limited signal and its periodic continuations from overlappling, prior to limiting the maximum signal



value, the sampling rate of the signal is increased following A/D conversion, resulting in greater separation between the limited signal spectral components and the fundamental spectrum of the signal.—DAP

6,628,795

43.66.Ts DYNAMIC AUTOMATIC GAIN CONTROL IN A HEARING AID

Carl Ludvigsen, assignor to Widex A/S
30 September 2003 (Class 381/321); filed 16 October 2000

Response times of an AGC function in a hearing aid are designed to provide relatively fast gain adjustment with short attack and release times at high input and/or output sound levels and relatively slow adjustment with long attack and release times at low input and/or output sound levels. Advantages are to prevent clipping and exceeding the wearer's pain threshold at high sound levels and to reduce audible pumping artifacts at lower sound levels.—DAP

6,633,645

43.66.Ts AUTOMATIC TELEPHONE SWITCH FOR HEARING AID

Mark A. Bren and Timothy S. Peterson, assignors to Micro Ear Technology, Incorporated
14 October 2003 (Class 381/1); filed 7 August 2002

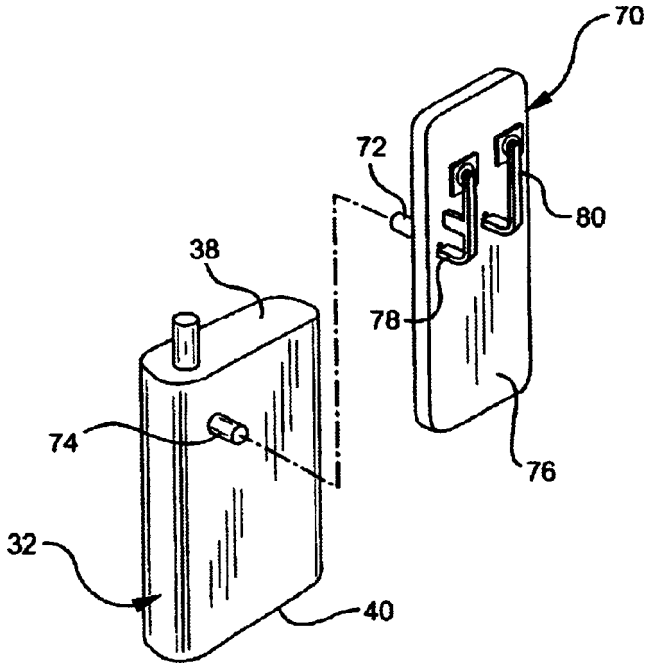
A switch in a hearing aid utilizes the presence of a magnetic field from a telephone handset to automatically switch between acoustic input via microphone and inductive input via an induction coil in the hearing aid. The switch function may be implemented with a reed switch.—DAP

6,639,564

43.66.Ts DEVICE AND METHOD OF USE FOR REDUCING HEARING AID RF INTERFERENCE

Gregory F. Johnson, Aptos, California
28 October 2003 (Class 343/702); filed 30 September 2002

An apparatus that forms a directional antenna may be incorporated within a personal wireless device (PWD) during manufacture or as an aftermarket accessory on PWDs having an external antenna port. The antenna



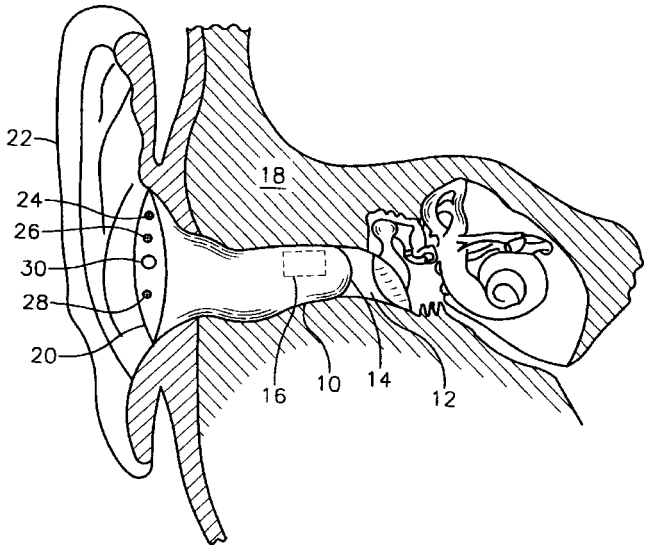
radiation pattern is directed away from the body which reduces rf interference into hearing aids as well as radiation into the wearer's body.—DAP

6,643,378

43.66.Ts BONE CONDUCTION HEARING AID

Daniel R. Schumaier, Johnson City, Tennessee
4 November 2003 (Class 381/326); filed 2 March 2001

A vibratory transducer **16** is positioned in the ear canal adjacent to the mastoid bone **18**. A microphone signal is amplified and sent to the vibrator,



which transfers the signal to the opposite cochlea via the mastoid bone.—DAP

6,643,373

43.71.Bp SPEAKER VOLUME INDICATOR FOR TELEPHONE HANDSETS

Robert D. Hughes, Andover, Massachusetts
4 November 2003 (Class 379/433.04); filed 6 October 2000

The aim of this loudness indicator, which can be incorporated into a telephone handset, is to assist hard-of-hearing persons speaking into a telephone receiver to keep their voices down to a moderate level. A microphone detects the sound pressure level of the user's speech and a series of colored lights indicate the speech loudness, e.g., green for ''normal,'' yellow for ''somewhat above normal,'' and red for ''really excessive loudness.''—DRR

6,511,324

43.71.Ft PHONOLOGICAL AWARENESS, PHONOLOGICAL PROCESSING, AND READING SKILL TRAINING SYSTEM AND METHOD

Janet M. Wasowicz, assignor to Cognitive Concepts, Incorporated
28 January 2003 (Class 434/167); filed 6 October 1999

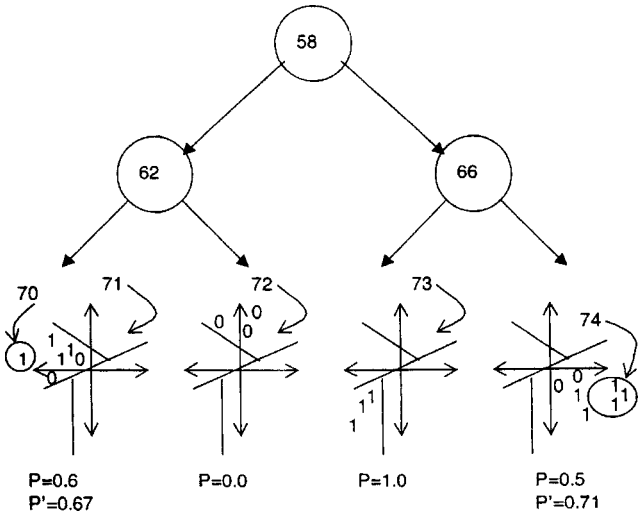
This computer-based, interactive phonetic environment provides a set of games which expose the child to various phonetic sequences, accompanied by suitable pictures and story line. A speech recognizer provides a limited degree of phonetic analysis, allowing the child to respond by speaking one of a limited range of sounds. The games are arranged to teach phonetic distinctions based on formant transitions, durations, and other like phonetic characteristics.—DLR

6,519,561

43.72.Fx MODEL ADAPTATION OF NEURAL TREE NETWORKS AND OTHER FUSED MODELS FOR SPEAKER VERIFICATION

Kevin Farrell and William Mistretta, assignors to T–Netix, Incorporated
11 February 2003 (Class 704/232); filed 3 November 1998

The patent describes a method by which the speaker models for any of several types of speaker verification systems can be dynamically updated to reflect the voice characteristics of the latest verified speech sample. The method is said to be applicable to neural network trees, Gaussian mixture



models, dynamic time warping, or combination systems using any of these methods. What is described is a straightforward technique of updating the weights or occurrence counts in the model once a given point in the feature space has been located. The figure shows how the node weights would be altered in a neural network tree system.—DLR

6,600,740

43.72.Gy VOICE QUALITY OPTIMIZATION ON MULTI-CODEC CALLS

Eric Valentine and Walter Lee Davidson, assignors to Ericsson Incorporated
29 July 2003 (Class 370/365); filed 3 October 1998

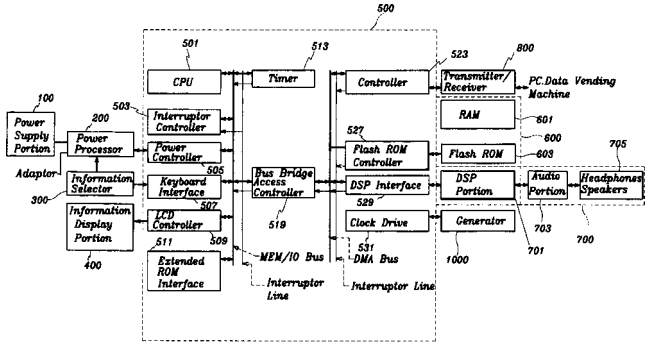
A communication network is proposed with signals indicating the encoding technique employed at the sending end. This is to enable the receiving end to optimize its decoding algorithm based on the encoding algorithm.—RSM

6,629,000

43.72.Gy MPEG PORTABLE SOUND REPRODUCING SYSTEM AND A REPRODUCING METHOD THEREOF

Kwang-su Moon and Jung-ha Hwang, assignors to MPMan.com Incorporated
30 September 2003 (Class 700/94); filed in the Republic of Korea 24 November 1997

A methodology is outlined to control all operations of downloading, converting, reproducing, and storing sound data compressed with the MPEG



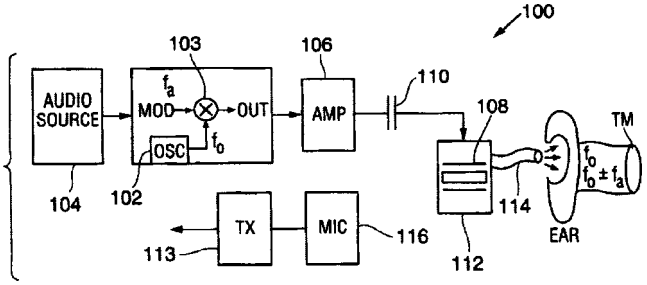
method in a portable configuration. Means are also provided to transmit and receive sound and program data from external devices.—DAP

6,631,196

43.72.Gy METHOD AND DEVICE FOR USING AN ULTRASONIC CARRIER TO PROVIDE WIDE AUDIO BANDWIDTH TRANSDUCTION

Jon C. Taenzer and Robert J. Higgins, Jr., assignors to GN ReSound North America Corporation
7 October 2003 (Class 381/316); filed 7 April 2000

An inaudible ultrasonic signal is transmitted into the ear canal. The nonlinearities of the middle and/or inner ear demodulate the audio via an



ultrasound-to-audio sound conversion in a confined volume. Ultrasound difference frequency pressure is converted into audible pressure, producing a constant pressure sound intensity across the entire audio bandwidth.—DAP

6,647,369

43.72.Gy READER TO DECODE SOUND AND PLAY SOUND ENCODED IN INFRA-RED INK ON PHOTOGRAPHS

Kia Silverbrook *et al.*, assignors to Silverbrook Research Pty Limited
11 November 2003 (Class 704/273); filed 20 October 2000

A camera system has means for recording sound associated with a photograph. The prerecorded audio is encoded on the photograph using an array of printed infrared ink dots. Reed–Solomon encoding may be used to provide error correction. After decoding, the audio is stored in memory for playback via a sound processing chip and loudspeaker.—DAP

6,502,074

43.72.Ja SYNTHESISING SPEECH BY CONVERTING PHONEMES TO DIGITAL WAVEFORMS

Andrew Paul Breen, assignor to British Telecommunications public limited company
31 December 2002 (Class 704/260); filed in the European Patent Office 4 August 1993

This system for speech synthesis by concatenation of phonetic segment waveforms looks up strings of up to five phonetic symbols in a phonetic network. Each node in the symbol network is linked to a waveform fragment. Waveform fragments are selected according to the longest context available, using a feature vector system to choose between available candidate sequences. The patent describes procedures for building the symbol/waveform datebase from an extended speech sample as well as the search methods for locating appropriate waveform segments during synthesis.—DLR

6,505,158

43.72.Ja SYNTHESIS-BASED PRE-SELECTION OF SUITABLE UNITS FOR CONCATENATIVE SPEECH

Alistair D. Conkie, assignor to AT&T Corporation
7 January 2003 (Class 704/260); filed 5 July 2000

This system for speech synthesis by concatenation of phonetic segment waveforms includes prosodic and spectral information in a database of triphone sequences. During synthesis, the phonetic string to be synthesized includes prosodic targets, which are matched against the available database candidates. A novel aspect of the system is that the waveform segment database is compiled by creating millions of phonetic sequences using a high-quality synthesis algorithm. This allows the construction of an extensive triphone inventory with each triphone represented with a variety of prosodic features. The result is a large-memory synthesizer which requires minimal processor time during execution.—DLR

6,510,413

43.72.Ja DISTRIBUTED SYNTHETIC SPEECH GENERATION

Mark R. Walker, assignor to Intel Corporation
21 January 2003 (Class 704/258); filed 29 June 2000

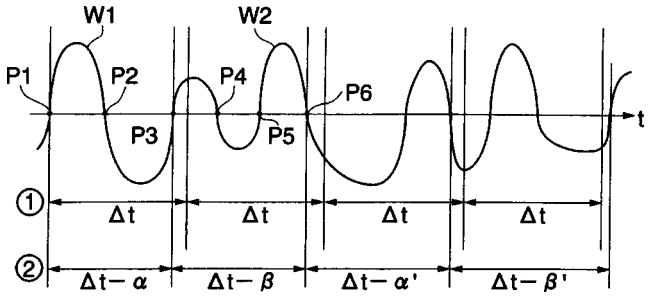
This network-based speech synthesis system is intended to allow optimization of network data transmission loads with minimal computational requirements at the client's end of the connection. Facilities are provided for converting portions of a web page into annotated phonetic text. The text may be converted to audio waveforms at a server site or a client site, depending on available processing capabilities. Various alternate forms of synthesis are discussed.—DLR

6,513,007

43.72.Ja GENERATING SYNTHESIZED VOICE AND INSTRUMENTAL SOUND

Akio Takahashi, assignor to Yamaha Corporation
28 January 2003 (Class 704/258); filed in Japan 5 August 1999

This is an alternative implementation of the old channel vocoder technique by which a human voice and a musical instrument sound are combined to produce a talking instrument. Disadvantages are cited for known methods of channel vocoding, filter-bank analysis and reconstruction, terminal analog synthesis, linear prediction coding, etc. A method referred to as



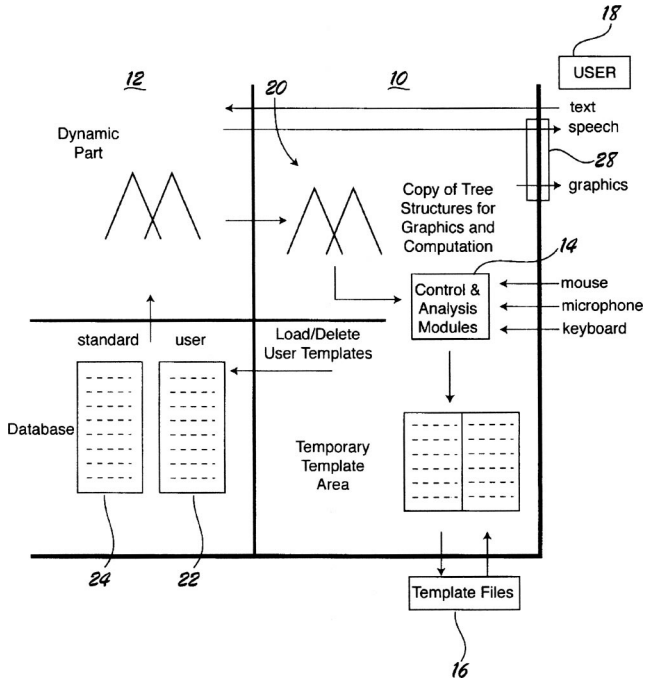
"dynamic cutting" extracts coefficients from the speech input signal. The coefficients then drive a FIR-based, time-domain convolution of the second input signal to synthesize the output signal. Details of the dynamic cutting process seem a bit sketchy, but a reference to an earlier Japanese patent is provided.—DLR

6,513,008

43.72.Ja METHOD AND TOOL FOR CUSTOMIZATION OF SPEECH SYNTHESIZER DATABASES USING HIERARCHICAL GENERALIZED SPEECH TEMPLATES

Steve Pearson *et al.*, assignors to Matsushita Electric Industrial Company, Limited
28 January 2003 (Class 704/260); filed 15 March 2001

This speech synthesis system is intended for generating spoken directions in an automobile navigation system. The system described is said to be more flexible than prior synthesizers in having dynamic, hierarchical data structures which allow modification of prosodic and other features late in the



synthesis cycle. The lexical database includes sets of templates which are used to transfer phonetic, prosodic, and semantic elements into the active synthesis structure. The method is said to be suitable for additional applications, including foreign language teaching and general Internet content synthesis.—DLR

6,516,207

43.72.Ja METHOD AND APPARATUS FOR PERFORMING TEXT TO SPEECH SYNTHESIS

Vishwa N. Gupta and Paul Boucher, assignors to Nortel Networks Limited
4 February 2003 (Class 455/563); filed 7 December 1999

This speech synthesizer, intended for use in a cell phone or similar environment, allows the remote unit to specify speaker characteristics, such as vocal tract length or pitch range, which are sent from the remote unit to the base station using a low-rate data channel. Text from a distant terminal is then synthesized at the base station using the speaker information and transmitted as voice to the remote unit. Alternative embodiments would support a synthesis capability in the remote unit.—DLR

6,516,298

43.72.Ja SYSTEM AND METHOD FOR SYNTHESIZING MULTIPLEXED SPEECH AND TEXT AT A RECEIVING TERMINAL

Takahiro Kamai *et al.*, assignors to Matsushita Electric Industrial Company, Limited
4 February 2003 (Class 704/260); filed in Japan 16 April 1999

This speech synthesizer, for use in a receiving terminal such as a cell phone, provides a variety of capabilities centered around a CELP synthesizer. The synthesizer is based on line spectrum pair coefficients with an adaptive code book source. A variety of parameter and transmission schemes split up the synthesis parameters in various ways between a data channel and a text channel, allowing generation of prosodic information and conversion of text to phonetic streams at various alternate locations.—DLR

6,606,382

43.72.Kb SYSTEM AND METHOD FOR IMPLEMENTATION OF AN ECHO CANCELLER

Samir K. Gupta, assignor to Qualcomm Incorporated
12 August 2003 (Class 379/406.05); filed 27 January 2000

Echo cancellers possess residual echo cancellers that attentuate only when residual echo is present. This can result in abrupt changes in the attenuation applied to an output signal as the state changes from echo present to near-end speech present or double talk. This patent proposes that these transitions be smoothed in order to reduce annoyance for the users and load on the vocoder.—RSM

6,625,279

43.72.Kb APPARATUS AND METHOD OF ESTIMATING ECHO PATH DELAY

Kwang-Seop Eom, assignor to Samsung Electronics Company, Limited
23 September 2003 (Class 379/406.01); filed in the Republic of Korea 19 July 1999

A method is proposed for reducing the amount of computation for echo cancellation by estimating time delay to the echo-causing impedance discontinuity. An adaptive filter is used to estimate the delay from transmitted and received low-passed and decimated signals.—RSM

6,628,781

43.72.Kb METHODS AND APPARATUS FOR IMPROVED SUB-BAND ADAPTIVE FILTERING IN ECHO CANCELLATION SYSTEMS

Patrik Grundström and Fredrik Ehrenstråle, assignors to Telefonaktiebolaget LM Ericsson (publ)
30 September 2003 (Class 379/406.14); filed 3 June 1999

This patent proposes that when using subband adaptive filtering, neighboring bands overlap at the frequency values for which they are 3 dB down from the all-band response. Aliasing will be mitigated with such a method.—RSM

6,636,604

43.72.Kb METHOD AND DEVICE FOR SUPPRESSING ECHO IN A HANDS FREE DEVICE SUCH AS A TELEPHONE

Peter Taege, assignor to Deutsche Telekom AG
21 October 2003 (Class 379/406.01); filed in Germany 1 December 1997

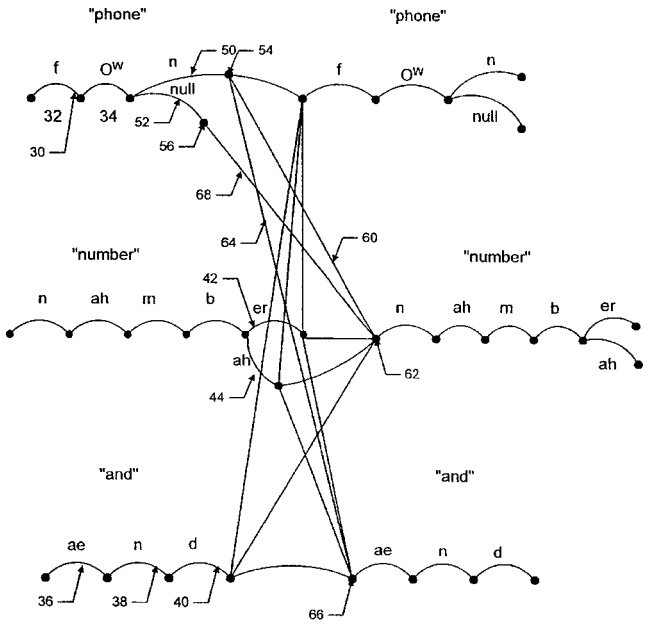
Residual echo suppression is attained by controlling the gain of amplifiers in the receiving and sending paths. The control signals are the sum of the far- and near-signal powers and the ratio of the near-signal power to the far-signal power. The change of state in amplifier gains is produced using exponential changes in time for a natural sounding signal.—RSM

6,501,833

43.72.Ne METHOD AND APPARATUS FOR DYNAMIC ADAPTATION OF A LARGE VOCABULARY SPEECH RECOGNITION SYSTEM AND FOR USE OF CONSTRAINTS FROM A DATABASE IN A LARGE VOCABULARY SPEECH RECOGNITION SYSTEM

Michael S. Phillips and John N. Nguyen, assignors to Speech Works International, Incorporated
31 December 2002 (Class 379/88.07); filed 3 October 1997

A highly structured lexical network is intended for use in a large-vocabulary speech recognition system. Three types of nodes, phonetic constraints, word classes, and word connections, are combined to form subnetwork fragments, which are then precompiled and cross-linked to other subnetworks. The overall network consists of three levels of words distinguished by usage frequency, with the most active subnetworks comprising up to perhaps 1000 most commonly used words. A few hundred thousand



words may be stored in the form of phonetic baseforms, which may be rapidly compiled into active subnetworks. Millions of words may be available in the form of phonetic spellings. New vocabulary may be added to the networks during usage by first storing the new item in spelled form, then promoting it into the compiled forms as usage indicates. The patent provides considerable detail on the organization of the various nodes and networks.—DLR

6,505,154

43.72.Ne METHOD AND DEVICE FOR COMPARING ACOUSTIC INPUT SIGNALS FED INTO AN INPUT DEVICE WITH ACOUSTIC REFERENCE SIGNALS STORED IN A MEMORY

Hermann Bottenbruch and Michael Mertens, assignors to Primasoft GmbH
7 January 2003 (Class 704/236); filed in Germany 13 February 1999

This relatively simple acoustic recognizer is intended for use in control applications, such as appliance control or automotive device control. A simple or complex tone or a vowel sound is analyzed as a Fourier spectrum and various spectral features, such as highest peak frequency, envelope amplitude, and pitch, are extracted. These features are used to search a reference vector space. There appears to be some mechanism for accepting sequential feature vector frames, however, most of the discussion relates to classification of individual frames.—DLR

6,507,815

43.72.Ne SPEECH RECOGNITION APPARATUS AND METHOD

Hiroki Yamamoto, assignor to Canon Kabushiki Kaisha
14 January 2003 (Class 704/231); filed in Japan 2 April 1999

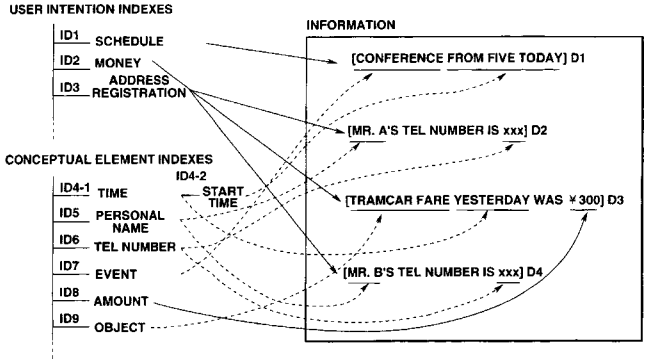
This large-vocabulary speech recognizer uses a well-known phonetic tree structure for storing the lexical database. Each node contains a list of words having a common initial phoneme sequence and branching according to the next phoneme. At each node, a language model is consulted to determine the likelihood of each possible pathway through that node.—DLR

6,510,412

43.72.Ne METHOD AND APPARATUS FOR INFORMATION PROCESSING, AND MEDIUM FOR PROVISION OF INFORMATION

Takashi Sasai *et al.*, assignors to Sony Corporation
21 January 2003 (Class 704/257); filed in Japan 2 June 1998

This personal information management system uses speech recognition to identify certain keywords in the input speech and uses these to determine certain conceptual elements from which the user's intentions are



determined. Given a fairly limited range of expectations, the assumption is that memos pertaining to schedules, ticket purchases, etc., can be appropriately classified.—DLR

6,513,006

43.72.Ne AUTOMATIC CONTROL OF HOUSEHOLD ACTIVITY USING SPEECH RECOGNITION AND NATURAL LANGUAGE

John Howard and Jean-Claude Junqua, assignors to Matsushita Electronic Industrial Company, Limited
28 January 2003 (Class 704/257); filed 6 June 2001

This natural-language speech recognizer provides a natural user interface to the operating manual and activity guide for a complex home electronic services system. A version of the operating manual is provided in the form of a semantic representation which is accessible by the natural-language parser activated during the analysis of spoken inputs. The system allows easy access to control home automation or entertainment systems. A dialog manager and synthesis system provide spoken feedback to the user. The parser appears to be based on a slot grammar, although few details are provided. Recognition and synthesis system techniques are not described.—DLR

6,519,562

43.72.Ne DYNAMIC SEMANTIC CONTROL OF A SPEECH RECOGNITION SYSTEM

Michael S. Phillips *et al.*, assignors to Speechworks International, Incorporated
11 February 2003 (Class 704/240); filed 25 February 1999

This speech recognizer works in cooperation with a set of semantic rules provided by the application to improve the recognition result. The initial output of the recognizer is an *n*-best list of candidate word strings, each marked with a probability value and a set of semantic element/value pairs for keywords in the string. The entire list is passed to the application, where a set of semantic rules is applied to the semantic elements and may modify the associated values based on the current situation, external conditions, etc. The entire list is then passed back to the recognizer, which adjusts the string probabilities and, if necessary, reorders the *n*-best list. The application then selects the most probable word string.—DLR

6,629,075

43.72.Ne LOAD-ADJUSTED SPEECH RECOGINTION

Johan Schalkwyk, assignor to SpeechWorks International, Incorporated
30 September 2003 (Class 704/270); filed 9 June 2000

Processor loading during speech recognition operations for interactive voice response systems is here determined in four categories depending on the number of callers at one time: idle, normal, busy, and pegged. The speech processor is configured for a particular usage limit associated with each loading category. Goals are to accommodate peak periods of speech recognition system use and improve accuracy of speech recognition during nonpeak periods.—DAP

6,627,802

43.75.De REINFORCING BRACES FOR STRINGED MUSICAL INSTRUMENTS AND METHOD FOR POSITIONING SAME

Grady Jones, Burton, Michigan
30 September 2003 (Class 84/290); filed 13 February 2002

Plate vibrations in stringed instruments exhibit moding. As the author remarks, these studies have been performed over several decades by a variety of researchers. The author proposes testing the plates for modes and using these results to place bracing where the best sound will result. Since the proof of the pudding is in the eating, final results would be advantageous but, alas, these are sadly lacking.—MK

6,627,808

43.75.Gh ACOUSTIC MODELING APPARATUS AND METHOD

Elon Ray Coats *et al.*, assignors to Peavey Electronics Corporation
30 September 2003 (Class 84/723); filed 3 September 2002

A solid body electric guitar lacks a body resonance. This patent teaches that a body vibration sensor located near the neck/body junction can be amplified and recombined with the string vibrations as shown in the

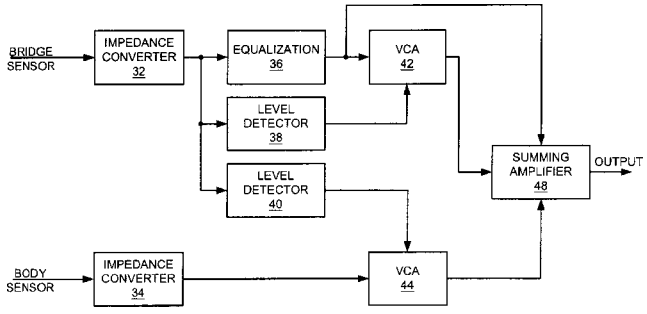


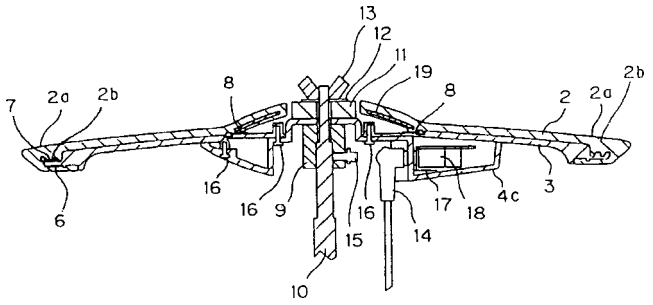
figure. Furthermore, it can all be done with analog circuitry as shown in schematics presented in the patent. However, Karjalainen and others have a DSP scheme that directly models the body resonance without tricks. Which would you choose?—MK

6,632,989

43.75.Kk ELECTRONIC PAD WITH VIBRATION ISOLATION FEATURES

Kiyoshi Yoshino and Masato Katsuda, assignors to Roland Corporation
14 October 2003 (Class 84/402); filed in Japan 22 August 2000

Cymbals can be used as sensors if some fundamental changes are made as shown in the figure. First, the cymbal surface is coated with rubber



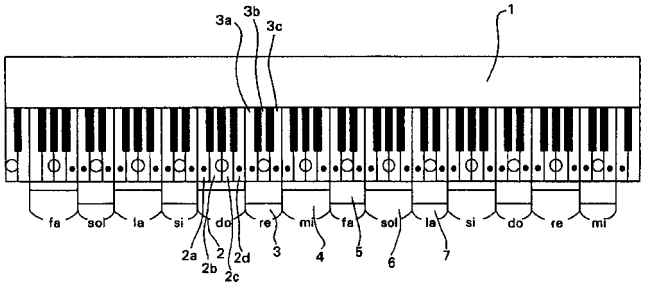
2 to reduce the hit sound. Second, piezoelectric sensors are added at the edge **6**, **7** and in the center **8**. They are connected to output cable **14**.—MK

6,624,348

43.75.Mn ELECTRONIC PIANO HAVING VARIABLE KEYS

Nobukado Matsutsuka, assignor to Kabushiki Kaisha Matsuya Sougou Kenkyusyo
23 September 2003 (Class 84/719); filed in Japan 27 June 1996

The basic idea is to divide the piano keyboard into zones, or ''big keys,'' **3**, **4**, **5**, etc. The abstract states that ''One can play the electronic piano in a sense of play. Besides, since each of the zones can be struck in a



lump, this electronic piano is suitable for a person having large hands and fingers or a person declined in dexterity such as an old person or a physically handicapped person.''—MK

6,629,067

43.75.St RANGE CONTROL SYSTEM

Tsutomu Saito *et al.*, assignors to Kabushiki Kaisha Kawai Gakki Seisakusho
30 September 2003 (Class 704/207); filed in Japan 15 May 1997

Karoke is the well-known entertainment of singing along with a soundtrack. Unfortunately, untrained singers lack the vocal range of the original performer. And so, the first idea is to pitch shift the untrained vocalist. If the computer knows the reference pitch of the original singer, then the pitch-shifted output can be closer to the original recording. In addition, it is well known that due to formant placement, pitch shifting will sound wrong unless the formants are moved into their expected places. Accordingly, the formants will be shifted into the right place. It's unfortunate that Kawai couldn't be bothered to find the many references to pitch shifting and left it all to the examiner.—MK

6,640,126

43.80.Qf ACOUSTIC GATING MONITOR FOR MAGNETIC RESONANCE IMAGING SYSTEM

Hsu Chang, assignor to Toshiba America MRI, Incorporated
28 October 2003 (Class 600/413); filed 26 February 2001

This is a system that monitors a patient undergoing magnetic resonance imaging (MRI) and triggers an acoustic signal in response to a physi-

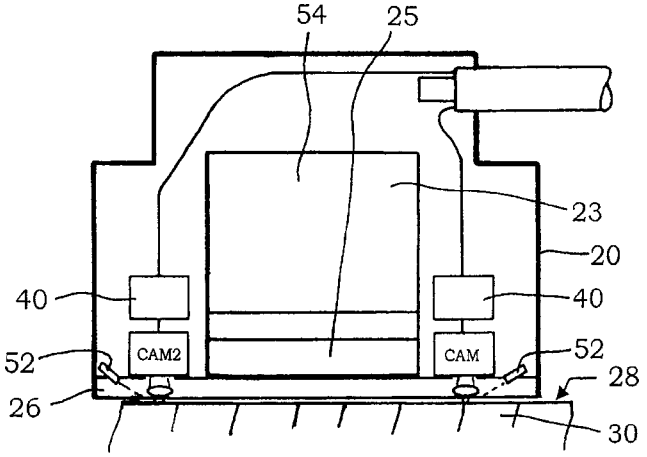
ological sensor on the patient. The acoustic feedback signal is generated by one of the MRI gradient coils and may be indicative of the patient's heartbeat or some other physiological event.—DRR

6,641,535

43.80.Qf ULTRASONIC PROBE, IN PARTICULAR FOR MANUAL INSPECTIONS

Paul Buschke *et al.*, assignors to Agfa NDT GmbH
4 November 2003 (Class 600/437); filed in Germany 22 November 2000

This ultrasonic probe, intended for manual use, has a housing that accommodates an ultrasonic crystal and has a contact surface for contacting the surface of a body to be inspected. At least one miniature digital camera



is aligned to view the surface of the body and supply an image of portions of the surface at intervals. Images taken at two different times are compared to determine the displacement of the housing relative to the surface through a displacement memory.—DRR

6,645,145

43.80.Qf DIAGNOSTIC MEDICAL ULTRASOUND SYSTEMS AND TRANSDUCERS UTILIZING MICRO-MECHANICAL COMPONENTS

William R. Dreschel *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
11 November 2003 (Class 600/443); filed 24 September 2001

Instead of electro-mechanical or electronic switches, micro-mechanical switches are used to control ultrasound systems and transducers. The patent asserts that use of these purely mechanical devices avoids current leakages, effects of capacitive coupling, electrostatic discharges, loss of signal integrity, etc. With the use of the newer technology of micro-mechanical devices, costs may be reduced and smaller, denser ultrasound transducer arrays developed.—DRR

6,645,148

43.80.Qf ULTRASONIC PROBE INCLUDING POINTING DEVICES FOR REMOTELY CONTROLLING FUNCTIONS OF AN ASSOCIATED IMAGING SYSTEM

An Nguyen-Dinh and Aimè Flesch, assignors to Vermon
11 November 2003 (Class 600/459); filed 20 March 2001

A pointing device similar to a computer mouse is mounted on an ultrasonic probe so that a user can remotely control the basic functions of the associated imaging system. The pointing device is effectively an extension of the controls provided by the keyboard so that the user, at his or her option, may access either the major image settings from the remote pointing device or all of the functions of the system using the keyboard.—DRR

6,645,149

43.80.Qf ULTRASOUND DEVICES

Leonard Smith, assignor to Deltex (Guernsey) Limited
11 November 2003 (Class 600/459); filed in the United Kingdom 13 April 1999

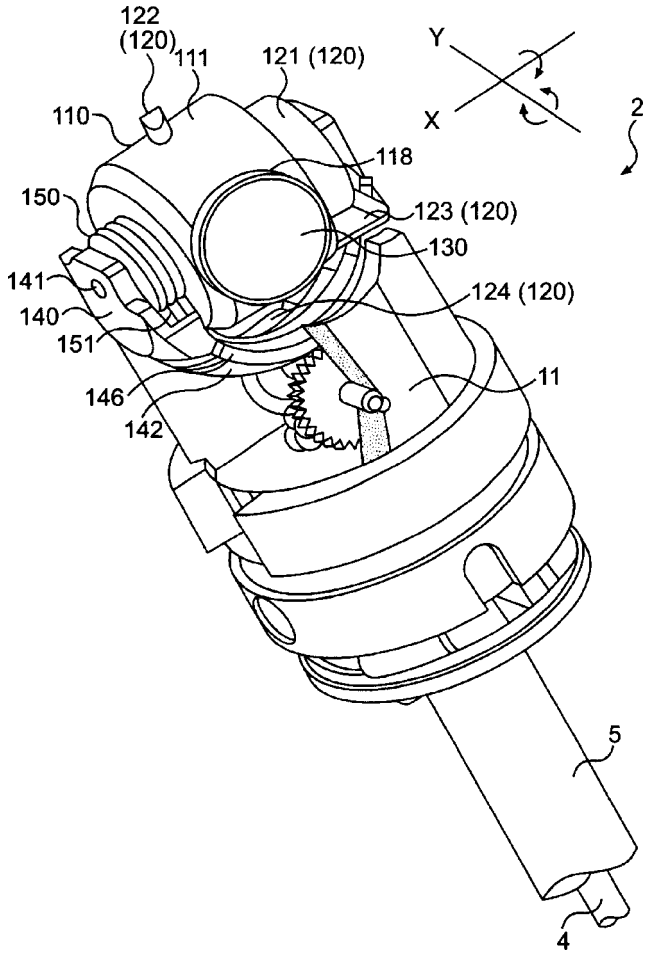
This ultrasound probe is intended for use as a haemodynamic monitor. It incorporates a host signal processor and an interconnect cable. The singular feature of the disposable probe is that it includes a memory device that communicates with the host processor to deliberately limit the probe's life, rendering it inoperable if an attempt is made to use it for more than one patient. Now that's designed-in obsolescence!—DRR

6,645,151

43.80.Qf ULTRASONIC PROBE

Kazuyoshi Irioka *et al.*, assignors to Matsushita Electric Industrial Company, Limited
11 November 2003 (Class 600/459); filed in Japan 26 November 1999

The object of this probe is to radiate an ultrasonic wave to noninvasively inspect the interior of a patient's body. The scan mechanism consists of rotating and swinging mechanisms. The rotating mechanism rotates a cylindrical holder carrying a piezoelectric element that emits an ultrasonic wave and receives the echo. The swinging mechanism swings a rotary base that supports the rotating mechanism, thus providing an additional degree of



freedom in allowing the cylindrical holder to swing about an axis of rotation extending perpendicular to the normal of the piezoelectric element.—DRR

6,650,940

43.80.Qf ACCELEROMETER-BASED HEART SOUND DETECTION FOR AUTOCAPTURE

Qingsheng Zhu *et al.*, assignors to Cardiac Pacemakers, Incorporated
18 November 2003 (Class 607/28); filed 2 February 2000

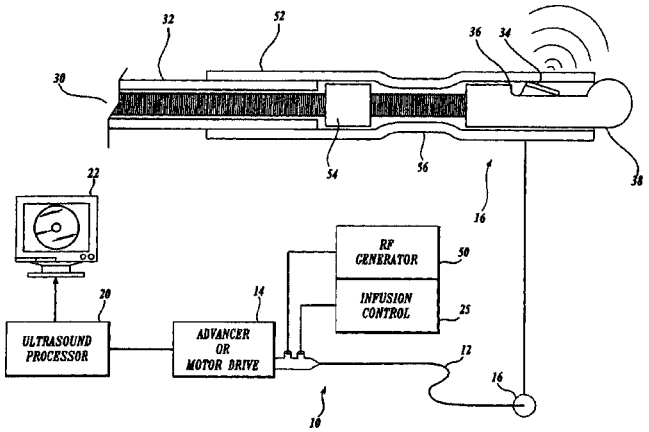
This implantable cardiac rhythm management device automatically detects intrinsic and evoked responses of a patient's heart. In the automatic capture verification mode, a global accelerometer signal is used to identify heart sounds S1 and S2. The presence or absence of one or more of these sounds indicates whether a stimulation pulse has evoked a heart response. The device may also automatically adjust the stimulation output by following a step-down stimulation protocol in which the presence of a predetermined heart sound indicates capture. If the patient's physical activity level exceeds a predetermined threshold, the device may also suspend the automatic capture verification sequence.—DRR

6,638,222

43.80.Sh RF ABLATION AND ULTRASOUND CATHETER FOR CROSSING CHRONIC TOTAL OCCLUSIONS

Chandru V. Chandrasekaran *et al.*, assignors to SciMed Life Systems, Incorporated
28 October 2003 (Class 600/439); filed 8 February 2002

This catheter combines an ultrasonic transducer and a rf ablation electrode. The ultrasound transducer sends signals to a blood vessel and receives returned echo signals. The echo signals are used to produce an image of the tissue surrounding the catheter. A driveshaft rotates the transducer to yield a



360-deg view of the target vessel wall. At the distal end of the driveshaft is an electrode that is coupled to a rf generator. The rf energy through the electrode ablates the occluding material in the vessel. A number of tip shapes for the electrode may be selected according to the configuration of the vessel and composition of the occluding material.—DRR

6,645,162

43.80.Sh SYSTEMS AND METHODS FOR ULTRASOUND ASSISTED LIPOLYSIS

Zvi Friedman *et al.*, assignors to Insightec–Txsonics Limited
11 November 2003 (Class 601/2); filed 11 June 2001

An external ultrasound transducer placed against a patient's skin is used to destroy subcutaneous cells. The transducer emits acoustic energy focused in a linear focus zone within that tissue region with sufficient intensity to rupture cells with a minimum of heating. The transducer may consist of one or more elements having a partial cylindrical shape, a single planar element coupled to an acoustic lens, or a number of linear transducer elements positioned in an arcuate or planar configuration. Detectors for sensing cavitation in the focal zone may be included. A frame can be provided to accommodate movement of the transducer along the patient's skin in response to the extent of cellular destruction caused by the transducer.—DRR

6,649,702

43.80.Sh STABILIZATION AND ACOUSTIC ACTIVATION OF POLYMERIC MICELLES FOR DRUG DELIVERY

Natalya Rapoport and William G. Pitt, assignors to University of Utah Research Foundation; Brigham Young University
18 November 2003 (Class 525/299); filed 7 August 2002

Disclosure is made of the methods for stabilizing micelles against degradation upon dilution and of activating micelles through ultrasound to release a medical substance. A micelle consists of molecules of a block polymer containing a hydrophobic block and a hydrophilic block. The hydrophobic block forms a core of the micelle with a corona formed of a hydrophilic block. The core is stabilized by (1) chemical cross linking, (2) incorporating hydrophobic oil (e.g., vegetable oil) in the core, and (3) incorporating a cross-linked interpenetrating network of a stimulus-responsive hydrogel into the core. The hydrogel is responsive to any stimulus, but preferably temperature or *p*H value. A substance such as a drug can be injected into the dense inner core of the micelles. When subjected to ultrasound, the micelles release the substance and then reversibly revert to a stable dense core and re-encapsulate the substance when the ultrasound is turned off. It is possible to control the release of the substance by pulsating the ultrasound.—DRR

6,650,935

43.80.Sh ULTRASOUND DRIVEN DEVICES FOR ACCELERATED TRANSFER OF SUBSTANCES ACROSS POROUS BOUNDARIES

David J. Watmough, Inverness, the United Kingdom
18 November 2003 (Class 604/20); filed in the United Kingdom 9 October 1998

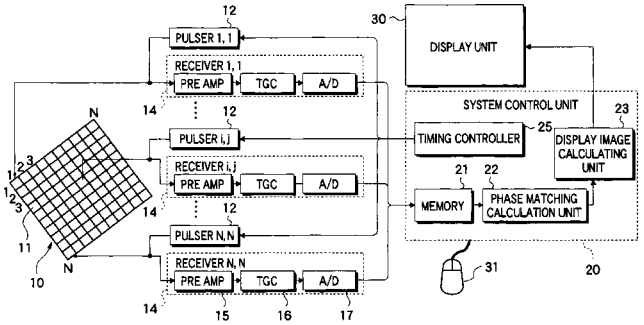
This is a method for the phonophoretic transfer of an active agent into or across a porous surface (such as the skin, eggshells, certain vegetables, and cellulosic surfaces). The active agent is put in contact with the porous surface and ultrasound in the frequency range of 20 kHz to 3 MHz is applied. The active agent is ideally cationic and is disposed in a low-viscosity carrier. The apparatus consists of an ultrasonic generator, a housing with a transducer mounted within, and a treatment chamber. The transducer is aimed toward the treatment chamber, which is equipped with a liquid-tight membrane and filled with the fluid carrying the active agent. The membrane retains the fluid in the chamber and is placed against the porous surface during sonification.—DRR

6,638,220

43.80.Vj ULTRASONIC IMAGING METHOD AND ULTRASONIC IMAGING APPARATUS

Tomoo Satoh, assignor to Fuji Photo Film Company, Limited
28 October 2003 (Class 600/437); filed in Japan 26 February 2001

Initial imaging is used to define range intervals of interest. In subsequent imaging, ultrasonic beams are emitted after receiving echoes from the maximum depth of interest to achieve an increased pulse repetition rate.



Beams can be emitted at angles that are not sequentially incremented to reduce interference from echoes arriving simultaneously from other directions or ranges.—RCW

6,638,221

43.80.Vj ULTRASOUND DIAGNOSTIC APPARATUS, AND IMAGE PROCESSING METHOD

Yasuhiko Abe and Ryoichi Kanda, assignors to Kabushiki Kaisha Toshiba
28 October 2003 (Class 600/437); filed in Japan 21 September 2001

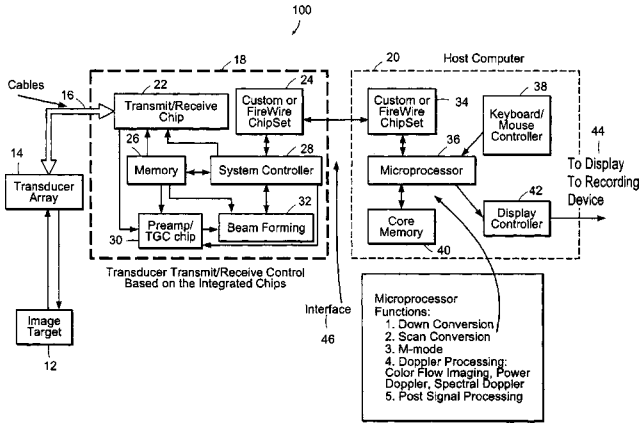
A field that defines the motion of tissue is stored, the motion of points in a sequence of fields is tracked, intermediate positions are computed, and images are produced using the values of the intermediate positions.—RCW

6,638,226

43.80.Vj ULTRASOUND IMAGING SYSTEM

Xingbai He *et al.*, assignors to TeraTech Corporation
28 October 2003 (Class 600/443); filed 28 September 2001

This system includes a special processing module and memory that are



used to map serial data into vectors, calculate autocorrelation functions and phase shifts from filtered data, and extract Doppler information.—RCW

6,638,227

43.80.Vj ULTRASOUND IMAGING METHOD AND APPARATUS USING ORTHOGONAL GOLAY CODES

Moo Ho Bae, assignor to Medison Company, Limited
28 October 2003 (Class 600/443); filed in the Republic of Korea 25 April 2001

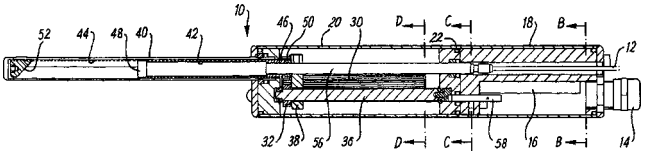
Transmit beams consisting of orthogonal Golay-coded pulses are used to permit imaging at an increased frame rate because mutually orthogonal codes can be transmitted at the same time and the orthogonality of the codes can be used to distinguish between echoes that are produced by each code.—RCW

6,638,229

43.80.Vj IMAGE PRODUCING APPARATUS

Lionel Reyes, Heanor, Derbyshire, the United Kingdom, *et al.*
28 October 2003 (Class 600/446); filed in the United Kingdom 19 October 2000

Intended for use during neurosurgery, this apparatus contains a piezo-



electric element that can be rotated and translated in a housing to obtain three-dimensional pulse-echo data for volumetric imaging.—RCW

6,641,536

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND IMAGING METHODS FOR EXTENDED FIELD OF VIEW

John A. Hossack *et al.*, assignors to Acuson Corporation
4 November 2003 (Class 600/443); filed 21 February 2002

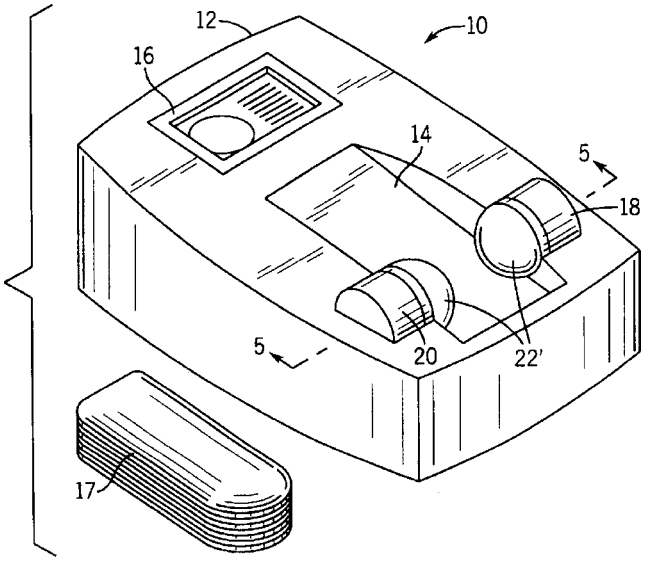
Two-dimensional essentially coplanar images are aligned and compounded. The compounding adaptively uses image regions, weighting, correlation of echoes, location within the image, and estimated motion. A user warning that depends on correlation is provided.—RCW

6,641,537

43.80.Vj MULTI-ZONE TRANSMITTER FOR QUANTITATIVE ULTRASOUND AND IMAGE MEASUREMENT

Richard Franklin Morris *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
4 November 2003 (Class 600/449); filed 20 July 2001

Two coaxial transducers are used. One transducer may be operated alone for quantitative measurements of bone. This transducer may also be



operated with a second transducer that provides broad illumination for imaging.—RCW

6,641,540

43.80.Vj MINIATURE ULTRASOUND TRANSDUCER

Aaron J. Fleischman *et al.*, assignors to The Cleveland Clinic Foundation
4 November 2003 (Class 600/459); filed 6 September 2001

This transducer is made on a substrate that has an aperture. Electronic circuitry is located on one surface of the substrate. Within the aperture and electrically connected with the circuitry is a piezoelectric element with a spherical curvature. The transducer can be fabricated using microelectromechanical system (MEMS) technology.—RCW

6,645,147

43.80.Vj DIAGNOSTIC MEDICAL ULTRASOUND IMAGE AND SYSTEM FOR CONTRAST AGENT IMAGING

John I. Jackson *et al.*, assignors to Acuson Corporation
11 November 2003 (Class 600/458); filed 25 November 1998

Using a transducer in a catheter, acoustic energy is transmitted and reflected energy from the contrast agent is received. An image of tissue and contrast is formed as a function of the reflected energy. Perfusion can be calculated without interference from other tissues.—RCW

6,647,135

43.80.Vj ULTRASONIC IMAGE PROCESSING METHOD AND SYSTEM FOR DISPLAYING A COMPOSITE IMAGE SEQUENCE OF AN ARTERY SEGMENT

Odile Bonnefous, assignor to Koninklijke Philips Electronics N.V.
11 November 2003 (Class 382/128); filed in the European Patent Office 7 December 1999

A sequence of color-coded bloodflow images is formed at one rate and another sequence of images showing wall movement is formed at another rate. Temporal markers related to the cardiac cycle are determined in each sequence. The sequences are superimposed into composite images by synchronizing corresponding markers.—RCW

6,648,824

43.80.Vj ULTRASOUND IMAGING SYSTEM FOR PERFORMING RECEIVING FOCUSING AT POINTS CORRESPONDING TO DISPLAY PIXELS AND METHOD THEREOF

Jae Sub Hwang and Tai Kyong Song, assignors to Medison Company, Limited
18 November 2003 (Class 600/437); filed in the Republic of Korea 16 February 2000

All rf echoes received in response to a transmit beam focused along a scan line are stored. The echoes are used to compute image data for points that do not necessarily correspond to pixels on the display. For the determination of data displayed, reflected signals from multiple points around the point of interest are used to detect the peak of the waveform produced by the echoes. In this way, distortion from conventional interpolation used with a limited number of scan lines can be reduced.—RCW

Acoustic backscattering form function of absorbing cylinder targets (L)

Farid G. Mitri,[a] Z. E. A. Fellah, and J. Y. Chapelon
National Institute of Health and Medical Research, INSERM Unit 556—Therapeutic Ultrasound Research Laboratory, 151 Cours Albert Thomas, 69424 Lyon, cedex 03, France

(Received 20 June 2003; accepted for publication 30 December 2003)

A new expression of the backscattering form function $|f_\infty(k_0a,\theta)|$ for cylindrical targets, suspended in an inviscid fluid in a plane incident sound field, is presented. The theory is modified to include the effects of absorption of shear and compressional waves in viscoelastic materials. The numerical results presented show how damping effect due to ultrasound absorption influences the cylinder's material properties. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1649332]

PACS numbers: 43.20.Fn [LLT] Pages: 1411–1413

I. INTRODUCTION

Echo reflected from spherical and cylindrical targets has been the subject of many theoretical (Faran, 1951; Hickling, 1962) and experimental (Hampton *et al.*, 1961; Neubauer *et al.*, 1974; Dragonette *et al.*, 1974) investigations. Dardy *et al.* (1977) computed the normalized steady-state reflected pressure or form function $|f_\infty(k_0a,\theta)|$ for aluminum cylinder that is quite in agreement with his experimental measurements.

However, the theory as it stands may not be applicable to absorbing materials such as lead, rubber, and polymeric materials, since it does not include induced loss effects due to shear and compressional waves' absorption in the solid cylinder.

In this work, a new expression of the scattered far-field form function is proposed and the theory is modified in order to calculate sound reflection from absorbing cylinders at normal incidence. This study follows the work done by Schuetz and Neubauer (1977), who calculated in a different way the far-field form function for aluminum and lucite cylinders. Here, a wide frequency range is explored and a systematic comparison with the backscattering form functions' curves for three viscoelastic materials (lucite, phenolic polymer, and polyethylene) is given with emphasis on the effect of absorption. The case of a lucite cylinder immersed in a high-density fluid is also discussed. It is difficult to include all types of absorption in solids since all the mechanisms involved are frequency dependent. In this analysis, the absorption of sound was considered of constant amplitude in function of frequency. This assumption accurately describes the behavior of many polymeric materials.

II. BACKSCATTERING FROM CYLINDRICAL TARGETS

The scattered pressure produced when an infinite elastic cylinder is placed in an incident plane acoustic field is given in the far field $r \gg a$ by (Faran, 1951)

$$P_s = P_0 \sqrt{\frac{2}{\pi k_0 r}} \sum_{n=0}^{\infty} \varepsilon_n \sin(\eta_n)\exp(i\eta_n)\cos(n\theta), \qquad (1)$$

where P_s is the scattered wave measured at the field point, P_0 is the incident pressure amplitude at the position of the cylinder center in the absence of the cylinder, r and θ are the polar coordinates, k_0 is the wave number in the fluid medium, and η_n is the phase-shift angle of the nth partial wave defined below. The backscattering form function evaluated at $\theta=\pi$ is given by (Dardy *et al.*, 1977)

$$|f_\infty(x_0,\pi)| = \sqrt{\left(\frac{2r}{a}\right)}\frac{P_s}{P_0}, \qquad (2)$$

where a is cylinder radius and $x_0 = k_0 a$.
The amplitude of the backscattered form function is then, from Eq. (1) and Eq. (2)

$$|f_\infty(x_0,\pi)| = \left|\frac{2}{\sqrt{\pi x_0}}\sum_{n=0}^{\infty}(-1)^n\varepsilon_n \sin(\eta_n)\exp(i\eta_n)\right|. \qquad (3)$$

Absorption is included by the standard method of introducing complex wave numbers into the theory (Hasegawa *et al.*, 1977).
As given by Faran (1951), the phase-shift angle is expressed by

$$\eta_n = \tan^{-1}\left[\frac{\Delta_n(x_0)(F_n+\Lambda_n(x_0))}{F_n+\Gamma_n(x_0)}\right], \qquad (4)$$

where $\Delta_n(x_0) = -[J_n(x_0)]/[Y_n(x_0)]$, with $J_n(x_0)$ and $Y_n(x_0)$ the Bessel function of the first and second kind, respectively, with order n and argument x_0.

$$F_n = \frac{\rho_0}{\rho^*}\frac{\tilde{x}_2^2}{2}\left[\frac{A_n(\tilde{x}_1)-B_n(\tilde{x}_2)}{C_n(\tilde{x}_1)-D_n(\tilde{x}_2)}\right],$$

with ρ_0 and ρ^* the mass densities of the fluid medium and cylinder material, respectively, $\tilde{x}_1 = x_1(1-i\gamma_1)$, $\tilde{x}_2 = x_2(1-i\gamma_2)$, $x_1 = (\omega/c_1)a$, $x_2 = (\omega/c_2)a$, with c_1 and c_2 the velocities of longitudinal and shear waves in the cylindrical material, and γ_1 and γ_2 the normalized longitudinal and shear absorptions, respectively (Schuetz and Neubauer, 1977).

$$A_n(\tilde{x}_1) = \frac{\Lambda_n(\tilde{x}_1)}{\Lambda_n(\tilde{x}_1)+1}; \quad B_n(\tilde{x}_2) = \frac{n^2}{\Lambda_n(\tilde{x}_2)+n^2-\frac{1}{2}\tilde{x}_2};$$

[a]Electronic mail: mitri@ieee.org

TABLE I. Material parameters used in the numerical calculations.

Material	Mass density [10^3 kg/m^3]	Compressional velocity c_1[m/s]	Shear velocity c_2[m/s]	Normalized longitudinal absorption γ_1	Normalized shear absorption γ_2
Lucite	1.191	2690	1340	0.0035	0.0053
Phenolic polymer	1.22	2840	1320	0.0119	0.0257
Polyethylene	0.957	2430	950	0.0073	0.022

$$C_n(\tilde{x}_1)=\frac{\Lambda_n(\tilde{x}_1)+n^2-\frac{1}{2}\tilde{x}_2}{\Lambda_n(\tilde{x}_1)+1};$$

$$D_n(\tilde{x}_2)=\frac{n^2(\Lambda_n(\tilde{x}_1)+1)}{\Lambda_n(\tilde{x}_2)+n^2-\frac{1}{2}\tilde{x}_2}, \tag{5}$$

with $\Lambda_n(\tilde{x}_i)=-[\tilde{x}_i J_n'(\tilde{x}_i)]/[J_n(\tilde{x}_i)]$; $i=0,1,2$, and $\Gamma_n(x_0)=-[x_0 Y_n'(x_0)]/[Y_n(x_0)]$; where $J_n'(x)$ and $Y_n'(x)$ are the first-order derivatives of Bessel function of the first and second kind, respectively, with order n and argument x.

III. NUMERICAL RESULTS AND DISCUSSION

The $|f_\infty(x_0,\pi)|$ relationship was evaluated numerically using Eq. (3). The results obtained by Dardy *et al.* (1977) were recalculated for an aluminum cylinder in the absence of absorption. Excellent agreement was found. The three materials' physical constants (Hartmann and Jarzynski, 1972; Hartmann, 1975) for which graphical results are shown, are listed in Table I. The fluid outside the cylinder was assumed to be water of density $\rho_0=1000$ kg/m^3 in which the sound velocity is $c_0=1500$ m/s. The $|f_\infty(x_0,\pi)|$ curves are plotted as function of the parameter k_0a. In addition, for ease of comparison, the same scales on the graphs have been used. A large range in k_0a was investigated (i.e., $0\leqslant k_0a\leqslant 40$) with a step of 0.004. Notice the importance of the statement of the k_0a step used in the numerical calculations, since the resonance phenomena encountered are very sharp.

The results are given in Fig. 1, Fig. 2, and Fig. 3 for lucite, phenolic polymer, and polyethylene cylinders, with and without the inclusion of absorption, respectively. Maxima and minima in $|f_\infty(x_0,\pi)|$ curves for these materials in the case of nonabsorption correspond to resonance frequencies of normal vibrational modes of the cylinder. As expected, the decrease in peak values in the case of absorption should undoubtedly be attributed to the damping due to ultrasound absorption by the cylinder material.

In an ingenious paper, Hefner and Marston (2000) studied the backscattering response of a lucite sphere in water using the ray approximation theory. They concluded that lucite material exhibits a large increase of low-frequency backscattering signals associated with tunneling to a subsonic Rayleigh wave. This can be verified also from the form-function curve of a lucite cylinder (Fig. 1) that elicits high backscattering amplitude at low frequency (low k_0a values). The presence of this backscattering enhancement may make lucite cylinders a potential candidate for passive sonar targets for cases where the narrow resonances may be used to advantage. Certain other polymers such as phenolic polymer and polyethylene may also be useful for this application (Fig. 2 and Fig. 3). Figure 4 shows an additional calculation of the backscattering form function for lucite cylinder immersed in a low-viscosity and high-density fluid (in this case mercury) with and without the effect of absorption. The fluid density is $\rho_{Hg}=13\,579.04$ kg/m^3 in which the sound velocity is $c_{Hg}=1407$ m/s (at 20°C) (Wikipedia Encyclopedia, 2004). (Here, the viscosity of the fluid surrounding the cylinder is

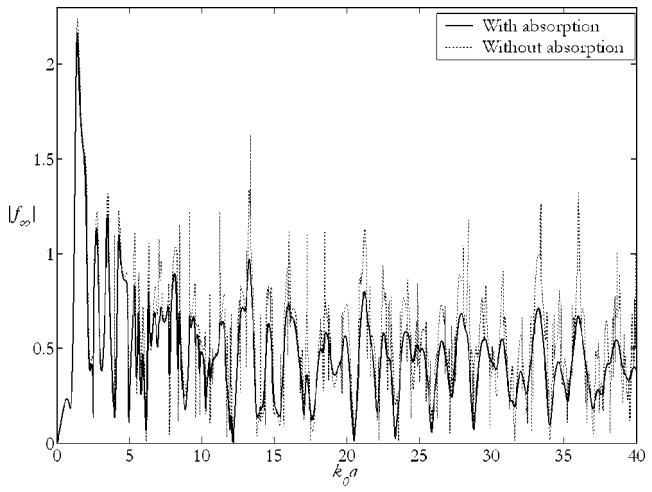


FIG. 1. Backscattering form function for a lucite cylinder immersed in water with and without the effect of absorption, respectively. One should note the damping of resonance peaks for the case of absorption.

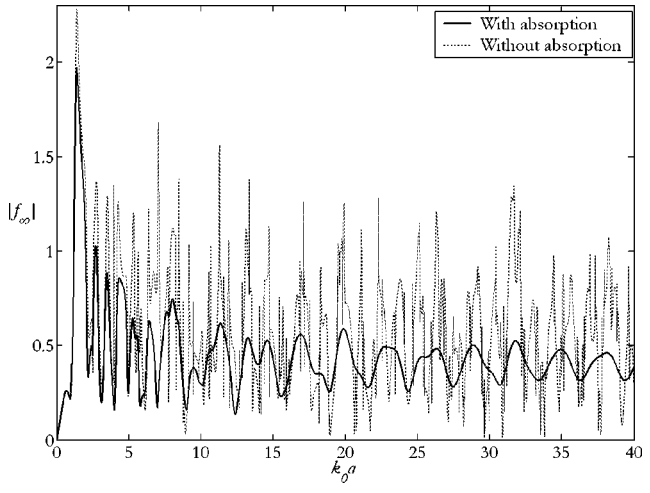


FIG. 2. Backscattering form function for a phenolic polymer cylinder immersed in water with and without the effect of absorption, respectively. The high attenuation of resonance peaks appears more clearly for this material whose absorption coefficients are greater than lucite.

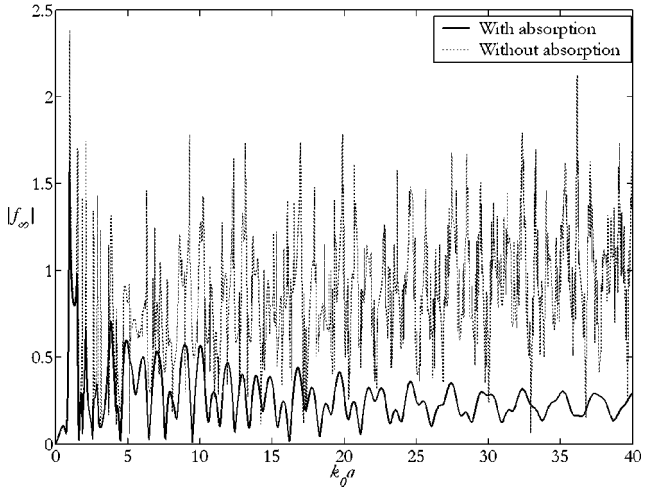


FIG. 3. Same as in Fig. 2 but for a polyethylene cylinder.

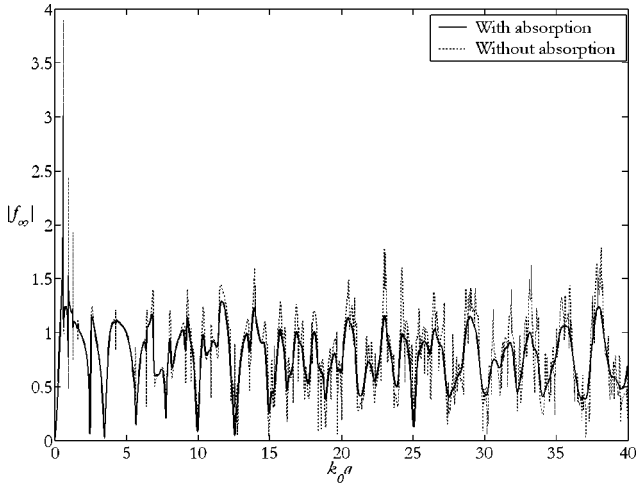


FIG. 4. Backscattering form function for a lucite cylinder immersed in a high-density fluid (mercury) with and without the effect of absorption, respectively. The high-resonance peak appears clearly at $k_0a=0.6$ (low frequency).

neglected.) As it was discussed by Hefner and Marston (2000), the fluid-loading effect produces interactions between various resonance modes that can have a significant effect on the frequency response of the cylinder. This is clearly observed in Fig. 4, where a ''giant resonance'' peak appears at low k_0a values and confirms the results of the work done on lucite spheres (Hefner and Marston, 2000).

Absorption effects upon $|f_\infty(x_0,\pi)|$ curves appear more clearly in Fig. 2 and Fig. 3 for phenolic polymer and polyethylene materials whose absorption coefficients are much greater than lucite. It is a noteworthy fact that many sharp resonance peaks tends to decrease gradually versus k_0a in the nonabsorbing case. The reason for this is that the absorption of sound (or ultrasound) by the cylinder material increases at high frequency, as it is well known.

IV. CONCLUSION

A new expression of the backscattering form function of absorbing cylinder is given and its frequency dependence in a large frequency bandwidth is analyzed for polymeric materials. The numerical calculation results are presented indicating how absorption affects the frequency dependence of the scattering properties of the cylinder material. These results are of significant importance in underwater acoustics to compute backscattered sound-field intensity from absorbing cylindrical targets. Additional work must focus on extending these computations to calculate the backscattering form function for tilted absorbent cylinders.

ACKNOWLEDGMENTS

The author gratefully acknowledges the associated editor and anonymous reviewers for their helpful comments.

This work was supported by Grant No. 01-2-93-0314 of an RNTS project from the French Ministry of Economy, Finance and Industry.

Dardy, H. D., Bucaro, J. A., Schuetz, L. S., and Dragonette, L. R. (**1977**). ''Dynamic wide-bandwidth acoustic form-function determination,'' J. Acoust. Soc. Am. **62**, 1373–1376.

Dragonette, L. R., Vogt, R. H., Flax, L., and Werner, G. (**1974**). ''Acoustic reflection from elastic spheres and rigid spheres and spheroids. II. Transient analysis,'' J. Acoust. Soc. Am. **55**, 1130–1137.

Faran, J. J. (**1951**). ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**, 405–418.

Hampton, L. D., and McKinney, C. M. (**1961**). ''Experimental study of the scattering of acoustic energy from solid metal spheres in water,'' J. Acoust. Soc. Am. **33**, 664–673.

Hartmann, B., and Jarzynski, J. (**1972**). ''Ultrasonic hysteresis absorption in polymers,'' J. Appl. Phys. **43**, 4304–4319.

Hartmann, B. (**1975**). ''Ultrasonic properties of phenolic and poly(phenilquinoxaline) polymers,'' J. Appl. Polym. Sci. **19**, 3241–3255.

Hasegawa, T., Kitagawa, Y., and Watanabe, Y. (**1977**). ''Sound reflection from an absorbing sphere,'' J. Acoust. Soc. Am. **62**, 1298–1300.

Hefner, B. T., and Marston, P. L. (**2000**). ''Backscattering enhancements associated with subsonic Rayleigh waves on polymer spheres in water: Observation and modeling for acrylic spheres,'' J. Acoust. Soc. Am. **107**, 1930–1936.

Hickling, R. (**1962**). ''Analysis of echoes from a solid elastic sphere in water,'' J. Acoust. Soc. Am. **34**, 1582–1592.

Neubauer, W. G., Vogt, R. H., and Dragonette, L. R. (**1974**). ''Acoustic reflection from elastic spheres. I. Steady-state signals,'' J. Acoust. Soc. Am. **55**, 1123–1129.

Schuetz, L. S., and Neubauer, W. G. (**1977**). ''Acoustic reflection from cylinders—nonabsorbing and absorbing,'' J. Acoust. Soc. Am. **62**, 513–517.

Wikipedia Encyclopedia, 2004.

Discussions on "Radial vibrations of orthotropic laminated hollow spheres" [J. Acoust. Soc. Am. 113, 847–851 (2003)] (L)

H. J. Ding[a] and H. M. Wang
Department of Civil Engineering, Zhejiang University, Hangzhou 310027, People's Republic of China

(Received 3 September 2003; revised 8 December 2003; accepted 30 December 2003)

With the assumption of radial vibration, there exists only a trivial solution for a general orthotropic hollow sphere and the nonzero solution can be obtained only for a spherically isotropic hollow sphere (a special case of the orthotropic hollow sphere) and the corresponding proofs are presented. So, the numerical results given by Stavsky and Greenberg [J. Acoust. Soc. Am. **113**, 847–851 (2003)] for a general orthotropic hollow sphere are unreasonable. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1649911]

PACS numbers: 43.40.Ey [ANN] Pages: 1414–1415

I. BASIC EQUATIONS

In a spherical coordinate system (r,θ,φ), for a general orthotropic hollow sphere the basic equations for three-dimensional problems are presented as follows.[1,2]

The equations of motion are

$$\begin{aligned}
&\frac{\partial\sigma_{rr}}{\partial r}+\frac{\partial\tau_{r\theta}}{r\partial\theta}+\frac{1}{r\sin\theta}\frac{\partial\tau_{r\varphi}}{r\partial\varphi}\\
&\quad+\frac{1}{r}(2\sigma_{rr}-\sigma_{\theta\theta}-\sigma_{\varphi\varphi}-\tau_{r\theta}\cot\theta)=\rho\frac{\partial^2 u_r}{\partial t^2},\\
&\frac{\partial\tau_{r\theta}}{\partial r}+\frac{\partial\sigma_{\theta\theta}}{r\partial\theta}+\frac{1}{r\sin\theta}\frac{\partial\tau_{\theta\varphi}}{r\partial\varphi}\\
&\quad+\frac{1}{r}[(\sigma_{\theta\theta}-\sigma_{\varphi\varphi})\cot\theta+3\tau_{r\theta}]=\rho\frac{\partial^2 u_\theta}{\partial t^2},\\
&\frac{\partial\tau_{\varphi r}}{\partial r}+\frac{\partial\tau_{\theta\varphi}}{r\partial\theta}+\frac{1}{r\sin\theta}\frac{\partial\sigma_{\varphi\varphi}}{r\partial\varphi}+\frac{1}{r}(2\tau_{\theta\varphi}\cot\theta+3\tau_{\varphi r})\\
&\quad=\rho\frac{\partial^2 u_\varphi}{\partial t^2}.
\end{aligned}\tag{1}$$

The strain-displacement relations are

$$\begin{aligned}
&\varepsilon_{rr}=\frac{\partial u_r}{\partial r},\quad \varepsilon_{\theta\theta}=\frac{1}{r}\frac{\partial u_\theta}{\partial\theta}+\frac{u_r}{r},\\
&\varepsilon_{\varphi\varphi}=\frac{1}{r\sin\theta}\frac{\partial u_\varphi}{\partial\varphi}+\frac{u_r}{r}+\frac{u_\theta}{r}\cot\theta,\\
&\varepsilon_{\theta\varphi}=\frac{1}{r}\left(\frac{\partial u_\varphi}{\partial\theta}-u_\varphi\cot\theta\right)+\frac{1}{r\sin\theta}\frac{\partial u_\theta}{\partial\varphi},\\
&\varepsilon_{\varphi r}=\frac{1}{r\sin\theta}\frac{\partial u_r}{\partial\varphi}+\frac{\partial u_\varphi}{\partial r}-\frac{u_\varphi}{r},\\
&\varepsilon_{r\theta}=\frac{1}{r}\frac{\partial u_r}{\partial\theta}+\frac{\partial u_\theta}{\partial r}-\frac{u_\theta}{r}
\end{aligned}\tag{2}$$

and the constitutive relations are

$$\begin{Bmatrix}\sigma_{rr}\\ \sigma_{\theta\theta}\\ \sigma_{\varphi\varphi}\\ \tau_{\theta\varphi}\\ \tau_{\varphi r}\\ \tau_{r\theta}\end{Bmatrix}=\begin{bmatrix}E_{11}&E_{12}&E_{13}&0&0&0\\ E_{12}&E_{22}&E_{23}&0&0&0\\ E_{13}&E_{23}&E_{33}&0&0&0\\ 0&0&0&E_{44}&0&0\\ 0&0&0&0&E_{55}&0\\ 0&0&0&0&0&E_{66}\end{bmatrix}\times\begin{Bmatrix}\varepsilon_{rr}\\ \varepsilon_{\theta\theta}\\ \varepsilon_{\varphi\varphi}\\ \varepsilon_{\theta\varphi}\\ \varepsilon_{\varphi r}\\ \varepsilon_{r\theta}\end{Bmatrix}.\tag{3}$$

II. PROOFS

In the following, we will prove that under the assumption of the radial vibration, only a trivial solution can be obtained for a general orthotropic hollow sphere and the nonzero solution exists only for a spherically isotropic hollow sphere (a special case of the orthotropic hollow sphere).

Under the assumption of the radial vibration, the displacement field is[3]

$$u_r=u_r(r,t),\quad u_\theta=0,\quad u_\varphi=0.\tag{4}$$

Utilizing Eq. (4), Eqs. (2) and (3) can be rewritten as

$$\varepsilon_{rr}=\frac{\partial u_r}{\partial r},\quad \varepsilon_{\theta\theta}=\varepsilon_{\varphi\varphi}=\frac{u_r}{r},\quad \varepsilon_{\theta\varphi}=\varepsilon_{\varphi r}=\varepsilon_{r\theta}=0,\tag{5}$$

$$\begin{Bmatrix}\sigma_{rr}\\ \sigma_{\theta\theta}\\ \sigma_{\varphi\varphi}\end{Bmatrix}=\begin{bmatrix}E_{11}&E_{12}&E_{13}\\ E_{12}&E_{22}&E_{23}\\ E_{13}&E_{23}&E_{33}\end{bmatrix}\begin{Bmatrix}\varepsilon_{rr}\\ \varepsilon_{\theta\theta}\\ \varepsilon_{\varphi\varphi}\end{Bmatrix}.\tag{6a}$$

$$\tau_{\theta\varphi}=\tau_{\varphi r}=\tau_{r\theta}=0.\tag{6b}$$

Utilizing Eqs. (4), (5), and (6a), we have

$$\sigma_{rr}=\sigma_{rr}(r,t),\quad \sigma_{\theta\theta}=\sigma_{\theta\theta}(r,t),\quad \sigma_{\varphi\varphi}=\sigma_{\varphi\varphi}(r,t).\tag{7}$$

By virtue of Eqs. (4), (6a), and (7), we can derive the following two equations from Eq. (1),

[a]Electronic mail: wanghuiming@zjuem.zju.edu.cn

 0001-4966/2004/115(4)/1414/2/$20.00

$$\frac{\partial \sigma_{rr}}{\partial r}+\frac{1}{r}(2\sigma_{rr}-\sigma_{\theta\theta}-\sigma_{\varphi\varphi})=\rho\frac{\partial^2 u_r}{\partial t^2}, \tag{8a}$$

$$\sigma_{\theta\theta}-\sigma_{\varphi\varphi}=0, \tag{8b}$$

where Eq. (8a) is just the same as Eq. (2) presented by Stavsky and Greenberg[3] and Eq. (8b) is the equation which is omitted by Stavsky and Greenberg. Substituting Eq. (6a) into Eqs. (8), we obtain

$$\frac{\partial^2 u_r}{\partial r^2}+2\frac{1}{r}\frac{\partial u_r}{\partial r}-\frac{K}{E_{11}}\frac{u_r}{r^2}=\rho\frac{\partial^2 u_r}{\partial t^2}, \tag{9a}$$

$$(E_{13}-E_{12})\frac{\partial u_r}{\partial r}+(E_{33}-E_{22})\frac{u_r}{r}=0, \tag{9b}$$

where

$$K=E_{22}+E_{33}+2E_{23}-E_{12}-E_{13}. \tag{10}$$

Next, we will discuss the solutions of Eqs. (9).

If $E_{13}\neq E_{12}$, the solution of Eq. (9b) is

$$u_r=C(t)r^{-\alpha}, \tag{11}$$

where

$$\alpha=(E_{33}-E_{22})/(E_{13}-E_{12}). \tag{12}$$

Substituting Eq. (11) into Eq. (9a), we have

$$\rho E_{11}\frac{\mathrm{d}^2C(t)}{\mathrm{d}t^2}+[E_{11}\alpha(\alpha-1)-K]\frac{1}{r^2}C(t)=0. \tag{13}$$

For the general case, $E_{11}\alpha(\alpha-1)-K\neq 0$, we have $C(t)=0$ from Eq. (13), and then $u_r(r,t)=0$ (trivial solution) is obtained. While if $E_{11}\alpha(\alpha-1)-K=0$, we obtain that $C(t)$ is a linear function with respect to time. It is not the problem of vibration that interested engineers and scientists.

If $E_{13}=E_{12}$ and $E_{33}\neq E_{22}$, also, only a zero solution can be obtained from Eqs. (9). So, for the above-mentioned two cases, only the trivial solutions can be obtained and they are nonsensical.

If $E_{13}=E_{12}$ and $E_{33}=E_{22}$, then Eq. (9b) is satisfied automatically, and Eq. (9a) is the unique equation which should be satisfied by the displacement field and from which the nonzero solution can be obtained.

III. CONCLUSIONS

From the above analysis, we can conclude that under the assumption of the radial vibration, only a trivial solution can be obtained for the general orthotropic hollow sphere and the nonzero solution exists only for a spherically isotropic hollow sphere (a special case of the orthotropic hollow sphere). So, using the radial vibration displacement field to discuss the radial vibrations of general orthotropic hollow spheres is unreasonable. It is noted here that in the first example considered by Stavsky and Greenberg[3] the first five frequencies for a single-layered isotropic sphere presented in their Table I agree well with those predicted by Love's classical theory. The reason is that there exist the relationships $E_{13}=E_{12}$ and $E_{33}=E_{22}$ for an isotropic sphere (a special case of the spherically isotropic sphere), and then the correct results can be obtained. In other examples, all the laminated hollow spheres involve the orthotropic materials, of which $E_{13}\neq E_{12}$ and $E_{11}\alpha(\alpha-1)-K=438.657\ (\mathrm{GPa})\neq 0$ for glass-epoxy (GE) and $E_{11}\alpha(\alpha-1)-K=-40.92\ (\mathrm{GPa})\neq 0$ for boron-epoxy (BE). As mentioned above there exist only trivial solutions for these cases. So, all the numerical results for laminated hollow spheres presented by Stavsky and Greenberg are unreasonable.

[1] A. E. H. Love, *Treatise on the Mathematical Theory of Elasticity*, 4th revised ed. (Cambridge University, Press, Cambridge, 1944).

[2] S. G. Lekhnitskii, *Theory of Elasticity of an Anisotropic Body* (Mir Publishers, Moscow, 1981).

[3] Y. Stavsky and J. B. Greenberg, "Radial vibrations of orthotropic laminated hollow spheres," J. Acoust. Soc. Am. **113**, 847–851 (2003).

Dynamic stability of a beam excited by a sequence of moving mass particles (L)

Seroj Mackertich
Engineering Science Graduate Program, Penn State University-Harrisburg, Middletown, Pennsylvania 17057

(Received 18 October 2002; revised 10 December 2003; accepted 13 January 2004)

The dynamic stability of an elastically supported Timoshenko beam exited by a continuous sequence of identical, mass particles attached to and traveling at a constant velocity across the beam has been investigated. The regions of dynamic stability are determined for different physical system parameters. Floquet theory is utilized to study the parametric regions of stability, which are displayed in graphical form. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1652035]

PACS numbers: 43.40.Cw [EGW] Pages: 1416–1419

I. INTRODUCTION

The stability of an elastically supported beam is a common technical problem in engineering and many solutions have been proposed in recent years (Abbas and Thomas, 1978; Benedetti, 1974; Bogaez, Nowakowaski, and Popp, 1986; Bolotin, 1964; Mackertich, 1999; Nelson and Conover, 1971; Newland 1970). This type of problem is interesting from a theoretical point of view, but has also many practical applications, e.g., a train traveling on a railroad track, pipeline containing flowing fluid, and flutter phenomena in aeroelastic structures. Newland (1970) studied the instability of an elastically supported beam under a traveling inertia load and concluded that the presence of moving train may therefore reduce the lateral stability of railway track, and allow buckling to occur at a reduced axial force. Nelson and Conover (1971) studied the dynamic stability of lateral response of a simply supported beam carrying a continuous series of equally spaced mass particles. They determined the dynamic stability region for different physical system parameters. Both problems were studied on the basis of the classical Bernoulli–Euler beam theory. Bogaez, Nowakowski, and Popp (1986) studied the problem of the stability of a Timoshenko beam on an elastic foundation under the moving spring-mass system and concluded that the effects of shear distortion and rotatory inertia are of great influence on the critical velocities of traveling waves. They can change the configuration of the stability regions. The problem treated here is the response of a finite, simply supported Timoshenko beam on elastic foundation excited by a sequence of constant velocity moving mass particles. The pursuing analysis determines parametric regions of dynamic stability for the transverse motion of the beam. The solution is presented within the framework of a beam theory, which includes the effects of shear deformation and rotatory inertia.

II. FORMULATION OF THE PROBLEM

Consider a finite, simply supported Timoshenko beam on an elastic foundation, initially undeformed and at rest, that is excited by a constant velocity moving masses as shown in Fig. 1. The equation of motion governing transverse vibration of the beam including shear deformation and rotatory inertia on elastic foundation can be written as (Mackertich, 1999)

$$EI\frac{\partial^4 y}{\partial x^4}-\left(P-\bar{m}\frac{\partial^2 y}{\partial t^2}\right)-\bar{m}r^2\frac{\partial^4 y}{\partial x^2\partial t^2}+\frac{EI}{k'A_0G}\frac{\partial^2}{\partial x^2}\times\left(P-\bar{m}\frac{\partial^2 y}{\partial t^2}\right)-\frac{\bar{m}r^2}{k'A_0G}\frac{\partial^2}{\partial t^2}\left(P-\bar{m}\frac{\partial^2 y}{\partial t^2}\right)+k\left(y-\frac{EI}{k'A_0G}\frac{\partial^2 y}{\partial x^2}+\frac{\bar{m}r^2}{k'A_0G}\frac{\partial^2 y}{\partial t^2}\right)+F\frac{\partial^2 y}{\partial x^2}=0, \tag{1}$$

where y is transverse displacement of the beam, F is axial force, P is the time-and-space-dependent applied force per unit length, E is Young's modulus, I is the moment of inertia of the cross-sectional area, $\bar{m}$ is the mass per unit length, k' is the shear coefficient, A_0 is the cross-sectional area, G is the shear modulus, r is the radius of gyration, ℓ is the length of the beam, and k is foundation stiffness per unit length. The general force $P(x,t)$ may be written as

$$P(x,t)=-M\left[\frac{\partial^2 y}{\partial t^2}+2\nu\frac{\partial^2 y}{\partial x\partial t}+\nu^2\frac{\partial^2 y}{\partial x^2}\right]\times\sum_{j=0}^{N-1}\delta(x-\nu t+jd), \tag{2}$$

where ν is the velocity of moving mass, M is the moving mass, N is the number of masses, d is the space between masses, $\delta(x)$ is the Dirac-delta function, and terms inside the bracket represent the transverse acceleration of the mass. To complete the formulation, the boundary conditions and initial conditions for a simply supported beam are chosen as

$$y(0,t)=y(\ell,t)=\frac{\partial\theta}{\partial x}(0,t)=\frac{\partial\theta}{\partial x}(\ell,t)=0, \tag{3}$$

$$y(x,0)=\frac{\partial y}{\partial t}(x,0)=\frac{\partial^2 y}{\partial t^2}(x,0)=\frac{\partial^3 y}{\partial t^3}(x,0)=0, \tag{4}$$

where θ is rotation of the beam. Using modal expansion method to solve the above partial differential equation, it is suggested that the solution is formulated as a superposition of the product of the modal shape function $\phi_n(x)$ and time varying modal amplitudes $A_n(t)$, that is,

 0001-4966/2004/115(4)/1416/4/$20.00

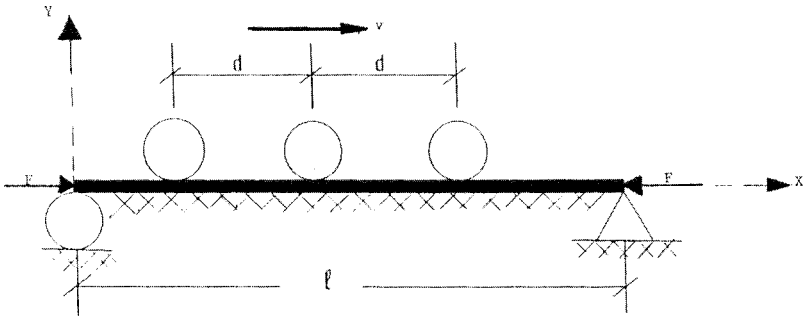


FIG. 1. Masses crossing the beam with constant velocity.

$$y(x,t)=\sum_{n}^{\infty} A_{\bar{n}}(t)\phi_{\bar{n}}(x), \tag{5}$$

where the natural mode shape functions are (Mackertich, 1992)

$$\phi_{\bar{n}}(x)=\sin\left(\frac{\bar{n}\pi x}{\ell}\right). \tag{6}$$

For a beam on elastic foundation we are interested only the buckling that will occurs at first mode of vibration. Therefore let $\bar{n}=1$

$$y(x,t)=A_1(t)\sin\left(\frac{\pi x}{\ell}\right). \tag{7}$$

Substituting (7) into (1) then integrating over space and considering the orthogonality of the mode shapes we obtain modal equation of motion for $\bar{n}=1$,

$$\beta_{1j}(t)\dddot{A}_1+\beta_{2j}(t)\ddot{A}_1+\beta_{3j}(t)\ddot{A}_1+\beta_{4j}(t)\dot{A}_1+\beta_{5j}(t)A_1=0, \tag{8}$$

where

$$\beta_{1j}(t)=1+2\Delta\sum_{j=0}^{N-1}\sin^2\left[\frac{\pi}{l}(\nu t-jd)\right],$$

$$\beta_{2j}(t)=12\alpha\Delta\sum_{j=0}^{N-1}\sin\left[\frac{\pi}{l}(\nu t-jd)\right]\cos\left[\frac{\pi}{l}(\nu t-jd)\right],$$

$$\beta_{3j}(t)=C_{n1}+12\alpha^2\Delta+2\Delta(a_1^2+a_2^2-14\alpha^2)\times\sum_{j=0}^{N-1}\sin^2\left[\frac{\pi}{l}(\nu t-jd)\right],$$

$$\beta_{4j}(t)=(4\alpha a_1^2\Delta+4\alpha a_2^2-26\alpha^3\Delta)\times\sum_{j=0}^{N-1}\sin\left[\frac{\pi}{l}(\nu t-jd)\right]\cos\left[\frac{\pi}{l}(\nu t-jd)\right],$$

$$\beta_{5j}(t)=C_{n2}-4\alpha^4\Delta+2\alpha^2\Delta(4\alpha^2-a_1^2-a_2^2)\times\sum_{j=0}^{N-1}\sin^2\left[\frac{\pi}{l}(\nu t-jd)\right],$$

$$\Delta=\frac{M}{\bar{m}\ell},\quad a_1=\frac{1}{r}\sqrt{\frac{k'A_0G}{\bar{m}}},\quad a_2=\frac{\alpha}{r\nu}\sqrt{\frac{EI}{\bar{m}}},$$

$$\alpha=\frac{\pi\nu}{\ell},\quad \gamma=\frac{k}{EI\left(\frac{\pi}{\ell}\right)^4},$$

$$C_{n1}=\frac{k'A_0G}{r^2\bar{m}}\left[\left(1+\frac{\gamma EIr^2\pi^4}{k'A_0G\ell^4}\right)+r^2\left(\frac{\pi}{\ell}\right)^2\left(1+\frac{E}{k'G}\right)\right],$$

and

$$C_{n2}=\frac{k'A_0G}{r^2\bar{m}^2}\left[EI\left(\frac{\pi}{\ell}\right)^4+\gamma EI\left(\frac{\pi}{\ell}\right)^4+\left(\frac{E^2I^2\gamma\pi^4}{k'A_0G\ell^2}-F\right)\times\left(\frac{\pi}{\ell}\right)^2\right].$$

It is convenient to write fourth-order modal differential equation of motion (8) as a system of first-order differential equations,

$$\beta_{1j}(t)\dot{\Gamma}_3=-\beta_{2j}(t)\Gamma_3-\beta_{3j}(t)\Gamma_2-\beta_{4j}(t)\Gamma_1-\beta_{5j}(t)A_1, \tag{9a}$$

$$\dot{\Gamma}_2=\Gamma_3, \tag{9b}$$

$$\dot{\Gamma}_1=\Gamma_2, \tag{9c}$$

$$\dot{A}_1=\Gamma_1. \tag{9d}$$

Above equations can be written in the matrix form

$$\{\dot{\Gamma}\}=[\Psi(t)]\{\Gamma\}, \tag{10a}$$

where

$$\{\dot{\Gamma}\}=\begin{Bmatrix}\dot{\Gamma}_3\\ \dot{\Gamma}_2\\ \dot{\Gamma}_1\\ \dot{A}_1\end{Bmatrix},\quad [\Psi(t)]=\begin{bmatrix}B_{11} & B_{12} & B_{13} & B_{14}\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix},$$

$$\{\Gamma\}=\begin{Bmatrix}\Gamma_3\\ \Gamma_2\\ \Gamma_1\\ A_1\end{Bmatrix},$$

and

$$B_{11}=-\frac{\beta_{2j}(t)}{\beta_{1j}(t)},\quad B_{12}=-\frac{\beta_{3j}(t)}{\beta_{1j}(t)},$$

$$B_{13}=-\frac{\beta_{4j}(t)}{\beta_{1j}(t)},\quad B_{14}=-\frac{\beta_{5j}(t)}{\beta_{1j}(t)}. \tag{10b}$$

The coefficient matrix $\Psi(t)$ is continuous and periodic with period τ. The response of the system can now be studied by applying Floquet theory (Cesari, 1971) to Eq. (10a), which is a first-order system of equations with periodic coefficients.

III. STABILITY ANALYSIS

According to Floquet theory (Cesari, 1971), the boundedness of the solution to (10a) is dependent on the characteristic values of

$$\det[\lambda I-\Phi(\tau)]=0, \tag{11}$$

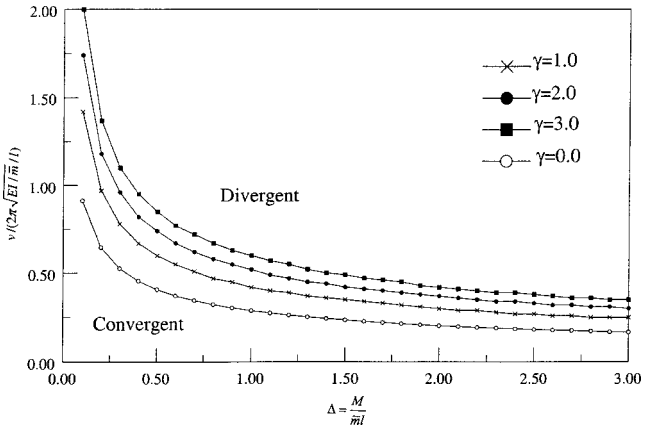


FIG. 2. Parameter plane—three particles, $\Omega=1$.

where λ is characteristic value, I is a unit matrix, and $\Phi(\tau)$ is the fundamental matrix evaluated at $t=\tau$.

To obtain the fundamental matrix we let $d=\Omega(\ell/N)$, where Ω is an arbitrary constant. There are two cases to be considered depending on the value of Ω.

Case 1: When $\Omega=1,2,3,4,...$ is an integer, Eq. (10a) can be reduced to a constant coefficient differential equation.

Case 2: When $\Omega\neq 1,2,3,...$ is not an integer. The coefficient of Eq. (10a) will remain time dependent.

First, we consider the case 1 when $\Omega=1$ is an integer. Substituting $d=(\ell/N)$ into (10b) and expanding the summation and simplifying we obtained the coefficients of the matrix $\Psi(t)$ to be equal to

$$B_{11}=B_{13}=0,$$

$$B_{12}=-\left\{\frac{\Delta[(a_1^2+a_2^2-14\alpha^2)N+12\alpha^2]+C_{n1}}{\Delta N+1}\right\},$$

and

$$B_{14}=-\left\{\frac{\Delta\alpha^2[(4\alpha^2-a_1^2-a_2^2)N-4\alpha^2]+C_{n2}}{\Delta N+1}\right\}.$$

These simplified coefficients of the matrix (10a) are constant therefore we have a system of first-order differential equations with constant coefficients,

$$\{\dot{\Gamma}(t)\}=[\Psi]\{\Gamma(t)\}, \tag{12}$$

where

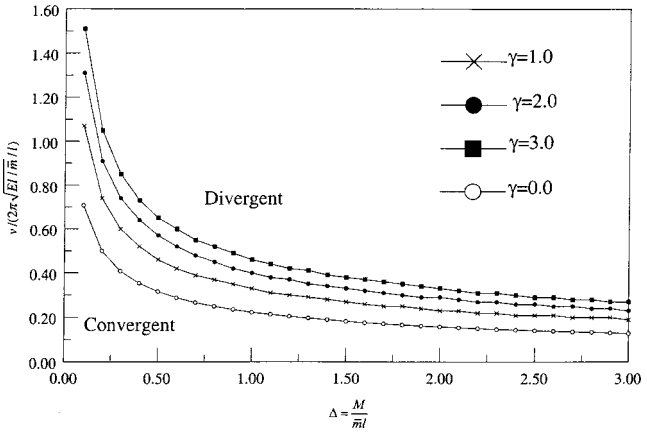


FIG. 3. Parameter plane—five particles, $\Omega=1$.

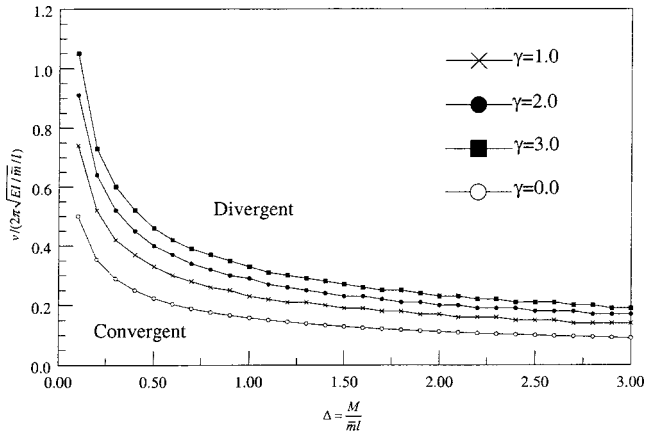


FIG. 4. Parameter plane—10 particles, $\Omega=1$.

$$\{\dot{\Gamma}\}=\begin{Bmatrix}\dot{\Gamma}_3\\ \dot{\Gamma}_2\\ \dot{\Gamma}_1\\ \dot{A}_1\end{Bmatrix},\quad [\Psi]=\begin{bmatrix}0 & B_{12} & 0 & B_{14}\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix},$$

$$\{\Gamma\}=\begin{Bmatrix}\Gamma_3\\ \Gamma_2\\ \Gamma_1\\ A_1\end{Bmatrix}.$$

Using the general method of solving systems of first-order differential equations with constant coefficients yields the fundamental matrix $[\Phi(t)]$. Substituting for $[\Phi(t=\tau)]$ in (11) and evaluating the determinate one obtains the characteristic equation, and the roots of the characteristic equation are the characteristic values. The solutions are bounded and the beam is stable if the magnitude of the characteristic value is greater than 0 and less than or equal to 1.

Next we consider case 2 when Ω is not an integer the coefficient of differential equations remain time dependent. In general the fundamental matrix cannot be determined in an analytical form, but there are several approximate methods to obtain a fundamental matrix [Sinha and Wu (1991)].

To obtain a fundamental matrix by an approximate method we replace the elements of the matrix $\Psi(t)$ by n number of piecewise constant values with the step size of τ/n in the following way:

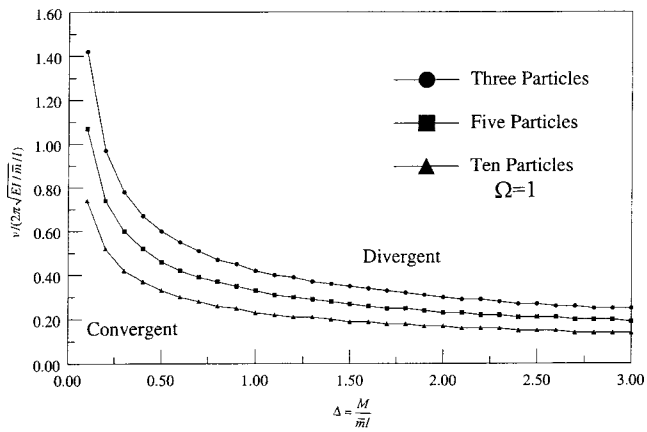


FIG. 5. Parameter plane $\gamma=1.0$.

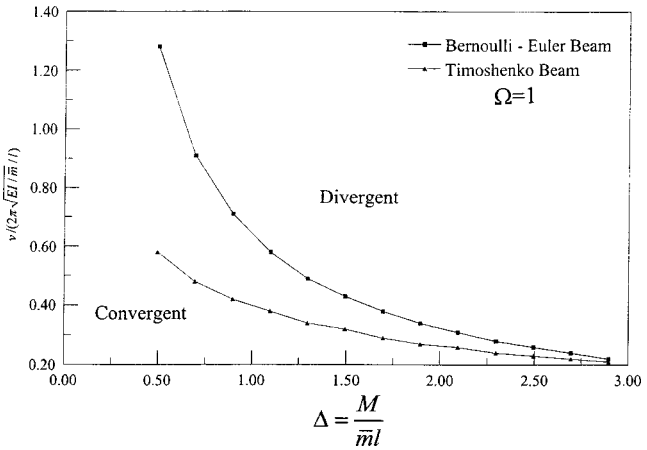


FIG. 6. Parameter plane—three particles, $\gamma=1.0$.

$$\tilde{\Psi}(t)=\begin{cases}\Psi_1=\Psi\left(\dfrac{\tau}{2n}\right) & \text{if } 0\leqslant t\leqslant\dfrac{\tau}{n},\\ \Psi_k=\Psi\left[\dfrac{\tau}{2n}(2k-1)\right] & \text{if } \dfrac{\tau}{n}(k-1)\leqslant t\leqslant\dfrac{\tau}{n}k,\\ \Psi_n=\Psi\left[\dfrac{\tau}{2n}(2n-1)\right] & \text{if } \dfrac{\tau}{n}(n-1)\leqslant t\leqslant\tau,\end{cases}$$

where matrix $\tilde{\Psi}(t)$ is also periodic with period τ. Then the fundamental matrix at $t=\tau$ is given by

$$\Phi(\tau)=\exp(t_n\Psi_n)\exp(t_{n-1}\Psi_{n-1})\cdots\exp(t_1\Psi_1),$$

where $t_1=t_2=\cdots t_n=\tau/n$.

Similarly to case 1 by examining the eigenvalues of the fundamental matrix, we can determine the stability regions. The larger the *n*, the more accurate the approximation.

IV. RESULTS AND CONCLUSIONS

In this analysis, it is of primary interest to determine dynamic stability regions for different combinations of elastic foundation constant γ and different numbers of mass particles. For a particular combination of system parameters, dynamic stability regions were determined. In order to accomplish this, each of the parameters was allowed to vary over a reasonable range of values and the Floquet theory was applied to each different combination. Figures 2–7 are the plots of the boundaries between stable and unstable regions for different combinations of the foundation constant γ and number of mass particles. The numerical method was used to obtain the plots. Figures 2–4 indicate that for keeping a constant number of mass particles on the beam, while increasing the foundation stiffness γ, the frequency of free vibration and boundary curve between stable and unstable regions rises. As the number of mass particles on the beam increases space between particles decreases, therefore one can approximate the loading as a continuous strip of moving mass along the beam. Housner (1952) studied the stability of uniform elastic pipe carrying flowing fluids. Housner determined that for a simply supported pipe the critical fluid velocity for the pipe to buckle is

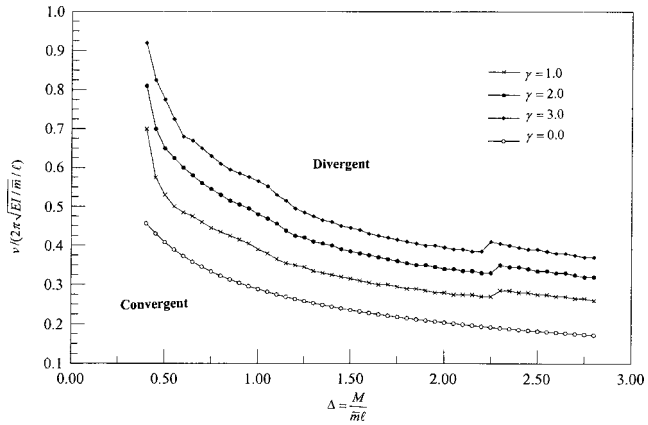


FIG. 7. Parameter plane—three particles, $\Omega=0.5$.

$$\nu_{\rm cr}=(EI/m_f)^{1/2}\pi/l. \tag{13}$$

Normalizing the critical velocity,

$$\nu_{\rm cr}/(\nu/2\pi\sqrt{EI/\bar{m}}/l)=\frac{1}{2}\sqrt{\frac{\bar{m}}{m_f}}, \tag{14}$$

where m_f is the fluid mass per unit length. An equivalent fluid mass per unit length is

$$m_f=N^*M/l, \tag{15}$$

and normalized critical velocity then becomes

$$\nu_{\rm cr}/(\nu/2\pi\sqrt{EI/\bar{m}}/l)=\frac{1}{2}\sqrt{\frac{1}{N^*\Delta}}. \tag{16}$$

The normalized critical velocity represents the lower bound for the stability and instability region in the parametric plane. In Figs. 2–7 zero foundation constant γ represents the normalized critical velocity. Figure 5 indicates that for keeping foundation stiffness γ constant and increasing the number of masses on the beam, the frequency of free vibration reduces and the boundary curve between stable and unstable regions lowers. Figure 6 is the comparison of classical Bernoulli–Euler beam and Timoshenko beam. Figure 7 represents the stable and unstable region when Ω is not an integer.

Abbas, B. A. H., and Thomas, J. (**1978**). "Dynamic stability of Timoshenko beam resting on an elastic foundation," J. Sound Vib. **60**, 33–44.

Benedetti, G. A. (**1974**). "Dynamic stability of a beam loaded by a sequence of moving mass particles," J. Appl. Mech. **41**, 1069–1071.

Bogaez, R., Nowakowski, S., and Popp, K. (**1986**). "On the stability of a Timoshenko beam on an elastic foundation under a moving spring-mass system," Acta Mech. **61**, 117–127.

Bolotin, V. V. (**1964**). Dynamic stability of elastic systems Holden-Day, San Francisco, CA.

Cesari, L. (**1971**). Asymptotic behavior and stability problems in ordinary differential equations (Springer-Verlag, New York).

Housner, G. W. (**1952**). "Bending vibrations of a pipe line containing flowing fluid," Trans. ASME, J. Appl. Mech. **74**, 205–208.

Mackertich, S. (**1992**). "Response of a beam to a moving mass," J. Acoust. Soc. Am. **92**, 1766–1769.

Mackertich, S. (**1999**). "Dynamic instability of elastically supported beam under traveling inertia load," J. Eng. Mech. 1327–1329.

Nelson, H. D., and Conover, R. A. (**1971**). "Dynamic stability of a beam carrying moving masses," J. Appl. Mech. 1003–1006.

Newland, E. (**1970**). "Instability of an elastically supported beam under a traveling inertia load," J. Mech. Eng. Sci. **12**, 373–374.

Sinha, S. C., and Wu, D. H. (**1991**). "An efficient computational scheme for the analysis of periodic systems," J. Sound Vib. **151**, 91–117.

Do tests for cochlear dead regions provide important information for fitting hearing aids?[a] (L)

Van Summers[b]
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307-5001

(Received 29 September 2003; revised 29 December 2003; accepted 6 January 2004)

For listeners with cochlear hearing loss, cochlear damage may include ''dead regions'' with no functioning inner hair cells and/or associated neurons. Recent studies indicate that amplifying frequencies more than 1.7 times the edge frequency ($1.7F_e$) of a high-frequency dead region is unlikely to improve (and may reduce) speech scores [Vickers *et al.*, J. Acoust. Soc. Am. **110**, 1164–1175 (2001); Baer *et al.*, J. Acoust. Soc. Am. **112**, 1133–1144 (2002)]. These results were taken as evidence that tests to identify dead regions could improve hearing aid fitting. In the current study, practicing audiologists examined audiograms of listeners diagnosed as having high-frequency dead regions. The audiologists were given no specific information regarding dead regions for any individual, and were asked to base amplification decisions entirely on the audiograms. Most audiologists did not recommend amplification of frequencies with hearing losses exceeding 90 dB HL. Reexamination of speech results reported by Vickers *et al.* and Baer *et al.* indicated that limiting amplification based on audiograms alone (90-dB rule) or on specific testing for dead regions ($1.7F_e$ rule) produced similar performance. Thus, testing for dead regions may not provide important information for hearing aid fitting that is not already available in the audiogram. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1649931]

PACS numbers: 43.71.Ky, 43.66.Ts [GK] Pages: 1420–1423

I. INTRODUCTION

For listeners with moderate-to-severe high-frequency hearing loss, the basal cochlea may contain ''dead regions'' where inner hair cells and/or associated neurons are entirely nonfunctional. The auditory nerve will not receive input from these regions. However, at high presentation levels, frequencies normally processed in these regions may still be detected as a result of spread of excitation to adjacent portions of the cochlea.

Vickers, Moore, and Baer (2001) and Baer, Moore, and Kluk (2002) examined whether listeners with high-frequency dead regions received benefit from amplified high-frequency speech. Psychophysical tuning curves (PTCs) and tone-detection thresholds in ''threshold-equalizing noise'' (TEN) were used to identify hearing-impaired listeners with and without high-frequency dead regions. Speech recognition performance was then tested in quiet (Vickers *et al.*, 2001) and in noise (Baer *et al.*, 2002) using amplified broadband speech and amplified low-pass speech with high-frequency regions removed [amplification based on the ''Cambridge'' fitting rule (Moore and Glasberg, 1998)]. The results of both studies indicated that the presence or absence of high-frequency dead regions was directly linked to whether subjects benefited from high-frequency speech cues. Subjects without high-frequency dead regions performed best in broadband speech. Subjects with dead regions did as well, and sometimes better, in low-pass speech than in broadband conditions. Both Vickers *et al.* (2001) and Baer *et al.* (2002) interpreted their results as suggesting that tests to diagnose the presence and frequency extent of dead regions may provide a useful clinical tool in making amplification decisions, and that amplification of frequencies falling well within high-frequency dead regions may not be advisable. Specifically, amplification of frequencies more than 1.7 times the edge frequency of a high-frequency dead region ($1.7F_e$) may not provide any benefit (Vickers *et al.*, 2001; Baer *et al.*, 2002).

A complication in interpreting these results is that in both Vickers *et al.* (2001) and Baer *et al.* (2002), listeners diagnosed as having high-frequency dead regions had significantly more high-frequency hearing loss than listeners without dead regions. Audiograms for the ten ears with no dead regions and nine ears with dead regions from the two studies are plotted in Fig. 1 (solid lines and dashed lines, respectively; the heavy solid line and the filled symbols in the figure will be described later). Clearly, the audiograms for these two sets of ears diverge above 2000 Hz, with ears diagnosed as having high-frequency dead regions showing greater high-frequency losses. This leads to the question of whether a separate test for dead regions is likely to be clinically valuable or if the relevant information provided by this testing is already available in the audiograms. Rankovic (2002) recently reported articulation index (AI) analyses of the speech results reported by Vickers *et al.* (2001) and concluded that ability to benefit from high-frequency amplification can be accurately predicted based on the audiogram and presentation levels alone [but see Moore (2002) for an alternative interpretation of Rankovic's findings].

Based on Rankovic (2002) and Moore (2002), it is unclear whether tests for high-frequency dead regions will lead to better clinical decisions about high-frequency amplification than can be made based on audiograms alone. To exam-

[a]These data were presented at the 15th Annual Convention of the American Academy of Audiology in San Antonio, Texas, April 2003.

[b]Electronic mail: walter.summers@na.amedd.army.mil

 0001-4966/2004/115(4)/1420/4/$20.00

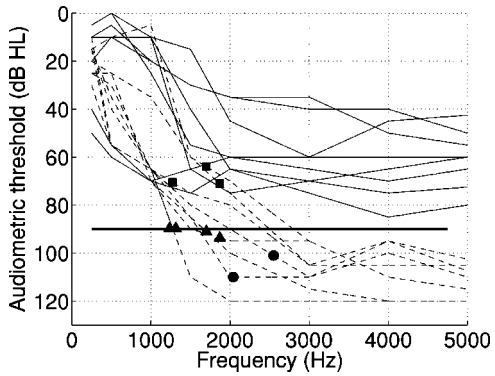


FIG. 1. Audiograms for nine ears with high-frequency dead regions (dashed lines) and ten ears with no dead regions (solid lines) as previously reported in Vickers *et al.* (2001) and Baer *et al.* (2002). Intersections of audiograms with the heavy line at 90 dB HL indicate cutoff frequencies for amplification based on the 90-dB rule. Symbols indicate cutoffs based on 1.7 times the estimated low-frequency edge of the dead region.

ine this question, practicing audiologists were shown audiograms for ears diagnosed as having high-frequency dead regions. Based solely on the audiograms, the audiologists were asked to assess whether they would prescribe broadband amplification, and, if not, to estimate the frequency limit of where amplification was likely to provide benefit. As reported below, experienced audiologists were unlikely to expect benefit from amplification once thresholds exceeded about 90 dB HL. The speech results reported by Vickers *et al.* (2001) and Baer *et al.* (2002) were then reexamined to see whether limiting high-frequency amplification based on testing for dead regions led to better speech performance than limiting based on the audiogram and a 90 dB HL ''rule of thumb.''

II. METHODS

Eleven audiologists working in the Army Audiology and Speech Center at Walter Reed Army Medical Center participated by completing a one-page handout. The handout contained a single graph with six lines representing audiograms for the six ears identified as having high-frequency dead regions in Baer *et al.* (2002, Table I). These are six of the nine ears with dead regions shown in Fig. 1. The following instructions appeared on the handout: ''The graph below shows thresholds for six hearing-impaired listeners. If these were the only data you had for each listener, would you be likely to attempt broadband amplification aimed at providing signal audibility all the way up to 5 kHz in every case? If not, at what frequency would you be likely to ''give up''? Please place an X on each of the six threshold lines at the frequency where you feel the loss has become severe enough so that amplification is probably unlikely to provide benefit. In cases where you would provide broadband amplification, put your X at 5 kHz.''

III. RESULTS

Figure 2 shows the six audiograms included in the handout and the responses provided by one of the audiologists (the X's). Note that these data show little variability on the *y* axis, and that this audiologist would be unlikely to provide

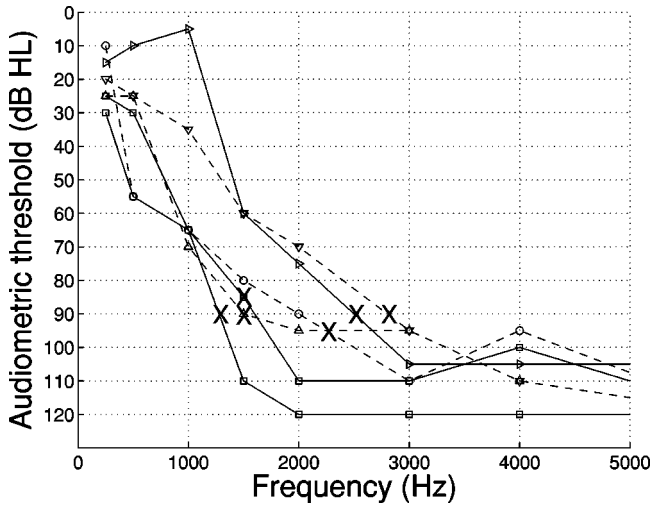


FIG. 2. The six audiograms as presented in the handout and responses provided by one audiologist (X's). Note that the handout used a linear frequency scale rather than the log scale more commonly seen in audiograms. This allowed for finer frequency distinctions to be indicated in the region above 1000 Hz.

high-frequency amplification once the hearing loss exceeded about 90 dB HL. For each audiologist, threshold values for the six X's on the completed handout (in dB HL) were averaged to get a mean amount of hearing loss where high-frequency amplification was likely to be discontinued. These mean values (and their standard deviations) are plotted in Fig. 3 as a function of years of experience as a practicing audiologist.

The data in Fig. 3 show some variability that appears to be related to years of clinical experience. Means displayed in the right-hand part of the figure, representing audiologists with more than 10 years of experience, indicate that amplifying high frequencies would not be attempted when losses exceeded about 90 dB HL. The left-hand portion of the figure shows that most of the audiologists with fewer years of experience would provide amplification for greater amounts of loss. The data from audiologists with over 10 years of

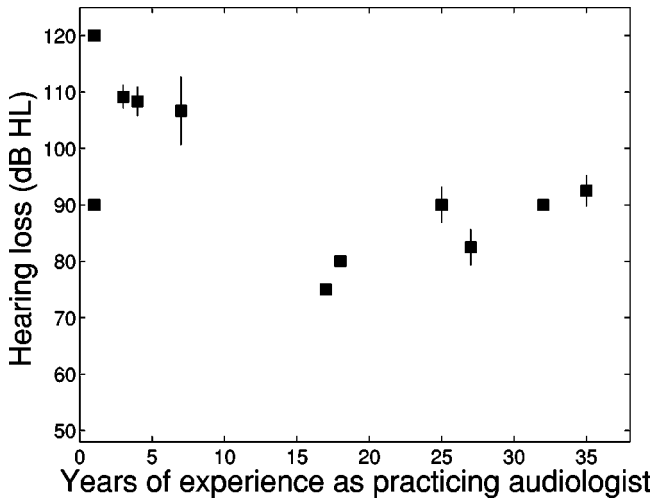


FIG. 3. Mean pure-tone hearing loss at the recommended cutoff frequency for amplification based on responses to the handout. Means are plotted as a function of years of experience as a practicing audiologist. Error bars indicate ± one standard deviation.

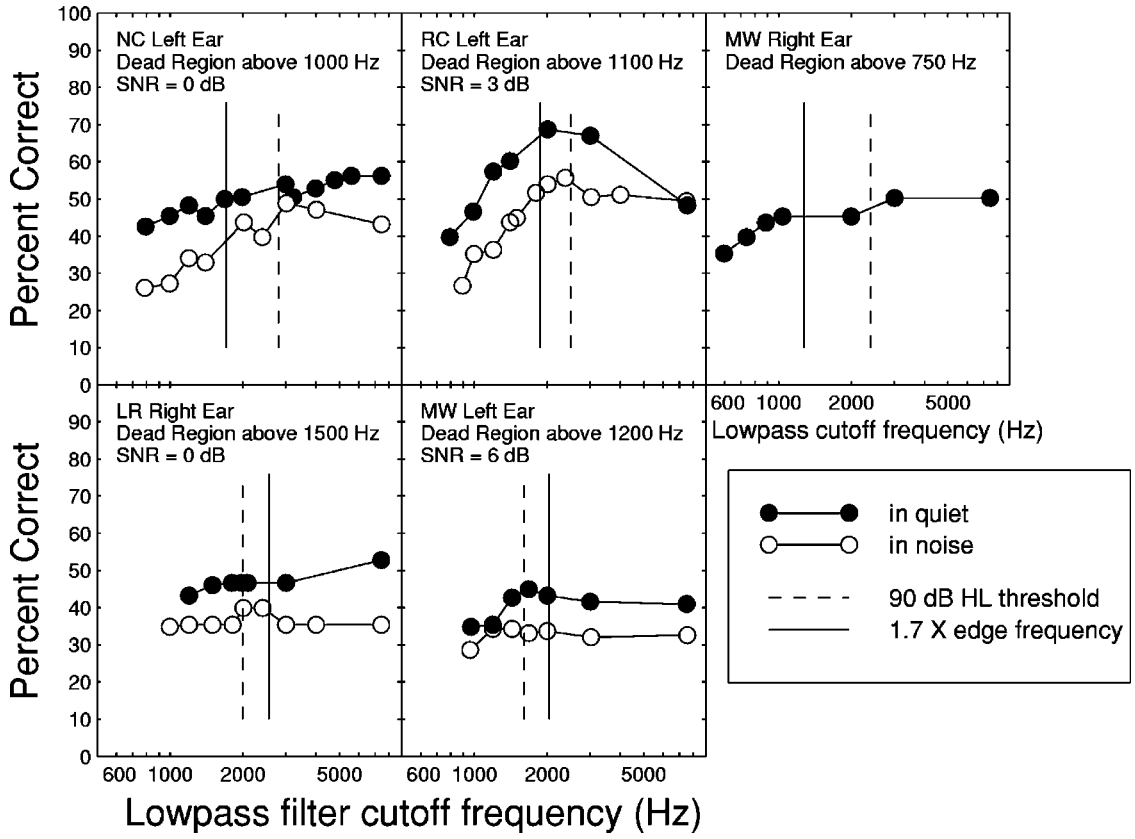


FIG. 4. Identification performance for amplified speech as a function of low-pass filter cutoff frequency [data originally reported in Vickers *et al.* (2001) and Baer *et al.* (2002)]. Vertical lines indicate recommended cutoff frequencies for amplification based on 90-dB and $1.7F_e$ rules (dashed and solid lines, respectively). Signal-to-noise ratios used during in-noise testing are indicated in figure panels.

experience suggest a simple rule. That is, limit high-frequency amplification to frequencies with losses no greater than 90 dB HL. This ''90-dB rule'' provides a recommended frequency cutoff for high-frequency amplification that can be compared to the cutoff indicated by the testing for dead regions ($1.7F_e$ rule). (It should be noted that most of the audiologists did not simply follow the 90-dB rule in responding to the survey—there were in fact only two instances where surveys were returned with all six responses at 90 dB HL. The 90-dB rule was selected for comparison to the $1.7F_e$ rule because the audiologists with more than 10 years of experience generally did not recommend amplification when thresholds exceeded 90 dB HL and because the rule provides a recommended frequency cutoff for high-frequency amplification directly from the audiogram.)

Figure 1 can now be reconsidered in terms of the recommended cutoff frequencies for high-frequency amplification based on the two rules (90-dB and $1.7F_e$). Using the 90-dB rule, amplification would only be provided for frequencies where thresholds fall above the heavy solid line in the figure (90 dB HL). Note first that all thresholds fall above this line for the ears with no dead regions. Thus, based on the 90-dB rule, broadband amplification would be indicated for these ears. This fits with the results reported by Vickers *et al.* (2001) and Baer *et al.* (2002): these subjects performed best with broadband amplification. For these ten ears, the 90-dB rule led to the same recommendation in terms of amplification as the additional testing required to specifically identify dead regions. Thus, for these ears, it appears that testing for dead regions would not have affected clinical decisions regarding amplification.

For the nine ears with high-frequency dead regions, recommended cutoff frequencies for high-frequency amplification based on the 90-dB rule can be compared to recommended cutoffs based on testing to locate the edge frequencies of dead regions. For these ears, the filled symbols in Fig. 1 indicate cutoff frequencies based on the $1.7F_e$ rule [using edge frequency values reported by Vickers *et al.* (2001) and Baer *et al.* (2002)]. For four of these ears, the recommended cutoff frequency based on $1.7F_e$ was nearly equal to 90 dB HL (triangles in Fig. 1). Thus, for these ears, the recommended cutoff frequency for high-frequency amplification would be essentially equivalent using either the 90-dB or the $1.7F_e$ rules. These ears represent additional instances where testing to assess dead regions would not have altered decisions about amplification.

Cutoff frequencies based on $1.7F_e$ were lower than cutoffs based on the 90 dB HL threshold for three ears in Fig. 1 (squares in the figure). That is, for these three ears, the 90-dB rule recommends extending amplification to slightly higher frequencies than $1.7F_e$. For two ears the $1.7F_e$ rule recommends amplifying frequency regions with more than 100 dB of hearing loss (circles in Fig. 1). In these instances, the 90-dB rule limits high-frequency amplification slightly more than the $1.7F_e$ rule.

The five ears considered in the previous paragraph are cases in which the 90-dB and $1.7F_e$ rules led to slightly different recommendations in terms of high-frequency amplifi-

cation. In these instances, speech recognition data can be examined to determine if either rule led to better overall performance. The speech results reported by Vickers *et al.* (2001) and Baer *et al.* (2002) for these ears allow this comparison and are replotted in Fig. 4. Each panel of the figure shows speech recognition performance by a single ear for amplified speech under various lowpass conditions. Within each panel, performance is plotted as a function of the lowpass cutoff frequency of amplified speech. Movement to the right within each panel shows how performance changed as more and more high-frequency portions of the signal were presented. The rightmost point in each panel represents amplified broadband speech. As noted by both Vickers *et al.* (2001) and Baer *et al.* (2002), for these and other listeners with high-frequency dead regions, there appears to be a point at which performance stops improving with the addition of more high-frequency cues. For some ears, (e.g., RC–left ear), lowpass filtering appears to result in better performance than broadband amplification. The question is whether the 90-dB rule provides as accurate an estimate of the appropriate cutoff frequency for amplification in these ears as the cutoff frequency indicated by specific testing for dead regions. The vertical lines in Fig. 4 indicate recommended cutoff frequencies for amplification based on the 90 dB and $1.7F_e$ rules and allow an examination of this question.

The upper panels in Fig. 4 represent ears where cutoff frequencies were slightly higher based on the 90-dB rule than the $1.7F_e$ rule (ears represented by squares in Fig. 1). For these ears, speech scores showed a small improvement when amplification was extended above $1.7F_e$ up to the 90-dB threshold. For the data in these three panels, speech scores (measured or estimated based on interpolation where necessary) were, on average, about 3% higher in the more broadband amplification provided by the 90-dB rule than the more narrow-band amplification prescribed by the $1.7F_e$ rule. This small improvement in performance was statistically significant $[t(4)=2.848, p<0.05]$.

The lower panels in Fig. 4 represent the two cases where amplification would extend to higher frequencies based on the $1.7F_e$ rule than based on the 90-dB rule (circles in Fig. 1). For these two ears, the $1.7F_e$ rule leads to amplification of frequency regions with more than 100 dB of hearing loss. The amplification of frequencies with this much hearing loss did not lead to improved performance over amplification restricted to frequencies with hearing loss no greater than 90 dB HL.

It was noted already that the 90-dB rule provides only an approximation to the actual frequency cutoff values recommended by the audiologists with more than 10 years experience who were surveyed. In general, the individual cutoff frequencies recommended by these clinicians produced similar percent correct scores in the speech tasks as the 90-dB and $1.7F_e$ rules. For these (and perhaps many other) experienced audiologists, specific testing for dead regions may not improve hearing aid fitting decisions that would be made without this testing.

IV. CONCLUSIONS

Testing for the presence of cochlear dead regions may be of clinical value if the results lead to improved decisions about amplification for hearing-impaired listeners. However, if the important clinical information provided by this testing is available directly from the audiogram, the additional testing is clearly unnecessary. In the current study, a simple rule based directly on the audiogram led to limiting high-frequency amplification for the same listeners and for similar frequency regions as indicated by specific testing for dead regions. The results suggest that tests to identify high-frequency dead regions do not improve clinical decisions relating to high-frequency amplification compared to a simple rule based on absolute thresholds.

ACKNOWLEDGMENTS

This research was supported by Grant No. DC 03553 from NIH to Walter Reed Army Medical Center. The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

Baer, T., Moore, B. C. J., and Kluk, K. (**2002**). ''Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies,'' J. Acoust. Soc. Am. **112**, 1133–1144.

Moore, B. C. J. (**2002**). ''Response to 'Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions','' J. Acoust. Soc. Am. **111**, 2549–2550.

Moore, B. C. J., and Glasberg, B. R. (**1998**). ''Use of a loudness model for hearing aid fitting,'' Br. J. Audiol. **32**, 301–319.

Rankovic, C. M. (**2002**). ''Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions,'' J. Acoust. Soc. Am. **111**, 2545–2548.

Vickers, D. A., Moore, B. C. J., and Baer, T. (**2001**). ''Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies,'' J. Acoust. Soc. Am. **110**, 1164–1175.

Fractional Laplacian time-space models for linear and nonlinear lossy media exhibiting arbitrary frequency power-law dependency

W. Chen[a)]
Simula Research Laboratory, P.O. Box 134, 1325 Lysaker, Oslo, Norway and Institute of Applied Physics and Computational Mathematics, P.O. Boc 8009, Division Box 26, Beijing 100088, P.R. China.

S. Holm
Simula Research Laboratory, P.O. Box 134, 1325 Lysaker, Oslo, Norway

(Received 5 February 2003; revised 11 July 2003; accepted 11 December 2003)

Frequency-dependent attenuation typically obeys an empirical power law with an exponent ranging from 0 to 2. The standard time-domain partial differential equation models can describe merely two extreme cases of frequency-independent and frequency-squared dependent attenuations. The otherwise nonzero and nonsquare frequency dependency occurring in many cases of practical interest is thus often called the anomalous attenuation. In this study, a linear integro-differential equation wave model was developed for the anomalous attenuation by using the space-fractional Laplacian operation, and the strategy is then extended to the nonlinear Burgers equation. A new definition of the fractional Laplacian is also introduced which naturally includes the boundary conditions and has inherent regularization to ease the hypersingularity in the conventional fractional Laplacian. Under the Szabo's smallness approximation, where attenuation is assumed to be much smaller than the wave number, the linear model is found consistent with arbitrary frequency power-law dependency. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1646399]

PACS numbers: 43.20.Bi, 43.20.Hq, 43.35.Bf, 43.35.Cg [MFH] Pages: 1424–1430

I. INTRODUCTION

Frequency-dependent attenuation has been observed in a wide range of important engineering areas such as acoustics (Blackstock, 1985; Szabo, 1994; Wojcik *et al.*, 1995), viscous dampers in seismic isolation of buildings (Makris and Constantinou, 1991), structural vibration (Enelund, 1996; Rusovici, 1999; Adhikari, 2000), seismic wave propagation (Caputo, 1967; Caputo and Mainardi, 1971), anomalous diffusions occurring in porous media (Hanyga, 1999), just to mention a few. This frequency dependency is described by

$$E = E_0 e^{-\alpha(\omega) z}, \tag{1}$$

where E denotes the amplitude of an acoustic field variable such as velocity or pressure, and ω represents angular frequency. Coefficient $\alpha(\omega)$ is often characterized with an empirical power law

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

for a wide range of frequencies of practical interest, in which α_0 and y are media-specific attenuation parameters obtained through a fitting of measured data.

The most straightforward strategy in computer simulation of the power-law lossy behavior is to do both mathematical and numerical modeling in the frequency domain via the Laplace transform (Ginter, 2000). The drawbacks of this approach are that the frequency-domain methods are often ineffective for nonlinear problems and the numerical inverse Laplace transform is very tedious and expensive. The time-domain simulation, in contrast, is feasible for general nonlinear problems and relatively easier to implement and less costly (Wismer and Ludwig, 1995). In addition, the time-domain models also outperform the frequency-domain models as they allow numerical simulation of various initial and boundary value problems (Hanyga, 2001a).

However, it has long been noted that common time-domain partial differential equations (PDE) can model merely two extreme cases of frequency-independent ($y=0$) and frequency-squared dependent ($y=2$) absorption behaviors. In many cases of practical interest such as acoustics in biomedical materials and fractal rock layers, $0<y<2$ mostly appears and the standard time-domain PDE modeling methodology does not apply (Blackstock, 1985; Nachman *et al.*, 1990; Szabo, 1994). In contrast to the $y=0,2$ attenuations well described by the standard PDEs, the attenuations obeying $0<y<2$ power law are thus often called the anomalous diffusion (Hanyga, 2001c), nonexponential relaxation, inelastic damping (Adhikari, 2000), hysteretic damping (Gaul, 1999), singular hereditary or singular memory media (Hanyga, 1999), originating from different engineering applications.

The recent decade has witnessed increasing attention to accurate time-domain mathematical modeling of such anomalous ($0<y<2$) attenuation phenomena, due to a dramatic increase in computer simulation of acoustic wave propagation through human tissues and irregular porous random media. Among these existing models are the adaptive proportional damping model (Wojcik *et al.*, 1995, 1999), the time-domain model via finite frequency decomposition (He, 1998; Chen and Holm, 2002a), the Z-transform model (Wismer and Ludwig, 1995), the multiple relaxation model

a)Electronic mail: chen_wen@iapcm.ac.cn

 0001-4966/2004/115(4)/1424/7/$20.00

(Nachman *et al.*, 1990; Mast *et al.*, 2001; Yuan *et al.*, 1999), the fractional time derivative models (Caputo 1967; Bagley and Torvik, 1983; Ochmann and Makarov, 1993), and Szabo's model via the singular convolution kernel (Szabo, 1993, 1994). As mentioned by Blackstock (1985), the space or space-time modeling of thermoviscous behavior is often replaced by a pure-time operation in the above models under the condition that the thermoviscous term is relatively small. Ochmann and Makarov (1993) further elaborate that this replacement is impossible in the general case where the interaction between two oppositely traveling sound waves cannot be neglected. The preference of the time-only expression is mostly due to its ease of analysis. For instance, the time-space representation $\partial^3 p/\partial t \partial z^2$ in the one-dimension thermoviscous wave equation, where t and z are, respectively, time and space variables [see Eq. (19) further below], is approximated by a triple time derivative $\partial^3 p/\partial t^3$ (Blackstock 1967; Pierce, 1989; Szabo, 1994). However, numerical implementations of the time-only models are still uncommon, and most research is now restricted to the related mathematical analysis partly due to great numerical difficulties involved. In addition, when $1<y$, fractional time derivative involves the initial condition of the second-order derivative which is unavailable in most practical problems. It is well known that anomalously attenuative and dispersive media often establish complicated microstructures in space; the spatial fractional derivative models may therefore instead be more suitable as a modeling approach, where the initial condition of second-order derivative is never required (Hanyga, 2001b).

The purpose of this study is to employ the spatial fractional Laplacian, also sometimes called the fractional Laplace operator and the Riesz derivative, instead of the Szabo's time convolutional integral and the time fractional derivative to develop linear and nonlinear mathematical models of anomalous thermoviscous behaviors characterized by nonzero and nonquadratic frequency dependency. It is known (Samko *et al.*, 1987) that the standard definition of the fractional Laplacian leads to a hypersingular convolution integral as in the Riemann–Liouville fractional derivative. We present the new definition of the fractional Laplacian which naturally includes the boundary conditions and has inherent regularization operation to ease the hypersingularity of the convolution kernel function. Therefore, it is more useful for engineering modeling.

In what follows, the new definition of the fractional Laplacian is introduced first in Sec. II, followed by a presentation and analysis of the linear fractional Laplacian thermoviscous models of wave equation in Sec. III. The corresponding nonlinear models are then developed in Sec. IV. Conclusions are presented in Sec. V. In the Appendix, a finite-element numerical model is briefly discussed.

II. FRACTIONAL LAPLACIAN

It is worth pointing out that the fractional Laplacian and the fractional derivative are two related but different mathematical concepts. Both are defined through a singular convolution integral, but the former is guaranteed to be the positive definition like the standard Laplace operator, while the latter (Diethelm, 2000) is not.

The conventions of Fourier transform used in Pierce (1989) and Szabo (1994) are also employed in this study. Namely

$$F_-\left(\frac{\partial^n \phi}{\partial z^n}\right) = (ik)^n \Phi(k,\omega), \tag{3}$$

$$F_+\left(\frac{\partial^n \phi}{\partial t^n}\right) = (-i\omega)^n \Phi(k,\omega), \tag{4}$$

where $\Phi(k,\omega)$ is the time and space two-dimensional Fourier transforms of a sufficiently good function $\phi(z,t)$

$$\Phi(k,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(z,t) e^{-i(kz-\omega t)} dz\, dt, \tag{5}$$

where k is the wave number. The inverse of the space Fourier transform is designated as F_-^{-1}, and the inverse of the time Fourier transform F_+^{-1}.

A common interpretation of the fractional Laplacian is to employ the inverse of its Fourier transform (e.g., see Samko *et al.*, 1987; Jespersen, 1999), i.e.,

$$F_-\{(-\nabla^2)_*^{s/2}\varphi\} = k^s \Phi, \quad 0<s<2, \tag{6}$$

$$(-\nabla^2)_*^{s/2}\varphi = F_-^{-1}\{k^s \Phi\} = \frac{1}{2\pi}\int \Phi k^s e^{ikx} dk. \tag{7}$$

The fractional Laplacian is also often called the Riesz fractional derivative in terms of the Riesz potential (Gorenflo and Mainardi, 1998). The Riesz potential I_d^s of order s of d dimensions reads (Zahle, 1997; Samko *et al.*, 1987)

$$I_d^s \varphi(x) = \frac{\Gamma[(d-s)/2]}{\pi^{s/2} 2^s \Gamma(s/2)} \int_\Omega \frac{\varphi(\xi)}{\|x-\xi\|^{d-s}} d\Omega(\xi), \quad 0<s<2, \tag{8}$$

where Γ denotes the Euler's gamma function, Ω is integral domain. The traditional definition of the fractional Laplacian involves the approximate finite difference expression (Samko *et al.*, 1987) and is not well suited for multidimensional irregular domain. By analogy with the fractional time derivative, we give an analytical definition below

$$(-\nabla^2)_*^{s/2}\varphi(x) = -\nabla^2[I_d^{2-s}\varphi(x)]. \tag{9}$$

It is known that the Laplacian operator has the expression

$$\nabla^2 \varphi(x) = \frac{d^2\varphi}{dr^2} + \frac{d-1}{r}\frac{d\varphi}{dr}, \tag{10}$$

where $r=\|x-\xi\|$. Equation (9) can then be reduced to

$$(-\nabla^2)_*^{s/2}\varphi(x)$$
$$= -\frac{\Gamma[(d-2+s)/2]}{\pi^{2-s/2}2^{2-s}\Gamma[(2-s)/2]}\nabla^2\int_\Omega \frac{\varphi(\xi)}{\|x-\xi\|^{d-2+s}}d\Omega(\xi)$$
$$= -\frac{(d-2+s)s\Gamma[(d-2+s)/2]}{\pi^{(2-s)/2}2^{2-s}\Gamma[(2-s)/2]}\int_\Omega \frac{\varphi(\xi)}{\|x-\xi\|^{d+s}}d\Omega(\xi). \tag{11}$$

It is noted that (11) encounters the detrimental hypersingularity, which means the singularity order $d+s$ is larger than the topological dimension d. An alternative way is thus presented below to define the fractional Laplacian

$$(-\nabla^2)^{s/2}\varphi(x) = -I_d^{2-s}[\nabla^2\varphi(x)]$$
$$= -\frac{\Gamma[(d-2+s)/2]}{\pi^{(2-s)/2}2^{2-s}\Gamma[(2-s)/2]}$$
$$\times\int_\Omega \frac{\nabla^2\varphi(\xi)}{\|x-\xi\|^{d-2+s}}d\Omega(\xi). \tag{12}$$

It is noted that the definition (12) has a weak singularity of order $d-2+s$ compared with the hypersingularity of order $d+s$ in (11). The Green's second identity is useful to connect (12) and (11), and can be stated as

$$\int_\Omega v\nabla^2\varphi\, d\xi = \int_\Omega \varphi\nabla^2 v\, d\Omega(\xi) - \int_S\left(\varphi\frac{\partial v}{\partial n} - v\frac{\partial\varphi}{\partial n}\right)dS(\xi), \tag{13}$$

where S represents the surface of the domain, and n is the unit outward normal. Let

$$v = 1/\|x-\xi\|^{d-2+s}, \tag{14}$$

and

$$\varphi(x)|_{x\in S} = D(x), \tag{15}$$

$$\left.\frac{\partial\varphi(x)}{\partial n}\right|_{x\in S} = N(x), \tag{16}$$

With the Green's second identity, the definition (12) is then reduced to

$$(-\nabla^2)^{s/2}\varphi(x) = -\frac{(d-2+s)s\Gamma[(d-2+s)/2]}{\pi^{(2-s)/2}2^{2-s}\Gamma[(2-s)/2]}\int_\Omega\frac{\varphi(\xi)}{\|x-\xi\|^{d+s}}d\Omega(\xi) + h\int_S\left[\varphi(\xi)\frac{\partial}{\partial n}\left(\frac{1}{\|x-\xi\|^{d+s-2}}\right) - \frac{1}{\|x-\xi\|^{d+s-2}}\frac{\partial\varphi(\xi)}{\partial n}\right]dS(\xi)$$
$$= (-\nabla^2)_*^{s/2}\varphi(x) + h\int_S\left[D(\xi)\frac{\partial}{\partial n}\left(\frac{1}{\|x-\xi\|^{d+s-2}}\right) - \frac{N(\xi)}{\|x-\xi\|^{d+s-2}}\right]dS(\xi), \tag{17}$$

where

$$h = \frac{\Gamma[(d-2+s)/2]}{\pi^{(2-s)/2}2^{2-s}[(2-s)/2]}. \tag{18}$$

It is seen from (17) that the fractional Laplacian definition $(-\nabla^2)^{s/2}$ is considered the fractional Laplacian derivative $(-\nabla^2)_*^{s/2}$ augmented with the boundary integral, which is a parallel to the fractional time derivatives in the Caputo sense relative to that in the Riemann–Liouville sense.

The above two definitions $(-\nabla^2)_*^{s/2}$ and $(-\nabla^2)^{s/2}$ involve only the symmetric fractional Laplacian for isotropic media. To simplify the illustration of the basic idea of this study without loss of generality, we only consider isotropic media in this paper. For the traditional definition of the anisotropic fractional Laplacian see Feller (1971) and Hanyga (2001). By analogy with the definitions (11) and (12), it will be straightforward to have the corresponding new expression of the anisotropic fractional Laplacian.

Albeit a long history, the research on the space fractional Laplacian still appears fairly poor in the literature (Gorenflo and Mainardi, 1998). In recent years, some interest has arisen from anomalous diffusion problems. Readers are advised to find more detailed description of the fractional Laplacian from Samko *et al.* (1987), Zaslavsky (1994), Gorenflo and Mainardi (1998), Hanyga (2001), and references therein. In the Appendix, we briefly discuss the finite-element numerical model of the fractional Laplacian.

III. LINEAR FRACTIONAL LAPLACIAN THERMOVISCOUS MODEL

Szabo (1994) started his time-domain model building and causality analysis of the attenuation power law with the thermoviscous wave equation (Blackstock, 1967; Lighthill, 1980; Pierce, 1989), also known as the augmented wave equation (Johnson and Dudgeon, 1993), which governs the propagation of sound through a viscous fluid and can be stated as

$$\nabla^2 p = \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{\mu}{c_0^2}\frac{\partial}{\partial t}(-\nabla^2 p), \tag{19}$$

where c_0 is the small signal sound speed, and $\mu = \lfloor 4\eta/3 + \eta_B + \kappa(\gamma_h - 1)/c_p\rfloor/\rho_0$ the collective thermoviscous coefficient, η and η_B the shear and bulk viscosity coefficients, respectively, ρ_0 the ambient density, κ thermal conductivity, γ_h ratio of specific heats, and c_p special heat at constant

pressure. Equation (19) describes both dispersion (waveform alternation with respect to frequency) and attenuation behaviors. Szabo (1994) pointed out that the low-frequency approximation of (19) leads to a square dependence of attenuation on frequency with constants

$$\alpha_0 = \mu/2c_0^3, \quad y = 2, \tag{20}$$

in terms of the power law (2). Szabo (1994) noted that the exponent y and the differential order of the lossy term in the generalized wave equation are related. Namely, the time derivative order of the lossy term is higher than y by 1. By analogy with this relationship, we generalize (19) via the space fractional Laplacian and intuitively get

$$\nabla^2 p = \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{2\alpha_0}{c_0^{1-y}}\frac{\partial}{\partial t}(-\nabla^2)^{y/2} p, \tag{21}$$

where $(-\nabla^2)^{y/2}$ is the fractional Laplace. When $y=2$, (21) is equivalent to the model equation (19). When $y=0$, (21) is reduced to the standard damped wave equation

$$\nabla^2 p = \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{2\alpha_0}{c_0}\frac{\partial p}{\partial t}, \tag{22}$$

which describes the frequency-independent attenuation (Szabo, 1994).

In order to verify that our intuitive fractional Laplacian wave equation (21) reflects the frequency power-law attenuation (2), the corresponding dispersion equation is derived below and analyzed. To facilitate the analysis without loss of generality, let us consider the 1D case of model equation (21)

$$\frac{\partial^2 p}{\partial z^2} = \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \frac{2\alpha_0}{c_0^{1-y}}\frac{\partial}{\partial t}\left(-\frac{\partial^2}{\partial z^2}\right)^{y/2} p. \tag{23}$$

The Fourier transforms of the time and space derivatives are given by

$$F_-\left\{\frac{\partial^2 p}{\partial z^2}\right\} = (ik)^2 P = -k^2 P, \tag{24}$$

$$F_-\left\{\left(-\frac{\partial^2}{\partial z^2}\right)^{y/2} p\right\} = (-(ik)^2)^{y/2} P = k^y P, \tag{25}$$

$$F_+\left\{\frac{\partial^2 p}{\partial t^2}\right\} = (-i\omega)^2 P = -\omega^2 P. \tag{26}$$

Applying the time and space Fourier transforms (24)–(26) to (23), we have the dispersion equation

$$k^2 - \omega^2/c_0^2 - i2\alpha_0\omega k^y/c_0^{1-y} = 0. \tag{27}$$

Since $k=\beta+i\alpha$ and $\beta=\omega/c$, it is straightforward to have

$$\beta^2 - \alpha^2 - \omega^2/c_0^2 + i2\alpha\beta - i2\alpha_0\omega\beta^y(1+i\alpha/\beta)^y/c_0^{1-y} = 0. \tag{28}$$

To move the analysis further, the Szabo conservative smallness approximation [Eqs. (17) and (18) of Szabo (1994)] is crucial, i.e.,

$$\alpha/\beta \approx \alpha_0|\omega|_{\lim}^y/\beta_0 = \alpha_0|\omega|_{\lim}^{y-1} c_0 \leqslant 0.1, \tag{29}$$

where $\omega_{\lim}$ is the frequency limit corresponding to 0.1. As discussed in Szabo (1994), the limit frequency range in terms of (29) is adequate enough to cover the frequency spectrum of practical interest in medical ultrasound applications. In terms of (29), (28) is then approximated by the binomial expansion as

$$\beta^2 - \alpha^2 - \omega^2/c_0^2 + i2\alpha\beta - i2\alpha_0\omega\beta^y/c_0^{1-y} + 2\alpha_0 y\omega\alpha\beta^{y-1}/c_0^{1-y} = 0. \tag{30}$$

Then, separating the real and imaginary parts of the above Eq. (30) produces

$$\alpha = \alpha_0\omega(\beta c_0)^{y-1}, \tag{31a}$$

$$\beta^2 = \omega^2/c_0^2 + (1-2y)\alpha^2. \tag{31b}$$

With the further help of the smallness approximation (29) and $\beta_0=\omega/c_0$, from (31a) we derive

$$\alpha \approx \alpha_0|\omega|^y, \tag{32}$$

It is noted that (31a) matches the power law (2). By now we have shown that the fractional Laplacian model (23) does have a power-law attenuation under the Szabo smallness approximation condition (29) corresponding to the time convolutional integral model (Szabo, 1994) and the time fractional models (Baglegy and Torvik, 1983). Szabo (1994) demonstrated his wave equation is causal indirectly by verifying its parabolic counterpart is causal. In the latter section III, we will show that the fractional Laplacian wave equation (21) can be reduced to the well-known parabolic anomalous diffusion equation. The causality of the latter is guaranteed (Hanyga, 2001c). We are therefore convinced that equation (21) is also causal. It is noteworthy that the fractional Laplacian is a positive definit operator (Gorenflo and Francesco, 1998), and the Duhamel's principle applies to the fractional Laplacian equations.

IV. NONLINEAR LOSSY MEDIA

Most nonlinear acoustic equations are only useful for describing lossy media with a quadratic dependence or independence on frequency (Szabo, 1993), and thus have limited practical utility. In this section, we are concerned with the extension of the previous linear fractional Laplacian modeling methodology to nonlinear media obeying the dissipative power law of arbitrary exponents. It is noted that the thermoviscous representation in the standard nonlinear acoustic PDE models, which characterizes the effects of absorption and dispersion, is mostly the same as in the corresponding linear models. Thus, Blackstock (1985) presented a straightforward strategy constructing the nonlinear anomalous attenuation model by simply replacing its attenuation term with that in the corresponding linear model, while keeping all other linear and nonlinear terms unchanged. The methodology is justified by a perturbation analysis (Blackstock, 1985). Following this strategy, Szabo (1993) extended his linear convolution integral modeling of arbitrary power law exponents (Szabo, 1994) to the Burgers, KZK, and Westervelt equations. By analogy with Blackstock (1985) and

Szabo (1993), we can generalize these nonlinear acoustic models by replacing their thermoviscous term with the fractional Laplacian lossy terms given in Eqs. (21) and (23).

It is stressed that the smallness approximation condition (29), crucial in the dispersion analysis of the preceding linear fractional Laplacian models, is also the foundation of the nonlinear acoustics modeling (Hamilton and Blackstock, 1998; Szabo, 1993). Thus, all derivations here are consistent.

The Burgers equation may be the best-known simple nonlinear acoustic model which describes the combined effects of nonlinearity and dissipation. The one-dimensional Burgers equation for plane progressive waves is stated as

$$\frac{\partial p}{\partial t}+Bp\frac{\partial p}{\partial z}-\varepsilon\frac{\partial^2 p}{\partial z^2}=0, \tag{33}$$

where B denotes the nonlinear coefficient (Sugimoto, 1991), and ε is a constant proportional to the coefficients of viscosity and heat conduction (Blackstock, 1985). It is known (Blackstock, 1985; Szabo, 1993) that the Burgers equation (31) describes lossy acoustic propagation of square frequency dependence. To extend the Burgers equation to accommodating power-law media of arbitrary exponent y, Blackstock (1985) suggests and verifies to some extent that only the third term of (33) needs to be modified, which involves the absorption, while keeping all others the same.

In terms of an approach detailed by Szabo (1994), the hyperbolic wave equation (23) can be approximated to the parabolic equation by removing the left-hand side term, namely

$$\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2}+\frac{2\alpha_0}{c_0^{1-y}}\frac{\partial}{\partial t}\left(-\frac{\partial^2}{\partial z^2}\right)^{y/2}p=0. \tag{34}$$

And then, integrating (34) with respect to time t and multiplying by c_0^2, we have

$$\frac{\partial p}{\partial t}+2\alpha_0 c_0^{1+y}\left(-\frac{\partial^2}{\partial z^2}\right)^{y/2}p=0. \tag{35}$$

The model (35) is a generalized diffusion equation and corresponds to the Burgers equation (33) without the nonlinear convection term. The analytical solution of Eq. (35) can be found in Hanyga (2001c). Applying the spatial Fourier transform (25) to (35), we have

$$\frac{dP}{dt}+2\alpha_0 c_0^{1+y}k^y P=0. \tag{36}$$

Thus, the transformation solution is

$$P(k,t)=Ce^{-2\alpha_0 c_0^{1+y}tk^y}, \tag{37}$$

where C depends on the initial condition. When $y=2$, (37) exhibits the normal frequency-squared diffusion. It is clear that Eq. (37) is a parabolic model originating from Eq. (23) while holding the capability describing arbitrary power (y) law attenuation.

By analogy with the generalizing methodology presented by Blackstock (1985) and Szabo (1993), the fractional Laplacian Burgers equation is presented below by simply adding the nonlinear convection term of Burgers equation (33) to Eq. (35), i.e.,

$$\frac{\partial p}{\partial t}+Bp\frac{\partial p}{\partial z}+2\alpha_0 c_0^{1+y}\left(-\frac{\partial^2}{\partial z^2}\right)^{y/2}p=0. \tag{38}$$

Note that the nonlinear term in (38) can be considered the source term in the sense of an inhomogeneous equation (Szabo, 1993).

The nonlinear equation model (38) belongs to the so-called fractal Burgers equations or the fractional advection–dispersion equation. A detailed analysis of such equations is given in Biler *et al.* (2001). In higher dimensional cases, (37) is restated as

$$\frac{\partial p}{\partial t}+Bp\cdot\nabla p+2\alpha_0 c_0^{1+y}(-\nabla^2)^{y/2}p=0, \tag{39}$$

where ∇p represents the pressure gradient vector, and the dot stands for a scalar product. Ochmann and Makarov (1993) also developed the time fractional derivative Burgers equation to describe the power-law absorptions with arbitrary y. By using the same strategy, Chen and Holm (2002b) also developed the fractional Laplacian KZK, Westervelt, general second-order approximation model, incompressible Naiver–Stokes, and Boussinesq shallow-water wave equation to incorporate arbitrary power-law frequency-dependent dissipations.

V. CONCLUDING REMARKS

Attenuation plays an essential part in many acoustics applications, for instance, the ultrasound second harmonic imaging and high-intensity focused ultrasound beam for therapeutic surgery. Compared with the Szabo's time convolutional integral model of the power-law attenuation, the present fractional Laplacian time-space model has a uniform and simpler expression. On the other hand, it is also not a simple task for the fractional time derivative models to obtain the initial conditions of the second-order derivative when $y>1$, since most physical systems only provide the zero- and first-order initial conditions. More importantly, most anomalous thermoviscous attenuations occur in spatially inhomogeneous environments (Henry and Wearne, 1999), notably biomaterials and geological random media, whose microgeometry largely have fractal dimension structures in space. The power-law formula (2) also shows that y is independent of frequency ω (time scale). It is therefore reasonable to think that y may in fact underlie the spatial fractal. For example, y varies with different human body tissues, which have different spatial microstructures. We thus conclude that a spatial representation of the dissipation via the fractional Laplacian is physically more valid than the fractional time derivative representations.

ACKNOWLEDGMENTS

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

APPENDIX: FEM DISCRETIZATION OF THE LINEAR FRACTIONAL LAPLACIAN LOSSY WAVE EQUATION

Let the FEM approximate discretization of a Laplacian operator be expressed as

$$-\nabla^2 p \Rightarrow K\mathbf{p}, \tag{A1}$$

where $\mathbf{p}$ represents the pressure value vector at the discrete nodes. Since the Laplacian is the positive definite operator, it is well known that its FEM discretization matrix K is also positive definite. The corresponding FEM formulation of the fractional Laplacian of $y/2$ order is then obtained by

$$(-\nabla^2)^{y/2} p \Rightarrow K^{y/2}\mathbf{p}. \tag{A2}$$

It is worth noting here that despite the fact that the FEM discretization matrix K is sparse, $K^{y/2}$ will be a full matrix underlying the non-local property of the fractional Laplacian, which models the global interactions in space. The FEM discretization of the fractional Laplacian wave equation model (21) is thus stated as

$$\mathbf{p}_{tt} + 2\alpha_0 c_0^{1+y} K^{y/2}\mathbf{p}_t + c^2 K\mathbf{p} = g(t), \tag{A3}$$

where the subscript t represents the temporal derivative, and $g(t)$ is due to the source term. It is obvious that (A3) readily takes into account frequency-dependent viscous effects for a multitude of frequency components (broadband signal) with empirical coefficients α_0 and y of the power-law attenuation.

For a little complicated fitting of measurement fitting, the empirical formula (2) of frequency dependent attenuation can be technially replaced by (He, 1998)

$$\alpha(\omega) = \alpha_1 + \alpha_0 |\omega|^y, \quad y \in [0,2], \tag{A4}$$

where α_1 is an empirical parameter. Thus, equation (A3) can be accordingly restated as

$$p_u + 2(\alpha_1 c_0 I + \alpha_0 c_0^{y+1} K^{y/2}) p_t + c_0^2 K p = g(t). \tag{A5}$$

When $y=2$, the semi-discrete model (A5) brings out the square frequency dependence and is reduced to the classical Rayleigh proportional damping model.

The temporal discretization of (A3) can easily be done via the standard finite difference time integrators. The major issue here is about computer resource requirements in the evaluation of $K^{y/2}$. The orthodox analytical approach for this task is costly singular value decomposition, i.e.,

$$K^{y/2} = \Phi^T \left(\sum \lambda_i^2 \right)^{y/2} \Phi, \tag{A6}$$

where Φ is the orthogonal matrix, the superscript T represents the matrix transpose, and Σ denotes a diagonal matrix with eigenvalues λ. The popular numerical methods for evaluating $K^{y/2}$ are the Schur decomposition, Padé approximation, and iterative method, which usually require $O(n^3)$ operations (Lu, 1998). In order to overcome such enormous cost, a parallel numerical model is under study. The results will be reported in a separate subsequent paper. As a matter of fact, the numerical solution of fractional Laplacian equations has not been researched well, and very few related reports are known to the authors.

It is also interesting to mention that the matrix power function of the fractional order [e.g., (A2)] can be considered a clear algebraic correspondence to the fractional calculus in analysis and the fractal in geometry. All three of these methodologies may consist of a complete set of mathematic apparatus in modeling, analyzing, simulating, and visualizing complex phenomena, where the traditional mathematic methods of integer order do not work well.

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

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

Biler, P., Karch, G., and Woyczynski, W. A. (**2001**). ''Asymptotics and high dimensional approximations for nonlinear pseudodifferential equations involving Levy generators,'' Demonstratio Math. **34**(2), 403–413.

Blackstock, D. T. (**1967**). ''Transient solution for sound radiated into a viscous fluid,'' J. Acoust. Soc. Am. **41**, 1312–1319.

Blackstock, D. T. (**1985**). ''Generalized Burgers equation for plane waves,'' J. Acoust. Soc. Am. **77**(6), 2050–2053.

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

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

Chen, W., Holm, S., Bounaim, A., Ødegård, Å., and Tveito, A. (**2002a**). ''Frequency decomposition time-domain model of broadband frequency-dependent absorption,'' The 9th Workshop of the Finite Element Method in Biomedical Engineering, Biomechanics and Related Fields, Ulm, Germany, pp. 40–48.

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

Diethelm, K. (**2000**). *Fractional Differential Equations, Theory and Numerical Treatment* (preprint).

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

Feller, W. (**1971**). *An Introduction to Probability Theory and its Applications*, 2nd ed. (Wiley, New York), Vol. 2.

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

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

Gorenflo, R., and Mainardi, F. (**1998**). ''Random walk models for space-fractional diffusion processes,'' Fract. Calc. Appl. Anal. **1**, 167–191.

Hamilton, M. F., and Blackstock, D. T., Editors (**1998**). *Nonlinear Acoustics* (Academic, New York).

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

Hanyga, A. (**2001a**). ''Multi-dimensional solutions of time-fractional diffusion-wave equations,'' Proc. R. Soc. London, Ser. A (in press).

Hanyga, A. (**2001b**). ''Fractional diffusion and wave equations,'' Conference of ''Modelli Matematici e Problemi Analitici per Materiali Speciali,'' Cortona.

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

He, P. (**1998**). ''Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**(1), 114–125.

Henry, B. I., and Wearne, S. L. (**1999**). ''Fractional reaction-diffusion,'' *Elsevier preprint*.

Jespersen, S. (**1999**). ''Anomalous diffusion,'' Progress Report of Institute of Physics and Astronomy, University of Aarhus.

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

Lighthill, J. (**1980**). *Waves in Fluids* (Cambridge University Press, Cambridge).

Lu, Y. Y. (**1998**). ''A Padé approximation method for square roots of sym-

metric positive definite matrices,'' Research Report, City University of Hong Kong.
Makris, N., and Constantinou, M. C. (**1991**). ''Fractional-derivative Maxwell model for viscous dampers,'' J. Struct. Eng. **117**(9), 2708–2724.
Mast, T. D., Souriau, L. P., Liu, D. L., Tabei, M., Nachman, A. I., and Wagg, R. C. (**2001**). ''A *k*-space method for large-scale models of wave propagation in tissue,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(2), 341–354.
Nachman, A. I., Smith, J., and Waag, R. C. (**1990**). ''An equation for acoustic propagation in inhomogeneous media with relaxation losses,'' J. Acoust. Soc. Am. **88**(3), 1584–1595.
Ochmann, M., and Makarov, S. (**1993**). ''Representation of the absorption of nonlinear waves by fractional derivative,'' J. Acoust. Soc. Am. **94**(6), 3392–3399.
Pierce, A. D. (**1989**). *Acoustics, an Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY).
Rusovici, R. (**1999**). ''Modeling of Shock Wave Propagation and Attenuation in Viscoelastic Structures,'' Ph.D. thesis, Virginia Polytechnic Institute and State University.
Samko, S. G., Kilbas, A. A., and Marichev, O. I. (**1987**). *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach Science, New York).
Sugimoto, N. (**1991**). ''Burgers equation with a fractional derivative; hereditary effects on nonlinear acoustic waves,'' J. Fluid Mech. **225**, 631–653.
Szabo, T. L. (**1993**). ''Time Domain Nonlinear Wave Equations for Lossy Media,'' in *Advances in Nonlinear Acoustics: Proceedings of 13th ISNA*, edited by H. Hobaek (World Scientific, Singapore), pp. 89–94.
Szabo, T. L. (**1994**). ''Time domain wave equations for lossy media obeying a frequency power law,'' J. Acoust. Soc. Am. **96**(1), 491–500.
Wismer, M., and Ludwig, R. (**1995**). ''An explicit numerical time domain formulation to simulate pulsed pressure waves in viscous fluids exhibiting arbitrary frequency power law attenuation,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(6), 1040–1049.
Wojcik, G., Mould, J., Lizzi, Jr., F., Abboud, N., Ostromogilsky, M., and Vaughan, D. (**1995**). ''Nonlinear modelling of therapeutic ultrasound,'' 1995 IEEE Ultrasonics Symposium Proceedings, 1617–1622.
Wojcik, G., Mould J., Carcione, L. M., Ostromogilsky, M., and Vaughan, D. (**1999**). ''Combined transducer and nonlinear tissue propagation simulations,'' 1999 International Mechanical Engineering Congress & Exposition Proceedings.
Yuan, X., Borup, D., and Wiskin, J. (**1999**). ''Simulation of acoustic wave propagation in dispersive media with relaxation losses by using FDTD method with PML absorbing boundary condition,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**(1), 14–23.
Zähle, M. (**1997**). ''Fractional differentiation in the self-affine case. V. The local degree of differentiability,'' Math. Nachr. **185**, 279–306.
Zaslavsky, G. M. (**1994**). ''Fractional kinetic equation for Hamiltonian chaos,'' Physica D **76**, 110–122.

Frequency–domain wave equation and its time–domain solutions in attenuating media

Nikolai V. Sushilov and Richard S. C. Cobbold[a)]
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, M5S 3G9, Canada

(Received 28 October 2003; revised 20 January 2004; accepted 21 January 2004)

Our purpose in this paper is to describe the wave propagation in media whose attenuation obeys a frequency power law. To achieve this, a frequency–domain wave equation was developed using previously derived causal dispersion relations. An inverse space and time Fourier transform of the solution to this algebraic equation results in a time–domain solution. It is shown that this solution satisfies the convolutional time–domain wave equation proposed by Szabo [J. Acoust. Soc. Am. **96**, 491–500 (1994)]. The form of the convolutional loss operator contained in this wave equation is obtained. Solutions representing the propagation of both plane sinusoidal and transient waves propagating in media with specific power law attenuation coefficients are investigated as special cases of our solution. Using our solution, comparisons are made for transient one-dimensional propagation in a medium whose attenuation is proportional to frequency with recently obtained numerical solutions of Szabo's equation. These show good agreement. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1675817]

PACS numbers: 43.20.Bi, 43.20.Hq, 43.80.Cs [JJM] Pages: 1431–1436

I. INTRODUCTION

It is well established that the attenuation of plane acoustic waves propagating in a wide variety of lossy media obey an empirical power law frequency dependence of the form

$$\alpha(\omega)=\alpha_0|\omega|^n, \tag{1}$$

where ω is angular frequency and α_0 and n are real positive constants. Experimental measurements on water over a very wide range of frequencies have shown that the value of n is almost exactly 2, which is the value predicted by the classical viscous loss theory. Most liquids have values in the range from 1 to 2 and, for many soft biological media in the diagnostic frequency range of 0.5–10 MHz, it is close to unity.[1]

To accurately predict how a given transient waveform will be propagated in such media, it is necessary to properly account for the effects of dispersion associated with the attenuation characteristics. For transients with a narrow bandwidth the effects of dispersion on the propagated waveform can generally be neglected. However, for wideband pulses propagating into media with relatively large attenuation, such as those used in B-mode ultrasound tissue imaging, the changes in the phase velocity with frequency, even though they may be relatively small, must be correctly accounted for. Measurements on the changes in phase velocity with frequency have been reported for at least 70 years, and much of this data, together with the results of more recent measurements, have been reported by Szabo.[2] He compared the experimental results with those predicted from absorption measurements based on causal dispersion relations.

A number of attempts have been made to obtain a causal wave equation in the spatial–time domain that can satisfactorily describe wave propagation in media obeying a power law. Based on the assumption of classical viscous losses and adiabatic conditions, Stokes[3] derived a wave equation from first principles. The plane-wave sinusoidal solution of this equation for small viscous losses yields $n=2$ and is accompanied by very small dispersion. Based on the assumption that the attenuation arises from a multiplicity of relaxation processes, Nachman *et al.*[4] derived a causal linearized wave equation for fluids from first principles. By a suitable choice of the relaxation processes, it is possible to obtain a good fit to a variety of the experimentally observed frequency-dependent attenuation characteristics.[5] However, for N relaxation mechanisms, the equation is of order $N+2$, making analytical solutions difficult to obtain. Moreover, evidence is not currently available to support the view that multiple physical relaxation mechanisms are responsible for absorption in media such as tissue in which n differs from 2.

Based on the space–time Fourier transform of the classical viscous loss wave equation, Szabo[6] proposed a frequency–domain wave equation for media whose attenuation obeyed a frequency power law. His equation is given by

$$\left[\kappa^2-\frac{\omega^2}{c_0^2}-2i\alpha_0(\omega/c_0)|\omega|^n\right]\Phi(\mathbf{k},\omega)=0, \tag{2}$$

where κ is a complex wave number, $\Phi(\mathbf{k},\omega)$ is the four-dimensional (space and time) Fourier transform of velocity potential, c_0 is a phase velocity constant, and $\mathbf{k}$ is a spatial transform vector. Szabo pointed out that this equation is noncausal. To make it causal, he added an extra term to the real part of κ and showed that both attenuation and the extra term are related through the Hilbert transform. Based on this modified causal frequency–domain wave equation, he derived a time–domain wave equation that involves a convolution integral that accounts for propagation in a power law attenuating medium and that guarantees causality. His equa-

a)Electronic mail: cobbold@ecf.utoronto.ca

tion, with an added source term, can be written as

$$\left[\nabla^2-\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}+L*\right]\phi(\mathbf{r},t)=\delta(\mathbf{r}-\mathbf{r}_s)s(t), \tag{3}$$

where L is a loss operator, $*$ denotes convolution, $\phi(\mathbf{r},t)$ is the velocity potential, $\delta(\mathbf{r}-\mathbf{r}_s)$ is the Dirac delta function at the source location, and $s(t)$ is the source function. Subsequently, Waters *et al.*[7] proposed a similar wave equation that Norton and Novarini[8] rewrote as

$$\left[\nabla^2-\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}+\frac{1}{c_0}\frac{\partial\Gamma(t)*}{\partial t}\right]\phi(\mathbf{r},t)=\delta(\mathbf{r}-\mathbf{r}_s)s(t), \tag{4}$$

where the propagation factor $\Gamma(t)$ is given by $L(t)=\Gamma(t)*\delta'(t)$. Through numerical methods Norton and Novarini[8] obtained solutions to (4) in a single spatial coordinate, and have shown that the effects of attenuation and dispersion are almost exactly accounted for by the causal propagation factor.

Chen and Holm[9] stated that appreciable difficulties arise in implementing the initial conditions of Szabo's equation due to an improper integral. They suggested a modification to his equation by replacing the convolution integral with a fractional derivative operator yielding

$$\left[\nabla^2-\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}-\frac{2\alpha_0}{c_0^{1-n}}\frac{\partial}{\partial t}(-\nabla^2)^{n/2}\right]\phi(\mathbf{r},t)=\delta(\mathbf{r}-\mathbf{r}_s)s(t). \tag{5}$$

In this equation, the term denoted by $(-\nabla^2)^{n/2}$ is a fractional Laplacian that can be defined through its inverse Fourier transform.[10,11] Chen and Holm[11] suggest that because the fractional Laplacian is a positive operator, the wave equation (5) is causal.

Although numerical methods can be used to obtain solutions to the above equations, to the best of our knowledge no analytical solutions have yet been obtained for either of these models for an arbitrary n in the range $0<n<3$. Based on the nonattenuating frequency–domain wave equation and the causal dispersion relations, we propose a new equation that accounts for causal attenuation and is valid over the above range for n. From this, a time–domain solution is obtained and some particular cases are investigated. In particular, we will compare our results for the one-dimensional spatial solution considered by Norton and Novarini,[8] in which they assume a transient source function and $n=1$.

II. FREQUENCY–DOMAIN WAVE EQUATION AND A TIME–DOMAIN SOLUTION

To account for a power law dependence of attenuation on frequency, it is convenient to make use of the complex wave number, defined by

$$K(\omega)=\frac{\omega}{c(\omega)}+i\alpha(\omega), \tag{6}$$

where the real and imaginary parts correspond to the effects of frequency-dependent dispersion and attenuation, respectively. Based on the Kramers–Kronig dispersion relations, Ben-Menahem and Singh[12] in their 1981 book on seismic wave propagation, derived a dispersion relation for an isotropic medium whose absorption is given by (1) and for which $0<n<1$. Making use of the generalized functions, Szabo[13] subsequently obtained a similar relation valid over a wider range and that is given by

$$\frac{1}{c(\omega)}=\frac{1}{c(\omega_0)}+\alpha_0\tan\left(\frac{\pi n}{2}\right)(|\omega|^{n-1}-|\omega_0|^{n-1}),$$
$$\text{for } 0<n<1 \quad \text{and} \quad 1<n<3, \tag{7}$$

where $c(\omega_0)=c_0$ is the phase velocity at a reference frequency of ω_0. Based on the differential Kramers–Kronig relations and just positive frequencies Waters *et al.*[14,7] also derived dispersion relations but without magnitude signs. One of these is the same as (7): the second was given by

$$\frac{1}{c(\omega)}=\frac{1}{c(\omega_0)}-\frac{2}{\pi}\alpha_0\ln\left|\frac{\omega}{\omega_0}\right|, \quad \text{for } n=1. \tag{8}$$

In fact, the logarithmic dependence of the phase velocity on frequency appears to have been first predicted by Futterman[15] in 1961. Recently, Waters *et al.*[16] pointed out that (7) reduces to (8) as n approaches unity.

To obtain a frequency–domain wave equation corresponding to the complex wave number as given by (6), we shall start with the frequency–domain lossless wave equation for the space and time Fourier transform of velocity potential, given by

$$[k^2-(\omega/c_0)^2]\Phi(\mathbf{k},\omega)=0, \tag{9}$$

where the spatial transform vector components are (k_x,k_y,k_z) and $k=\sqrt{k_x^2+k_y^2+k_z^2}$. Following the same procedure used in Ref. 17, we shall assume that the real wave number ω/c_0 in this equation can be replaced by the complex wave number given by (6) and that the resulting equation describes propagation in a lossy media characterized by different n. This frequency–domain wave equation can be expressed as

$$\left[k^2-\left(\frac{\omega}{c(\omega)}+i\alpha(\omega)\right)^2\right]\Phi(\mathbf{k},\omega)=0. \tag{10}$$

To obtain a time–domain solution, we shall make use of the technique developed by Donnelly and Ziolkowski.[18] First, we find the solution to the algebraic equation given by (10) and then make the inverse four-dimensional (space and time) Fourier transform. Based on the solution proposed by Donnelly and Ziolkowski for a real wave number, it is reasonable to assume that the same form of solution can be used for a complex wave number. Consequently, the solution of (10) can be written as

$$\Phi(\mathbf{k},\omega)=a(\omega,k)\,\delta[k_x-f(\omega)]\,\delta[k_y-f(\omega)]$$
$$\times\,\delta[k_z-f(\omega)], \tag{11}$$

where $a(\omega)$ and $f(\omega)$ are arbitrary functions. By taking the inverse space–time Fourier transform the time domain velocity potential can be written as

$$\phi(\mathbf{r},t)=\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)e^{i[f(\omega)(x+y+z)-\omega t]}\,d\omega. \tag{12}$$

Then, by substituting (11) into (10), it can be shown that $f(\omega)=[\omega/c(\omega)+i\alpha(\omega)]/\sqrt{3}$. This enables (12) to be expressed in the form

$$\phi(\mathbf{r},t)=\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)e^{i[\mathbf{K_r}(\omega)\cdot\mathbf{r}-\omega t]}\,d\omega, \tag{13}$$

where, in order to simplify the presentation, the vector $\mathbf{K_r}(\omega)$ has been introduced whose components are all equal to $[\omega/c(\omega)+i\alpha(\omega)]/\sqrt{3}=K(\omega)/\sqrt{3}$.

It can be seen that (13) is in the form of a Fourier transform amenable to computation by standard techniques. It consists of a superposition of plane waves of different frequencies and amplitudes each propagating in a dispersive medium in direction normal to the plane $x+y+z=\text{const}$. It should be noted that no approximations were made in deriving (13). Moreover, the function $a(\omega)$ represents the frequency distribution of a point source at $\mathbf{r}=0$. Thus, for a spatially distributed source, additional integrations are needed over the source distribution in order to find the total velocity potential.

It can be verified that (13) satisfies the following equation:

$$\nabla^2\phi(\mathbf{r},t)-\frac{1}{c_0^2}\frac{\partial^2\phi(\mathbf{r},t)}{\partial t^2}+\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)\times\left\{K(\omega)^2-\frac{\omega^2}{c_0^2}\right\}e^{i[\mathbf{K_r}(\omega)\cdot\mathbf{r}-\omega t]}\,d\omega=0. \tag{14}$$

If we let

$$L(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left\{K(\omega)^2-\frac{\omega^2}{c_0^2}\right\}e^{-i\omega t}\,d\omega, \tag{15}$$

then, from the convolution theorem, it can be shown that $L(t)*\phi(\mathbf{r},t)$, where $\phi(\mathbf{r},t)$, as given by (13), is exactly equal to the integral term in (14). Therefore, (13) represents the solution to the Szabo's wave equation (3) with the specific loss operator given by (15). However, (13) only satisfies the Chen and Holm equation, i.e., (5) when dispersion is absent, prompting us to suggest that their equation may be noncausal.

To illustrate the application of (13) a number of particular cases will now be described.

A. Monochromatic propagation

If we chose $a(\omega)=(2\pi)^4\phi_0\delta(\omega-\omega_0)$, where ϕ_0 is a constant, (13) reduces to

$$\phi(\mathbf{r},t)=\phi_0 e^{i(\omega_0/c_0+i\alpha_0|\omega_0|^n)(x+y+z)/\sqrt{3}-i\omega_0 t}. \tag{16}$$

This represents a monochromatic plane wave propagating in a direction normal to the plane $x+y+z=\text{const}$, in a medium whose attenuation is given by (1).

B. Propagation in a nonattenuating medium

In the absence of any attenuation, (13) reduces to

$$\phi(\mathbf{r},t)=\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)e^{i(\omega/c_0)[(x+y+z)/\sqrt{3}-c_0 t]}\,d\omega. \tag{17}$$

In fact, this describes a family of solutions to the time–domain wave equation describing wave propagation in homogeneous media without attenuation, i.e.,

$$\left[\nabla^2-\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right]\phi(\mathbf{r},t)=0. \tag{18}$$

Many particular solutions can be obtained from (17) depending on the choice for $a(\omega)$. For example, if $a(\omega)=(2\pi)^4\phi_0\delta(\omega-\omega_0)$, then (17) reduces to

$$\phi(\mathbf{r},t)=\phi_0 e^{i(\omega_0/c_0)[(x+y+z)/\sqrt{3}-c_0 t]}, \tag{19}$$

which describes a monochromatic plane wave propagating in media without attenuation. Alternatively, if we chose

$$a(\omega)=\frac{8\phi_0\pi^{7/2}}{\omega_0}e^{-\omega^2/4\omega_0^2}, \tag{20}$$

where ω_0 is a real positive constant, by making use of Eq. (3.323.2) in Ref. 19, we obtain from (17),

$$\phi(\mathbf{r},t)=\phi_0 e^{-\{(\omega_0/c_0)[(x+y+z)/\sqrt{3}-c_0 t]\}^2}. \tag{21}$$

This represents the propagation of a Gaussian-like wave whose beamwidth depends on ω_0.

C. Propagation in a medium with $n=2$

This particular case corresponds to a medium with classical viscous losses, and as will be seen, it provides some justification for the methodology used in obtaining the solution given by (13). According to (7), there should be no dispersion, i.e., $c(\omega)=c_0$. Consequently, the complex velocity potential, as given by (13), simplifies to

$$\phi(\mathbf{r},t)=\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)e^{i\omega(p_a-t+q_a\omega)}\,d\omega, \tag{22}$$

where $p_a=(x+y+z)/(c_0\sqrt{3})$ and $q_a=i\alpha_0(x+y+z)/\sqrt{3}$. Provided the attenuation is small, i.e., $\omega\alpha_0 c_0\ll 1$, it can be shown that (22) satisfies the time–domain wave equation for the media with classical viscous losses (for example, see Ref. 20), i.e.,

$$\left[\left(1+\tau_c\frac{\partial}{\partial t}\right)\nabla^2-\frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right]\phi(\mathbf{r},t)=0, \tag{23}$$

where $\tau_c=2\alpha_0 c_0=(\rho_0 c_0^2)^{-1/2}[\mu_B+(4/3)\mu]$, ρ_0 is the equilibrium density of the medium, and μ and μ_B are the shear and bulk viscosity, respectively.

For the particular case in which $a(\omega)$ is given by (20), with the help of Eq. (3.323.2) in Ref. 19, it can be shown that

$$\phi(\mathbf{r},t)=\frac{\phi_0}{\sqrt{1+4\alpha_0\omega_0 c_0(\mathbf{k}_0\cdot\mathbf{r})}}\times\exp\left\{-\frac{(\mathbf{k}_0\cdot\mathbf{r}-\omega_0 t)^2}{1+4\alpha_0\omega_0 c_0(\mathbf{k}_0\cdot\mathbf{r})}\right\}, \tag{24}$$

where $\mathbf{k}_0\cdot\mathbf{r}=(\omega_0/c_0)(x+y+z)/\sqrt{3}$. This represents an extension of the nonattenuating Gaussian-like solution given by (21) for a medium with classical viscous losses.

D. Attenuating medium, $n=1$

When the attenuation varies linearly with frequency ($n=1$), the solution as given by (13) can be obtained in an analytic form for the monochromatic case. If we let $a(\omega)=(2\pi)^4\phi_0\delta(\omega-\omega_0)$, in (13) and make use of (8) the velocity potential can be expressed as

$$\phi(\mathbf{r},t)=\phi_0 e^{i(\omega_0/c_0)((x+y+z)/\sqrt{3}-c_0 t)-\alpha_0|\omega_0|(x+y+z)/\sqrt{3}}. \tag{25}$$

This is a particular case of (16). Other solutions in cylindrical symmetrical coordinates for $n=1$ have been obtained and investigated for both infinite and finite aperture in our previous paper.[17]

E. Attenuating medium, $0<n<1$ and $1<n<3$

For the noninteger case, the solution given by (13) reduces to

$$\phi(\mathbf{r},t)=\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)\exp\{i\omega(p_b-t)+q_b|\omega|^n[i\,\mathrm{sign}(\omega)\tan(n\pi/2)-1]\}d\omega, \tag{26}$$

where $p_b=[c_0^{-1}-\alpha_0|\omega_0|^{n-1}\tan(n\pi/2)](x+y+z)/\sqrt{3}$ and $q_b=\alpha_0(x+y+z)/\sqrt{3}$.

If the attenuation is small such that $|q_b|\omega|^n[i\,\mathrm{sign}(\omega)\tan(n\pi/2)-1]|\ll 1$, then the second term in the exponent of (26) can be approximated by $1+q_b|\omega|^n[i\,\mathrm{sign}(\omega)\tan(n\pi/2)-1]$, enabling (26) to be written as

$$\phi(\mathbf{r},t)\approx\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)e^{i\omega(p_b-t)}d\omega+\frac{1}{(2\pi)^4}q_b\int_{-\infty}^{\infty}a(\omega)[i\,\mathrm{sign}(\omega)\times\tan(n\pi/2)-1]|\omega|^n e^{i\omega(p_b-t)}\,d\omega. \tag{27}$$

Let us assume that $a(\omega)$ is chosen to be the same as for cases **B** and **C**, i.e., by (20). By splitting the second integral into two integrals from 0 to ∞ and then making use of Eq. (3.462.1) in Ref. 19, the velocity potential can be obtained as

$$\phi(\mathbf{r},t)=\phi_0 e^{-\omega_0^2(p_b-t)^2}+\phi_0\frac{(\sqrt{2}\omega_0)^n}{\sqrt{2\pi}}\Gamma(n+1)\times q_b e^{(1/2)\omega_0^2(p_b-t)^2}\{[1+i\tan(n\pi/2)]\times D_{-1-n}[i\sqrt{2}\omega_0(p_b-t)]+[1-i\tan(n\pi/2)]D_{-1-n}[-i\sqrt{2}\omega_0(p_b-t)]\}. \tag{28}$$

In this equation, $D_{-1-n}(\cdot)$ is the parabolic cylinder function[19] and $\Gamma(\cdot)$ is the gamma function. This solution represents a sum of Gaussian-like pulses—one with a constant amplitude ϕ_0, the other with the space and time-modulated amplitude. In the particular case of $n=2$ (28) reduces to

$$\phi(\mathbf{r},t)=\phi_0 e^{-\omega_0^2(p_b-t)^2}+2q_b\omega_0^2\phi_0\times D_2[-\sqrt{2}\omega_0(p_b-t)]e^{-(1/2)\omega_0^2(p_b-t)^2}. \tag{29}$$

The same expression can be obtained directly from (22) by assuming $|q_b\omega^2|\ll 1$ and by using Eqs. (3.323.2) and (3.462.3) in Ref. 19.

III. COMPARISION WITH NUMERICAL RESULTS

As mentioned earlier, Norton and Novarini[8] numerically calculated a solution to Szabo's time–domain wave equation (3) in one dimension for a point source function given by

$$s(0,t)=t\exp(-\beta t^2), \tag{30}$$

where β is a constant that governs the time interval between the negative and positive peaks of $s(t)$. For $n=1$, they calculated the pressure at two different locations, z_1 and z_2, separated a distance d. The spectrum was then found enabling the attenuation and phase velocity to be obtained as functions of frequency.

For the same source function, the one-dimensional (1-D) form of (13) can be used to obtain the velocity potential and its spectrum at an arbitrary z location. Starting from the 1-D form of the frequency–domain wave equation and following the same steps as used to obtain (13), it can be shown that

$$\phi(z,t)=\frac{1}{(2\pi)}\int_{-\infty}^{\infty}a(\omega)e^{i\{[\omega/c(\omega)+i\alpha(\omega)]z-\omega t\}}\,d\omega.$$

Consequently, the signal spectrum can be written as

$$\Psi(z,\omega)=a(\omega)e^{izK(\omega)}, \tag{31}$$

so that $a(\omega)=\Psi(0,\omega)$, i.e., the Fourier transform of the source function $s(0,t)$. Thus, from (30)

$$a(\omega)=\int_{-\infty}^{\infty}te^{-\beta t^2}e^{i\omega t}\,dt=\frac{i\omega}{2\beta}\sqrt{\frac{\pi}{\beta}}e^{-\omega^2/4\beta}. \tag{32}$$

For a medium with linear frequency dependence of attenuation and corresponding dispersion ($n=1$), the spectrum of the propagating pulse can be obtained from (31), with the help of (32), (6), and (8) as

$$\Psi(z,\omega)=\frac{i\omega}{2\beta}\sqrt{\frac{\pi}{\beta}}e^{-\omega^2/4\beta}\times e^{i\omega z/c_0[1-(2/\pi)\alpha_0 c_0\ln|\omega/\omega_0|]}e^{-\alpha_0\omega z}. \tag{33}$$

An examination of this equation shows that for the assumed 1-D source propagating into a medium with $n=1$, the attenuation and dispersion correspond to (1) and (8), respectively.

We have used (33) to calculate the signal spectra for propagation in a medium with $n=1$ at $z=25$ m from the source, and the results are presented in Fig. 1(a) (dotted curve). The parameter values used were those taken from the paper by Norton and Novarini[8] together with the following additional and corrected information provided by the authors: $c_0=1372$ m/s was taken at the reference frequency, $\omega_0=2\pi\times 28\,567$ Hz, $\alpha_0=2.707\times10^{-6}$ Np/(m rad/s). The constant β for the source function, as used in (30), was set equal to $3.15\times10^6\ \mathrm{s}^{-2}$. The solid line corresponds to the spectrum when the effects of dispersion are neglected, i.e., by setting the logarithmic term in (33) to zero. By numerical inverse Fourier transform of these two spectra, the dispersive and nondispersive (time) wave forms were obtained, as

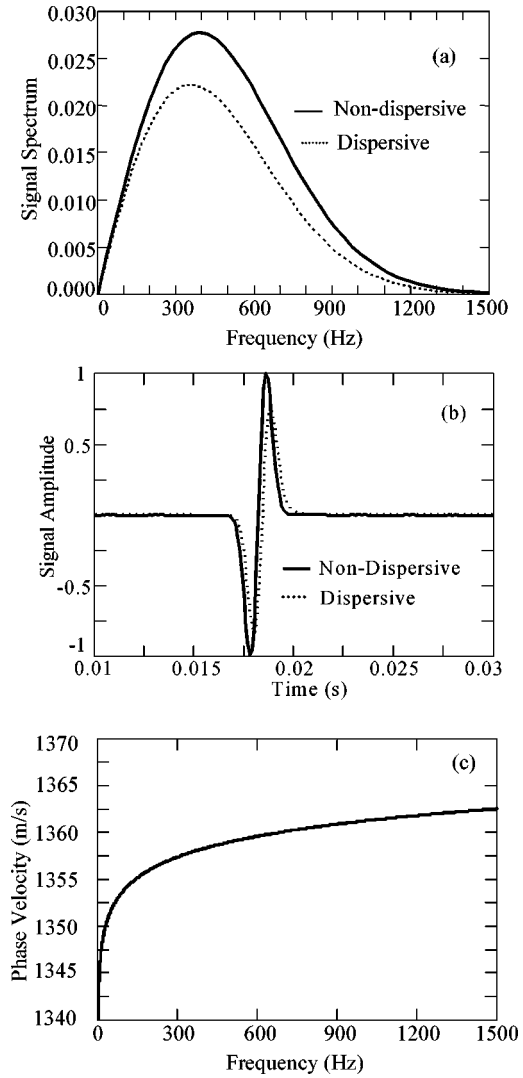


FIG. 1. Frequency spectra, wave forms, and phase velocity at a distance of $z=25$ m from a transient point source whose wave form is given by (30), for propagation in a medium with $n=1$. The parameter values are reference frequency, $f_0=28\,567$ Hz, speed of sound at the reference frequency, $c_0=1372$ m/s, $\alpha_0=2.707\times10^{-6}$ Np/(m rad/s) and $\beta=3.15\times10^6$ s^{-2}. (a) Signal spectra for propagation in nondispersive and dispersive media, as calculated from (33). (b) Time–domain waveforms as calculated from (a) using a numerical inverse Fourier transform. (c) Phase velocity versus frequency, as calculated from (8).

shown in Fig. 1(b). A comparison of Figs. 1(a) and 1(b) with the corresponding figures numerically calculated by Norton and Novarini [their Figs. 2(a) and 1(b), respectively] shows excellent agreement. In Fig. 1(c), we have plotted the frequency dependence of the phase velocity $c(\omega)$ as calculated from (8) using the same parameters given above. A comparison of this with that obtained by Norton and Novarini shows good agreement.

IV. DISCUSSION AND CONCLUSIONS

We have made use of the fact that knowledge of the time–domain wave equation is unnecessary to obtain a complete description of linear wave propagation. In media with attenuation obeying the frequency power law, the frequency-domain algebraic equation is sufficient. By taking an inverse four-dimensional space and time Fourier transform of the solution to this algebraic equation, it is possible to obtain the time–domain solution that describes the wave propagation. Based on the dispersion relations for the speed of sound and medium attenuation derived by Szabo,[13] we obtained a time–domain solution of the frequency–domain wave equation. If the loss operator is given by (15) this solution satisfies the time–domain wave equation introduced by Szabo[6] and represents a superposition of infinite numbers of plane waves of different frequencies and amplitudes. A number of particular cases was investigated to illustrate the application of the solution to plane wave propagation. In one particular case—when the attenuation is relatively small and proportional to the square of frequency, coresponding to classical viscous losses, our solution reduces to the classical result.

ACKNOWLEDGMENTS

We much appreciate the helpful comments and suggestions of Adam Weathermon, and wish to thank the National Science and Engineering Research Council of Canada and the Canadian Institute of Health Research for financial support. In addition, comments and suggestions by one of the reviewers have been very helpful.

[1] F. A. Duck, *Physical properties of tissue. A Comprehensive Reference Book* (Academic, London, 1990), Chap. 4, pp. 108–110.
[2] T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[3] G. G. Stokes, "On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids," Trans. Cambridge Philos. Soc. **8**, 287–319 (1845).
[4] A. J. Nachman, J. F. Smith III, and R. C. Waag, "An equation for acoustic propagation in inhomogeneous media with relaxation losses," J. Acoust. Soc. Am. **88**, 1584–1595 (1990).
[5] A. P. Berkhoff, J. M. Thijssen, and R. J. F. Homan, "Simulation of ultrasonic imaging with linear arrays in causal absorptive media," Ultrasound Med. Biol. **22**, 245–259 (1996).
[6] T. L. Szabo, "Time domain wave equation for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[7] K. R. Waters, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "On a time–domain representation of the Kramers–Kronig dispersion relations," J. Acoust. Soc. Am. **108**, 2114–2119 (2000).
[8] G. V. Norton and J. C. Novarini, "Including dispersion and attenuation directly in the time domain for wave propagation in isotropic media," J. Acoust. Soc. Am. **113**, 3024–3031 (2003).
[9] W. Chen and S. Holm, "Modified Szabo's wave equation models for lossy media obeying frequency power law," J. Acoust. Soc. Am. **114**, 2570–2574 (2003).
[10] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach, New York, 1987).
[11] W. Chen and S. Holm, "Fractional Laplacian, time–space models for linear and nonlinear media exhibiting arbitrary frequency power law dependency," J. Acoust. Soc. Am. **115**, 1424–1430 (2004).
[12] A. Ben-Menahem and S. J. Singh, *Seismic Waves and Sources* (Springer-Verlag, New York, 1981), pp. 893–897.
[13] T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[14] K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers–Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[15] W. I. Futterman, "Dispersive body waves," J. Geophys. Res. **67**, 5279–5291 (1962).
[16] K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, "Differential

forms of the Kramers–Kronig relations,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 68–76 (2003).
[17] N. V. Sushilov and R. S. C. Cobbold, ''Wave propagation in media whose attenuation is proportional to frequency,'' Wave Motion **38**, 207–219 (2003).
[18] R. Donnelly and R. Zilokowski, ''A method for constructing solutions of homogeneous partial differential equations: localized waves,'' Proc. R. Soc. London, Ser. A **437**, 673–692 (1992).
[19] I. S. Gradstein and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 4th ed. (Academic, New York, 1965).
[20] L. J. Ziomek, *Fundamentals of Acoustic Field Theory and Space–Time Signal Processing* (CRC Press, Boca Raton, 1995).

Modal expansions for sound propagation in the nocturnal boundary layer

Roger Waxler[a)]
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

(Received 28 June 2003; revised 9 November 2003; accepted for publication 15 December 2003)

A modal model is developed for the propagation of sound over an impedance ground plane in a stratified atmosphere which is downward refracting near the ground but upward refracting at high altitudes. The sound's interaction with the ground is modeled by an impedance with both real and imaginary parts so that the ground is lossy as well as compliant. Such sound speed profiles are typical of the atmospheric boundary layer at night and, together with the ground impedance, have been used extensively to model ground to ground sound propagation in the nocturnal environment. Applications range from community noise and bioacoustics to meteorology. The downward refraction near the ground causes the propagation to be ducted, suggesting that the long range propagation is modal in nature. This duct is, however, leaky due to the upward refraction at high altitudes. The modal model presented here accounts for both the attenuation of sound by the ground as well as the leaky nature of the duct.
[DOI: 10.1121/1.1646137]

PACS numbers: 43.20.Bi, 43.28.Js, 43.28.Fp [MSH] Pages: 1437–1448

I. INTRODUCTION

Sound propagation outdoors at night is characterized by the duct that forms in the first few hundred meters of the atmosphere. Typically, due to cooling of the atmosphere by the earth, the temperature of the air increases as one moves upward to an inversion point, generally at 100 to 300 m elevation, above which it decreases slowly for the next few kilometers. This temperature inversion causes a part of the sound field to propagate along the ground in a ducted fashion.

It should be noted that the sound speed depends on humidity as well as temperature and that atmospheric winds can have a profound effect on propagation.[2,3] For low angle propagation, however, these effects can be modeled by using the effective sound speed[3,4]

$$c = \sqrt{\gamma RT(1+0.511q)} + \hat{\mathbf{n}}\cdot\mathbf{v}.$$

Here c is the sound speed, γ is the ratio of the constant pressure to constant volume specific heats for dry air, R is the gas constant for dry air, q is the ratio of the mass of water vapor to dry air, $\hat{\mathbf{n}}$ is the horizontal propagation direction, and $\mathbf{v}$ is the wind speed vector.

Some nocturnal profiles for the effective sound speed, determined by direct measurements of temperature, humidity, and wind speed, can be found in Fig. 16 of Ref. 5. The distinguishing feature of such sound speed profiles is the presence of an inversion: the sound speed increases with increasing altitude up to the inversion height, typically found between 100 to 300 m elevation, above which it slowly decreases. The result is an atmosphere which is downward refracting near the ground, causing a duct, but upward refracting at high altitudes.

Below the inversion the rate of increase of the sound speed can be large. Care will be taken that the techniques developed here do not depend on any assumptions about the magnitude of the sound speed gradient near the ground. At high altitudes, however, the decrease of the sound speed is gradual. It will be assumed here that at sufficiently high altitudes the rate of decrease of the acoustic wavelength is small. Some model sound speed profiles are presented in Fig. 1.

Propagation in such sound speed profiles, together with a ground impedance condition[6] with both imaginary (compliant) and real (resistive) parts to model the sound's interaction with the ground, has been widely studied with intended applications including physical meteorology,[5] community noise modeling[7,4] and bioacoustics.[8] In modeling the sound propagation the parabolic equation and horizontal wave number integration techniques have been preferred.[4,9] These are robust numerical schemes which work well regardless of the sound speed profile, however, they are ''black box'' routines which obscure the fact that the physics of ducted sound propagation is modal.

This paper continues the study begun in Ref. 10 of modal expansions for sound propagation in horizontally stratified downward refracting atmospheres over lossy and compliant ground planes. The main advantage of modal expansions over other methods is that they give a transparent decomposition of the sound field into physically different, independently propagating parts. In addition, once the modes have been computed, they are essentially analytical solutions requiring negligible computation times.

In Ref. 10 sound speed profiles which become constant at sufficiently high altitudes were studied. A modal expansion similar to those commonly used in ocean acoustics[9] was developed and a technique for finding the corresponding horizontal wave numbers and attenuation coefficients was presented. It was shown that for asymptotically constant

a)Electronic mail: rwax@olemiss.edu

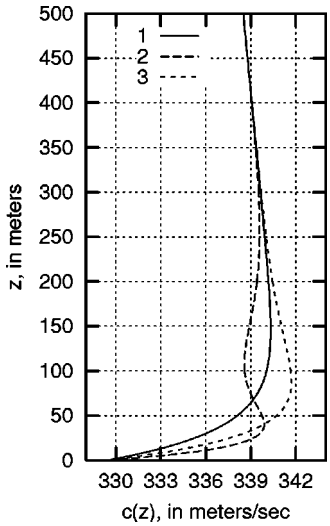


FIG. 1. Some model sound speed profiles. Profile 1 is used in the numerical simulations shown here.

sound speed profiles there is a clean separation into an upwardly propagating part and a ducted part, the ducted part of the sound field completely described by a sum over a small number of vertical modes whose amplitudes decrease exponentially with height. The upwardly propagating part is described by an integral whose computation can be intensive, but which is negligible at low angles and long ranges and thus need not be computed. The sum over modes will be referred to as the modal sum while the remaining integral will be referred to as the continuum integral.

In the asymptotically upward refracting case considered here the duct is imperfect in that sound trapped in the duct can leak upward to the region in which the sound speed profile is upward refracting and escape into the upper atmosphere. It follows that the separation into ducted and upwardly propagating parts cannot hold with complete rigor. In fact, in the lossless case, in which the real part of the ground impedance is zero, there is no modal sum at all.[11] Rather the entire sound field is contained in the continuum integral, the ducted behavior manifesting itself in sharp peaks in the integrand.

In the lossy case, in which the real part of the ground impedance is not zero, it will be seen that there is a modal sum. This is somewhat astonishing since there must be some mechanism through which the ducted part of the sound field leaks into the upper atmosphere. The vertical modes will be seen to extend very high into the atmosphere, although typically with negligible amplitude, before they begin to decrease exponentially with height. The physical picture is that the power flux into the ground due to the attenuation of sound by the ground is greater than the power flux upward into the atmosphere due to the asymptotic upward refraction. Thus, although the ducted part of the field does leak into the upper atmosphere, the attenuation by interaction with the ground is sufficient to prevent the field from extending upward without bound. A similar phenomenon has been recently found in an ocean acoustics propagation model in which some of the classical leaky modes[9] become trapped modes if the bulk attenuation in the acoustic medium is sufficient.[12]

It will be seen that for most frequencies the modal sum is sufficient to describe the ducted part of the sound field. There are, however, certain narrow frequency bands in which the continuum integral also contributes to the ducted part of the sound field. These are the frequency bands in which the number of terms in the modal sum changes. At these transition frequencies a mode emerges from the continuous spectrum, accompanied by a sharp peak in the integrand in the continuum integral, reminiscent of the peaks one finds in the lossless case. It will be shown that at such frequencies the continuum integral has a significant contribution which can, at low altitudes, be estimated by a discrete sum over terms of the same form as the terms in the modal sum, obtained from the residue theorem of complex analysis. These additional terms will be referred to as quasi-modes.

The resulting sum of modes and possible quasi-modes results in a model for sound propagation in the nocturnal boundary layer valid up to ranges of 10 km or more in a frequency range from a few Hz, the frequencies at which modes begin to appear, up to the kHz range, where the number of modes begins to become prohibitively large, making the method unwieldy. The method is particularly powerful at low frequencies for which the number of modes is small. Throughout, the method is illustrated by explicit computations for a specific model sound speed profile. These computations are restricted to frequencies below 100 Hz so that the number of modes is small and the required numerical techniques are straightforward.

II. FORMULATION OF THE PROBLEM

Consider the propagation of sound over a plane. Let $\mathbf{x}_H$ denote the horizontal coordinates (those parallel to the plane) and let z denote the vertical coordinate (the height above the plane). Let t denote time and let $\omega=2\pi f$ be the angular frequency corresponding to f cycles per second.

The sound speed, $c(z)$, is assumed to depend only on z. Further, it is assumed to be downward refracting near the ground but to be upward refracting for large z. Explicitly, $c(z)$ is assumed to have an absolute maximum at some height. Above or below the absolute maximum there may be local maxima, but generally speaking the sound speed is increasing at the ground and decreasing for sufficiently large z. Let H be the height above which $c(z)$ is strictly decreasing. The rate of decrease of $c(z)$ at high altitudes is assumed to be slow in the sense that if $z \gg H$, then

$$0 < -c'(z) \ll \omega. \tag{1}$$

This means that the relative rate of change of $c(z)$, $c'(z)/c(z)$, divided by 2π, is small compared with an acoustic wavelength f/c.

Recall the model sound speed profiles plotted in Fig. 1. In the numerical results presented in this paper sound speed profile 1 is used. The analysis presented here applies equally well to sound speed profiles of the form of profiles 2 and 3, however, the numerical analysis depends on the results of Ref. 10 which would require a slight generalization to handle sound speed profiles with multiple maxima like profile 2.

Introduce the notation

$$k(z)=\frac{\omega}{c(z)} \tag{2}$$

for the height dependent wave number. Note that if $z \gg H$, then $0<k'(z) \ll k(z)^2$. In addition it is assumed $k'(z)$ is bounded, which is to say that, as $z \to \infty$, $k(z)$ grow no faster than linearly in z. Finally, the mean density, $\rho_0(z)$, is assumed to vary so slowly with z that $c\rho_0'/\omega\rho_0 \ll 1$, so that the relative change in density over an acoustic wavelength is negligible.

The acoustic pressure, $P(\mathbf{x}_H,z,t)$, is assumed to be time harmonic

$$P(\mathbf{x}_H,z,t)=\mathrm{Re}\,\hat{P}(\mathbf{x}_H,z)e^{-i\omega t}, \tag{3}$$

where the amplitude $\hat{P}(\mathbf{x}_H,z)$ satisfies the Helmholtz equation

$$(\nabla^2+k(z)^2)\hat{P}(\mathbf{x}_H,z)=0. \tag{4}$$

To model the effect of the ground an impedance condition is assumed at $z=0$,

$$\left.\frac{\partial \hat{P}(\mathbf{x}_H,z)}{\partial z}\right|_{z=0} = -\mathcal{C}(\omega)\hat{P}(\mathbf{x}_H,0), \tag{5}$$

where $\mathcal{C}(\omega)$ is related to the impedance $Z(\omega)$ through

$$Z(\omega)=\frac{i\omega\rho_0(0)}{\mathcal{C}(\omega)}.$$

In general, both the real and imaginary parts of $Z(\omega)$ are nonzero. In the explicit computations presented here $\mathcal{C}(\omega)$ is chosen as in Eq. (24) of Ref. 10.

Equation (3) will be solved by separating variables and using an eigenfunction expansion[10,13] in the vertical coordinate. The vertical equation may be written

$$\left(\frac{d^2}{dz^2}+k(z)^2-\eta\right)\psi(z)=0, \tag{6}$$

where η is the separation parameter. The impedance condition (5) leads to

$$\psi'(0)=-\mathcal{C}\psi(0). \tag{7}$$

Consider the large z behavior of the solutions ψ of the differential equation (6). If z is large enough so that (1) is valid, then the WKB approximation[9,14] may be used. For real valued η one obtains the asymptotic form

$$\psi(z)\approx(k(z)^2-\eta)^{-1/4}\left(Ae^{i\int_{z_0}^{z}\sqrt{k(z')^2-\eta}\,dz'} + Be^{-i\int_{z_0}^{z}\sqrt{k(z')^2-\eta}\,dz'}\right) \tag{8}$$

as $z\to\infty$; here z_0 is large enough so that $k(z)^2>\eta$ for $z \geqslant z_0$ insuring that no turning points are encountered in the integration and A and B are constants. Choose the square roots in (8) so that they have non-negative real parts. Note that for real values of η the solutions are asymptotically oscillatory and remain bounded as $z\to\infty$.

If $\mathrm{Im}\,\eta\neq 0$, then the WKB approximation provides two types of solutions: those that grow exponentially as z increases,

$$\psi(z)\approx A\begin{cases}(k(z)^2-\eta)^{-1/4}e^{i\int_{z_0}^{z}\sqrt{k(z')^2-\eta}\,dz'} & \text{if Im } \eta>0,\\ (k(z)^2-\eta)^{-1/4}e^{-i\int_{z_0}^{z}\sqrt{k(z')^2-\eta}\,dz'} & \text{if Im } \eta<0,\end{cases}$$

and those that grow exponentially as z decreases, which is to say that they decrease exponentially as z increases. These later solutions play a prominent role in the analysis presented here and will require their own notation: let, for $\mathrm{Im}\,\eta\neq 0$, $\psi_+(\eta,z)$ be the solution of (6) which has the large z asymptotic form

$$\psi_+(\eta,z)\approx\begin{cases}(k(z)^2-\eta)^{-1/4}e^{-i\int_{z_0}^{z}\sqrt{k(z')^2-\eta}\,dz'} & \text{if Im } \eta>0,\\ (k(z)^2-\eta)^{-1/4}e^{i\int_{z_0}^{z}\sqrt{k(z')^2-\eta}\,dz'} & \text{if Im } \eta<0,\end{cases} \tag{9}$$

for some z_0 sufficiently large so that $k(z)^2>\mathrm{Re}\,\eta$ for $z \geqslant z_0$.

The solutions $\psi_+(\eta,z)$ do not generally satisfy the boundary condition (7). However, there may be a discrete set of values of η, $\{\eta_1,\eta_2,\ldots,\eta_N\}$, for which the solution $\psi_+(\eta,z)$ given above does satisfy (7).[13] The values $\{\eta_1,\eta_2,\ldots,\eta_N\}$ will be referred to as the point spectrum. Here N may be any whole number from 0 to ∞.

It follows from the large z asymptotics (8) that the η_j must have nonzero imaginary parts. In particular, if $\mathcal{C}$ were real valued, then the boundary condition (7) would make the eigenvalue problem (6) self-adjoint forcing the point spectrum to be real valued.[13] It follows that if $\mathcal{C}$ is real valued, then there is no point spectrum and $N=0$.

Given an η_j in the point spectrum, the mode corresponding to η_j is given by

$$\phi_j(z)=\left(\int_0^\infty \psi_+(\eta_j,z)^2 dz\right)^{-1/2}\psi_+(\eta_j,z). \tag{10}$$

This normalization is chosen so that

$$\int_0^\infty \phi_j(z)^2\,dz=1. \tag{11}$$

Note that in Eq. (11) the function squared, $\phi_j(z)^2$, rather than the complex modulus squared, $|\psi(z)|^2$, is integrated. This is the normalization appropriate to the eigenfunction expansion which arises from the non-self-adjoint boundary condition (7).[10,13] In practice it can be difficult to compute the integral in (10). An alternative formula, (A8), is presented in the Appendix.

The decrease in the magnitude of $\psi_+(\eta_j,z)$ as $z\to\infty$ is not quite exponential. From the asymptotic form (9) one obtains

$$\psi_+(\eta,z) \approx (k(z)^2-\eta)^{-1/4} \times e^{-i\int_{z_0}^{z}\sqrt{k(z')^2-\mathrm{Re}\,\eta}\,dz' - (1/2)\mathrm{Im}\,\eta\int_{z_0}^{z}[1/\sqrt{k(z')^2-\mathrm{Re}\,\eta}]\,dz'}.$$

Note that $\sqrt{k(z)^2-\mathrm{Re}\,\eta}$ is an increasing function of z for large z. It follows that the first integral in the exponential function above grows more rapidly with z than a linear function of z,

$$\lim_{z\to\infty}\frac{1}{z}\int_{z_0}^{z}\sqrt{k(z')^2-\mathrm{Re}\,\eta}\,dz'=\infty,$$

while the second integral grows more slowly than a linear function of z,

$$\lim_{z\to\infty}\frac{1}{z}\int_{z_0}^{z}\frac{1}{\sqrt{k(z')^2-\mathrm{Re}\,\eta}}dz'=0$$

[for example, if $k(z)^2$ is linear for large z, then this last integral$\sim$const$\cdot z^{1/2}$]. Thus, $\psi_+(\eta,z)$ asymptotically oscillates more rapidly than a sinusoid with an amplitude that decreases more slowly that an exponential. In addition, to reach this asymptotic form, z must be large enough so that $k(z)^2\gg\mathrm{Re}\,\eta$. Thus the asymptotic decrease in the magnitude of $\psi_+(\eta_j,z)$ does not generally begin until quite high up in the atmosphere.

The set of values of η for which the solutions of (6) subject to (7) are not exponentially decreasing but do remain bounded as $z\to\infty$ is called the continuous spectrum. It follows from (8) that the continuous spectrum is the entire real line. The associated solutions of (6) subject to (7) are called continuum modes. Let $\psi_\epsilon(z)$ be the continuum mode associated to the real number ϵ. If appropriately normalized, then one has biorthonormality in the sense that

$$\int_0^\infty \psi_\epsilon(z)\psi_{\epsilon'}(z)\,dz=\delta(\epsilon-\epsilon'),$$

$$\int_0^\infty \phi_j(z)\phi_k(z)\,dz=\delta_{jk},$$

and

$$\int_0^\infty \psi_\epsilon(z)\phi_j(z)\,dz=0,$$

and completeness[15,13]

$$\delta(z-z')=\sum_{j=1}^{N}\phi_j(z)\phi_j(z')+\int_{-\infty}^{\infty}\psi_\epsilon(z)\psi_\epsilon(z')\,d\epsilon. \tag{12}$$

In the Appendix it is shown how one may normalize $\psi_\epsilon(z)$. Let $\psi_-(\mathcal{C},\eta,z)$ be the solution of (6) satisfying the boundary condition (7) normalized by $\psi_-(\mathcal{C},\eta,0)=1$ and let $\psi_+(\eta,z)$ be as above. Let ϵ be a real number and let

$$F_\pm(\mathcal{C},\epsilon)=\psi_+'(\epsilon\mp i0^+,0)+\mathcal{C}\psi_+(\epsilon\mp i0^+,0). \tag{13}$$

Here the prime indicates the derivative with respect to z and $\epsilon\pm i0^+$ indicates the limiting value as $\pm\mathrm{Im}\,\epsilon$ approaches zero while remaining positive. One then has

$$\psi_\epsilon(z)=\frac{1}{\sqrt{\pi F_-(\mathcal{C},\epsilon)F_+(\mathcal{C},\epsilon)}}\psi_-(\mathcal{C},\epsilon,z). \tag{14}$$

One can now express the pressure amplitude $\hat{P}(\mathbf{x}_H,z)$ in an eigenfunction expansion with respect to the eigenfunctions ϕ_j and ψ_ϵ. One obtains

$$\hat{P}(\mathbf{x}_H,z)=\sum_{j=1}^{N}p_j(\mathbf{x}_H)\phi_j(z)+\int_{-\infty}^{\infty}p(\epsilon,\mathbf{x}_H)\psi_\epsilon(z)\,d\epsilon \tag{15}$$

with expansion coefficients

$$p_j(\mathbf{x}_H)=\int_0^\infty\phi_j(z)\hat{P}(\mathbf{x}_H,z)\,dz \tag{16}$$

and

$$p(\epsilon,\mathbf{x}_H)=\int_0^\infty\psi_\epsilon(z)\hat{P}(\mathbf{x}_H,z)\,dz. \tag{17}$$

The expansion coefficients satisfy the two dimensional Helmholtz equations

$$(\nabla_H^2+\eta_j)p_j(\mathbf{x}_H)=0 \tag{18}$$

and

$$(\nabla_H^2+\epsilon)p(\epsilon,\mathbf{x}_H)=0 \tag{19}$$

for $\mathbf{x}_H$ in source free regions of space. Here

$$\nabla_H=\begin{pmatrix}\dfrac{\partial}{\partial x}\\[2ex]\dfrac{\partial}{\partial y}\end{pmatrix}.$$

Given appropriate boundary conditions or sources, (18) and (19) are straightforward to solve.

III. FINDING THE POINT SPECTRUM

Once the spectrum is known, producing the associated modes is a straightforward numerical exercise. In principle, given η, one solves (6) subject to (7). For the continuum modes this procedure works well since the continuous spectrum is known *a priori* and consists of real values of η for which the solutions of (6) are bounded.

The chief difficulty is in finding the point spectrum. Further, once the point spectrum is known, solving (6) subject to (7) is numerically unstable for complex valued η because of the exponential behavior of solutions to (6). It is preferable to solve (6) subject to the condition that

$$\frac{\psi'(z)}{\psi(z)}\approx\begin{cases}-i\sqrt{k(z)^2-\eta} & \text{if Im }\eta>0,\\ i\sqrt{k(z)^2-\eta} & \text{if Im }\eta<0,\end{cases} \tag{20}$$

for large z. This forces the solution to have the asymptotic form (9) and is numerically stable.[9,14]

A. Solving the eigenvalue equation

Consider the eigenvalue problem given by (6) and (7). Recall the solution, $\psi_+(\eta,z)$, of (6) chosen to have the asymptotic form (9) for large z. If $\psi_+(\eta,z)$ satisfies (7), then η is in the point spectrum and $\psi_+(\eta,z)$ is, up to normalization, the eigenfunction corresponding to η. Dividing both sides of (7) by $\psi(0)$ it follows that the point spectrum is precisely the set of all η with Im $\eta\neq0$ for which

$$\frac{\psi_+'(\eta,0)}{\psi_+(\eta,0)}+\mathcal{C}=0. \tag{21}$$

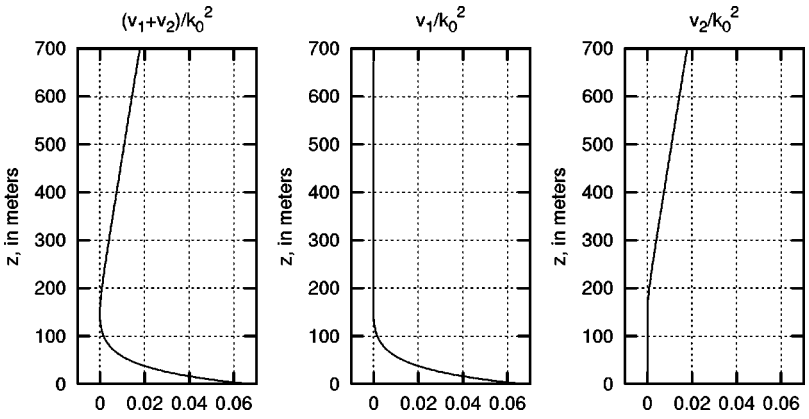


FIG. 2. The decomposition of the wave number squared for soundspeed profile 1 into asymptotically constant and asymptotically upward refracting parts.

To specify $\psi_+(\eta,z)$ completely it is necessary to fix z_0 in (9). However, with the eigenvalue condition written in the form (21) it becomes clear that specifying z_0 is not necessary since only the logarithmic derivative of $\psi_+(\eta,z)$ is required. One need only find any solution of (6) which satisfies (20) for large z, track its logarithmic derivative back to $z=0$, and then substitute into (21). In principle one could use this procedure to search for the roots of (21).

In practice such procedures are prone to failure.[9] Not only is it difficult to search blindly for roots in the complex plane, but such a search is further complicated here by the exponential growth and decay of the solutions to (6) which can make numerical procedures unstable. To produce stable and robust algorithms for finding the point spectra one must have some means to analytically estimate the roots with sufficient accuracy to allow standard numerical routines (the Newton–Rapheson method, for example) to converge rapidly from the estimate to the exact value.

The technique used in Ref. 10 to find the point spectrum, that of treating $\mathrm{Im}\,\mathcal{C}$ as a perturbation, cannot be applied here since if $\mathrm{Im}\,\mathcal{C}=0$, then, as pointed out in the previous section, there is no point spectrum to perturb. Instead decompose $k(z)^2$ as follows:

$$k(z)^2=k_0^2+v_1(z)+v_2(z), \tag{22}$$

where

$$k_0=k(H)$$

is the minimum value of k,

$$v_1(z)=\begin{cases} k(z)^2-k_0^2 & \text{if } z\leqslant H, \\ 0 & \text{if } z\geqslant H, \end{cases}$$

and

$$v_2(z)=\begin{cases} 0 & \text{if } z\leqslant H, \\ k(z)^2-k_0^2 & \text{if } z\geqslant H. \end{cases}$$

In Fig. 2 the decomposition (22) is depicted for the vertical sound speed profile 1 from Fig. 1.

Since $k(z)$ is slowly varying for large z it follows that $v_2(z)$ is as well. Write Eq. (6) in the form

$$\left(\frac{d^2}{dz^2}+v_1(z)+v_2(z)-(\eta-k_0^2)\right)\psi(z)=0.$$

It will be shown that the point spectrum can be obtained by treating $v_2(z)$ as a perturbation.

Let

$$\mathcal{C}=\mathcal{A}+i\mathcal{B}$$

be the decomposition of $\mathcal{C}$ into its real and imaginary parts. Let L_0 denote the differential operator obtained from

$$L_0=\frac{d^2}{dz^2}+v_1$$

by imposing (7) with $\mathcal{B}=0$,

$$\psi'(0)=-\mathcal{A}\psi(0),$$

let L_1 denote the differential operator obtained from

$$L_1=\frac{d^2}{dz^2}+v_1,$$

by imposing (7) with $\mathcal{B}\neq 0$, and let L denote the differential operator obtained from

$$L=\frac{d^2}{dz^2}+v_1+v_2$$

by again imposing (7) with $\mathcal{B}\neq 0$.

The operator L_0 is self-adjoint and numerous techniques exist for finding its spectrum.[9,16,17] The point spectrum of L_0 consists of M real, positive eigenvalues $\eta_1^{(0)},\ldots,\eta_M^{(0)}$. The continuous spectrum of L_0 is the negative real half-line, $(-\infty, 0)$. The techniques of Ref. 10 can then be applied to produce the point spectrum of L_1 from that of L_0; the continuous spectrum is unchanged. One finds that the point spectrum of L_1 consists again of M values, $\eta_1^{(1)},\ldots,\eta_M^{(1)}$, all with positive real and imaginary parts. Let $\phi_j^{(1)}(z)$ be the eigenfunction [normalized as in (11)] of L_1 corresponding to $\eta_j^{(1)}$.

It remains to be shown that the point spectrum of $L=L_1+v_2$ can be obtained from that of L_1. Note that $v_2(z)$ is a rather singular perturbation of L_1. In particular, the continuous spectra of L and L_1 are drastically different. As discussed in the previous section the continuous spectrum of L is the entire real line, $(-\infty,\infty)$, regardless of how slowly v_2 increases with height. Further, in the self-adjoint case, $\mathcal{B}$

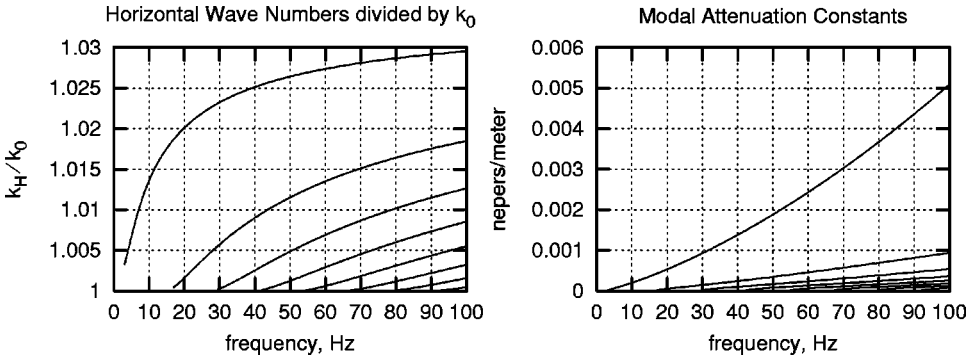


FIG. 3. The modal dispersion relations. Plotted are the horizontal wave numbers relative to k_0, $\mathrm{Re}\sqrt{k_0^2+\eta_j}/k_0$, and the attenuation constants, $\mathrm{Im}\sqrt{k_0^2+\eta_j}$.

$=0$, it was seen that L_1 has point spectrum while L does not. This self-adjoint case has been studied in detail in a classic paper by Titchmarsh.[11] What remains of the point spectrum in this case are poles, called resonances, in the analytic continuation in ϵ of the continuum eigenfunctions. It is clear from (13) that the resonances are solutions of the eigenvalue equation (21) for which the corresponding solution of (6) is not decreasing with height; it follows that the imaginary part of the resonance must be negative. Titchmarsh shows that the real part of a resonance is well approximated by formal perturbation theory[18] while the imaginary part is smaller than any power of $v_2(z)$ (which is to say that it is exponentially small in the size of the perturbation).

An approach similar to that of Titchmarsh can be taken here in this non-self-adjoint context. Perturbation theory can be adapted in a straightforward manner to the biorthogonality relevant to the current situation. One finds that, to leading order,

$$\eta_j = \eta_j^{(1)} + \int_0^\infty \phi_j^{(1)}(z)^2 v_2(z)\, dz + \cdots . \tag{23}$$

In practice the approximation (23) has proven to be sufficiently accurate to provide an initial point from which standard numerical routines converge rapidly to the solution of the eigenvalue equation (21).

As discussed in the previous section, the mode ϕ_j corresponding to η_j extends much higher into the atmosphere than $\phi_j^{(1)}$, although in most cases with very small amplitude. In particular ϕ_j has slightly less contact with the ground than $\phi_j^{(1)}$ and thus is attenuated slightly less by the ground. One thus expects that

$$\mathrm{Im}\ \eta_j < \mathrm{Im}\ \eta_j^{(1)},$$

just as in the self-adjoint case. This expectation is borne out by the numerical results presented below. Indeed, it is possible that the perturbation causes the imaginary part of η_j to be negative; as pointed out above, in the self-adjoint case all of the eigenvalues aquire a negative imaginary part. If the imaginary part becomes negative, then the corresponding solution of (6) and (7) grows exponentially as $z\to\infty$ and the eigenvalue becomes a resonance and ceases to be in the point spectrum. It follows that $N\leqslant M$: the number of eigenvalues in the point spectrum of L is less than or equal to the number of eigenvalues in the point spectrum of L_0, the operator relevant to a strictly downward refracting atmosphere with no attenuation at the ground.

Note that for a given v_2 the magnitude of the perturbation is determined by the vertical extent of the unperturbed mode. If the mode is concentrated near the ground and has insignificant magnitude for $z>H$, then the perturbation will be insignificant as well. If the magnitude of the mode decreases only slowly with height and is still appreciable for $z>H$, then the perturbation can be significant, perhaps sufficiently so that η_j is not in the point spectrum. Since one has

$$|\phi_j^{(1)}(z)| \sim \mathrm{const}\, e^{-\sqrt{\eta_j^{(0)}-k_0^2}z}$$

for large z the perturbation is significant only if $\eta_j^{(0)}$, the corresponding eigenvalue of L_0, is sufficiently close to k_0^2.

B. Numerical methods

The form (21) for the eigenvalue equation is not the most convenient for numerical analysis since it has poles at the values of η for which the Dirichlet problem $\psi_+(\eta,0)=0$ has a solution. An eigenvalue equation must be found which has no poles in the region of interest. A convenient choice is

$$\frac{\psi_+'(\eta,0)+\mathcal{C}\psi_+(\eta,0)}{\psi_+'(\eta,0)-i\psi_+(\eta,0)}=0. \tag{24}$$

This form still has poles, at the zeros of $\psi_+'(\eta,0)-i\psi_+(\eta,0)$, but they are in the lower half-plane sufficiently distant from the real line. To compute $\psi_+(\eta,0)$ fix some value of z_0 which is large enough for the asymptotic form (20) to be accurate and then solve (6), using any convenient numerical solver, subject to (20) imposed as a boundary condition at z_0.

In practice the estimate (23) is found to be quite accurate, except for those occasions in which $\eta_j^{(1)}$ is too close to k_0^2. Even in these cases, however, it has been found that (23) is sufficiently accurate to provide a starting point from which the Newton–Rapheson method applied to (24) converges rapidly, even when $\mathrm{Im}\ \eta_j$ turns out to be negative. In the examples worked in this paper the bisection method was employed to find the point spectrum of L_0. The perturbative estimate of Ref. 10 was then used as a starting point for a Newton–Rapheson solver applied to the eigenvalue equation for L_1.

The dispersion obtained from applying these methods to sound speed profile 1 from Fig. 1 is plotted in Fig. 3. Plotted are the real and imaginary parts of $k_{Hj}=\sqrt{k_0^2+\eta_j}$ as a function of frequency. The real part is normalized by k_0. The real

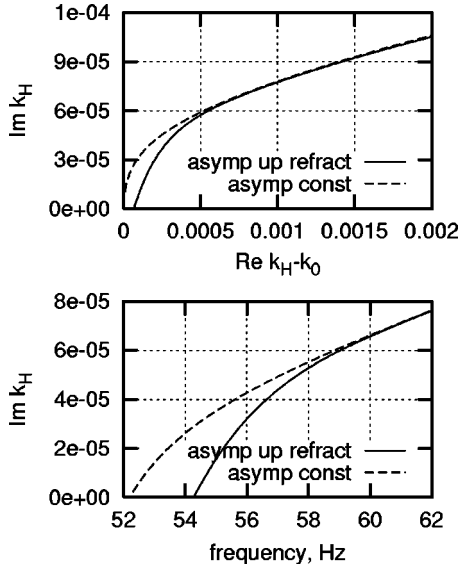


FIG. 4. The modal dispersion relations for the fifth mode in the asymptotically constant case and the asymptotically upward refracting case are compared near a transition frequency. Plotted are the attenuation constants, $\mathrm{Im}\sqrt{k_0^2+\eta_j}$, versus the horizontal wave numbers, $\mathrm{Re}\sqrt{k_0^2+\eta_j}-k_0$, and then the attenuation constants versus frequency.

part is referred to as the horizontal wave number and the imaginary part as the attenuation constant. In Fig. 4 the dispersion of mode 5 for the asymptotically constant approximation and the full asymptotically upward refracting profile are compared. In the upper plot $\sqrt{k_0^2+\eta_5}-k_0$ and $\sqrt{k_0^2+\eta_5^{(1)}}-k_0$ are tracked in the complex plane. In the lower plot the attenuation coefficients $\mathrm{Im}\sqrt{k_0^2=\eta_5}$ and $\mathrm{Im}\sqrt{k_0^2+\eta_5^{(1)}}$ are plotted as functions of frequency. Note that the asymptotically constant approximation can have more modes than in the asymptotically upward refracting case. In both cases the mode vanishes at the frequency at which the attenuation constant becomes zero; however, in the asymptotically constant approximation the horizontal wave number approaches k_0 while in the asymptotically upward refracting case the horizontal wave number remains slightly larger than k_0.

Once the eigenvalues are found, the corresponding modes are obtained by computing $\psi_+(\eta_j,z)$ and then normalizing according to (11). Computing the normalization from (10) can be difficult, however, since $\psi_+(\eta,z)$ can decrease very slowly as $z\to\infty$. Instead (A8) is used.

In Fig. 5 the modes at 50 Hz for sound speed profile 1 are plotted. They are labeled by their corresponding horizontal wave number and attenuation coefficient. In Fig. 6 the same plots are found, but for 41.6 Hz. This is a transition frequency at which the number of modes increases from three to four (see Fig. 3). Accordingly, the fourth mode has horizontal wave number very close to k_0.

Note that the fourth mode at 41.6 Hz extends high up into the atmosphere oscillating before it eventually slowly decays in accordance with (9). In fact, all of the modes have this behavior, although except in the exceptional case of a transition frequency at which the number of modes changes, the oscillations at high altitude have negligably small amplitude. This is demonstrated in Fig. 7 in which the fourth mode

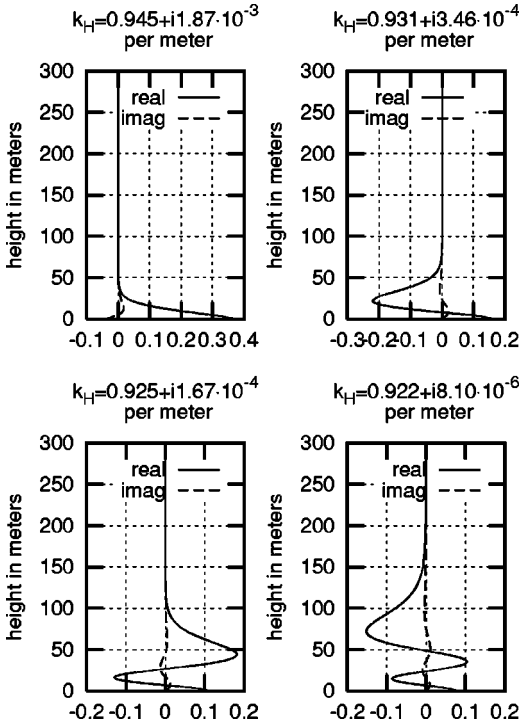


FIG. 5. The modes at 50 Hz for sound speed profile 1.

at 50 Hz is plotted at high altitudes. Physically this long tail at high amplitudes represents the leaking of acoustic energy from the duct up into the atmosphere. The eventual decay of the mode with height reflects the fact that the rate at which energy is lost into the ground is much greater than the rate at which it can leak into the upper atmosphere.

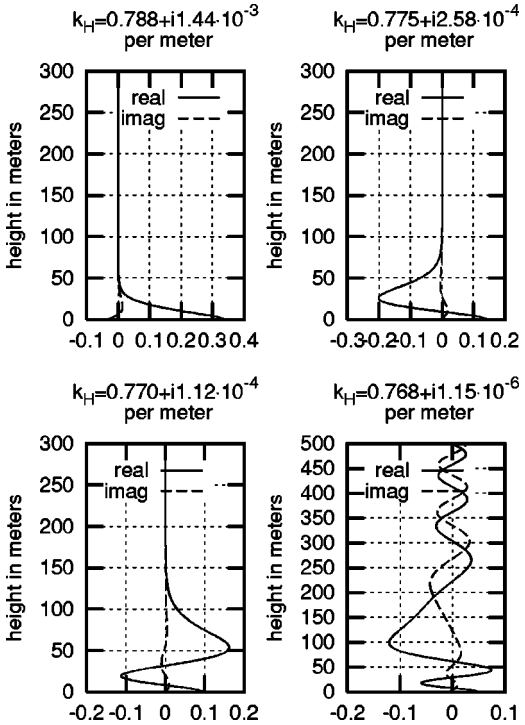


FIG. 6. The modes at the transition frequency 41.6 Hz for sound speed profile 1.

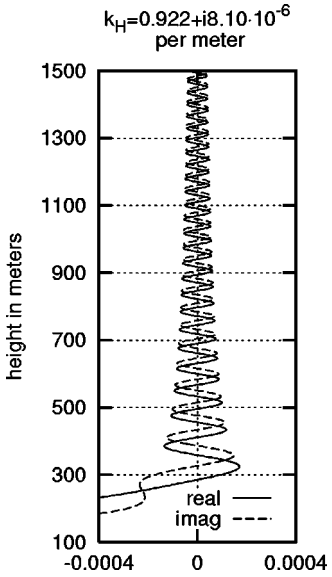


FIG. 7. The high altitude behavior of mode 4 at 50 Hz for sound speed profile 1.

IV. THE MODES OF THE CONTINUOUS SPECTRUM

Consider the modes of the continuous spectrum $\psi_\epsilon(z)$ given by (14). The dependence of $\psi_\epsilon(z)$ on ϵ may be divided into three general types according to whether ϵ is somewhat less than k_0^2, comparable to k_0^2, or somewhat greater that k_0^2. In Fig. 8 continuum modes at 50 Hz for sound speed profile 1 displaying the three types of behavior are plotted.

When ϵ is somewhat less than k_0^2 the behavior of the modes is oscillatory over the entire domain. These are the modes which contain that part of the sound field which propagates from low altitudes up into the atmosphere or the reverse. Note that in this case the magnitude of a mode does not change dramatically with increasing height.

When ϵ is somewhat greater than k_0^2 there are one or more turning points at height $\dot{z}$ for which $k(z)^2=\epsilon$. For sound speed profile 1 there is always one turning point, to be denoted $\tau_1(\epsilon)$ for purposes of discussion, above the sound speed inversion. If $\epsilon<k(0)^2$, there is a second, to be denoted $\tau_0(\epsilon)$, below the sound speed inversion. If $\epsilon\geqslant k(0)^2$, let $\tau_0(\epsilon)=0$. Above the turning points, for $z>\tau_1(\epsilon)$, the modes are oscillatory with behavior given, for sufficiently large z, by (8). Between the two turning points, for $\tau_0(\epsilon)<z<\tau_1(\epsilon)$, the modes grow exponentially with height.[14] It follows that in this case the modes have negligible amplitude near the ground. Their amplitude increases rapidly with height up to the vicinity of the turning point $\tau_1(\epsilon)$, above which they oscillate with gradually decreasing amplitude. These modes contain that part of the sound field radiated from a source above the sound speed inversion which is refracted upwards away from the ground, not penetrating below the inversion.

When ϵ is comparable to k_0^2 two factors can complicate the behavior of the continuum modes: there is the coalescence of the two turning points discussed in the last paragraph and there is the proximity of the point spectrum. At high altitudes the continuum modes will be oscillatory. At low altitudes the coalescence of the turning points means that no simple and general behavior can be expected. The behavior of the modes reflects the details of the low altitude form of the sound speed profile.

The proximity of the point spectrum produces sharp peaks in the low altitude behavior of those modes for which ϵ is in the vicinity of the real part of an eigenvalue in the point spectrum. This is seen clearly in (14): the eigenvalue condition (21) is precisely the condition that $F_+(\mathcal{C},\eta)=0$ so that ψ_ϵ has poles where the eigenvalue condition is satisfied. The effect of these poles is tempered by the fact that if ϵ is sufficiently greater than k_0^2, then the corresponding mode is still exponentially decreasing with decreasing z below the inversion.

This behavior is demonstrated in Figs. 9(a) and 9(b) in which the continuum mode amplitude on the ground, $|\psi_\epsilon(0)|$, for sound speed profile 1 is plotted as a function of ϵ for 50, 52, 54, 56, and 58 Hz. Both $|\psi_\epsilon(0)|$ and $\log|\psi_\epsilon(0)|$ are plotted in Figs. 9(a) and 9(b), respectively. Note that the peaks in $|\psi_\epsilon(0)|$ are sharpest for transition frequencies near which the number of modes changes. Indeed these are the frequencies for which one of the eigenvalues in the point spectrum can be arbitrarily close to the continuous spectrum. The plot of $\log|\psi_\epsilon(0)|$ shows that all of the eigenvalues in the point spectrum show up as peaks in the continuous spectrum; however, unless their real parts are sufficiently close to k_0^2, the exponential behavior of the mode overwhelms the peak.

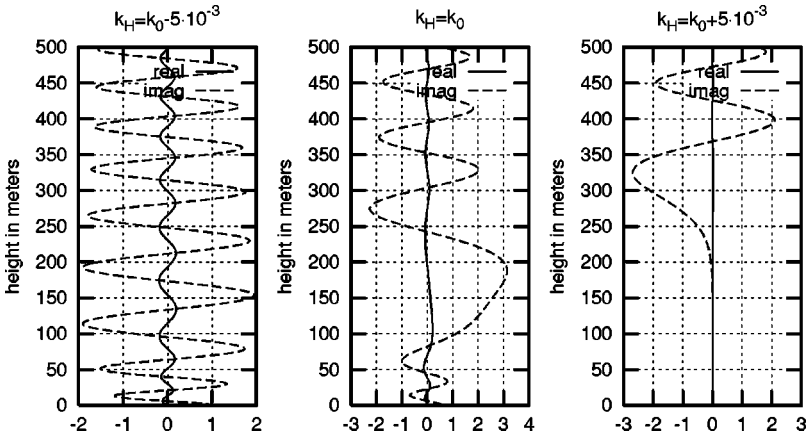


FIG. 8. Continuum modes at 50 Hz for sound speed profile 1. The three types of modal behavior, for $k_H=\sqrt{\epsilon}<k_0, =k_0, >k_0$, are displayed: below k_0 the modes are oscillatory with little variation in amplitude, near k_0 the modes are strongly influenced by the duct, above k_0 the modes have negligible amplitude near the ground.

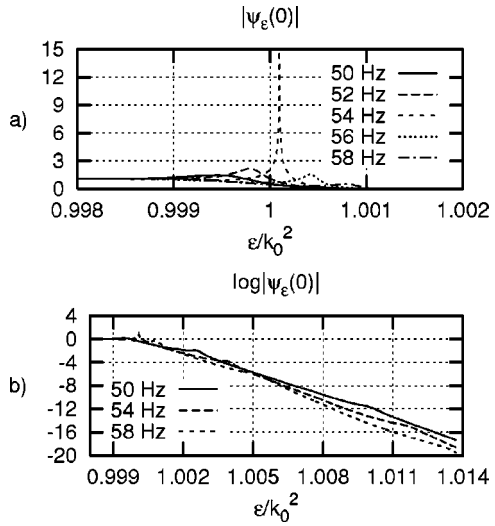


FIG. 9. Continuum mode amplitudes on the ground, (a) $|\psi_\epsilon(0)|$ and (b) $\log|\psi_\epsilon(0)|$, for sound speed profile 1 for several frequencies.

To extract the contribution to integrals such as those that appear in (15) from these peaks it is useful to know the residues of the poles that cause them. Consider

$$\psi_\epsilon(z_1)\psi_\epsilon(z_2) = -\frac{1}{\pi F_+(\epsilon)F_-(\epsilon)}\psi_-(\mathcal{C},\epsilon,z_1)\psi_-(\mathcal{C},\epsilon,z_2). \tag{25}$$

Let $\eta_1,\ldots,\eta_M$ be the solutions of the eigenvalue condition (7) which were obtained in Sec. III by perturbing off of the asymptotically constant case. Recall that, depending on the signs of their imaginary parts, not all of the η_j are necessarily in the point spectrum and thus not all represent true ducted modes. $F_\pm(\epsilon)$ are defined for $\pm\,\mathrm{Im}\,\epsilon\geqslant 0$, but have straightforward analytic continuations into regions with $\pm\,\mathrm{Im}\,\epsilon<0$. One then has $F_+(\eta_j)=0$, regardless of the sign of Im η_j. Thus, $\psi_\epsilon(z_0)\psi_\epsilon(z)$ has poles, as a function of ϵ, at the η_j. When $\mathrm{Im}\,\epsilon_j>0$, so that η_j is in fact in the point spectrum, then $-2\pi i$ times the residues at these poles gives the modes, ϕ_j, corresponding to η_j:

$$\phi_j(z_1)\phi_j(z_2) = \frac{2i}{F_-(\eta_j)F'_+(\eta_j)}\psi_-(\mathcal{C},\eta_j,z_1)\psi_-(\mathcal{C},\eta_j,z_2) \tag{26}$$

[this follows from the fact that $\psi_+(\epsilon+i0^+,0)\psi_-(\mathcal{C},\eta_j,z) = \psi_+(\epsilon+i0^+,z)$ and from (A4) and (A5); see the Appendix]. When $\mathrm{Im}\,\epsilon_j<0$, introduce the notion of the quasi-mode ϕ_j defined by (26).

V. LONG RANGE PROPAGATION FROM A POINT SOURCE

Consider the sound field produced by a point source at $\mathbf{x}_H=\mathbf{0}$, $z=z_0$. The solution to (4) for a point source at $\mathbf{x}_H=\mathbf{0}$, $z=z_0$ is proportional to the Green's function

$$G(\omega,\mathbf{x}_H,z_0,z) = \sum_{j=1}^{N} G_H(\eta_j,\mathbf{x}_H)\phi_j(z_0)\phi_j(z) + \int_{-\infty}^{\infty} G_H(\epsilon,\mathbf{x}_H)\psi_\epsilon(z_0)\psi_\epsilon(z)\,d\epsilon, \tag{27}$$

where $G_H(\epsilon,\mathbf{x}_H)$ is the free space Green's function satisfying

$$(\nabla_H^2+\epsilon)G_H(\epsilon,\mathbf{x}_H)=\delta(\mathbf{x}_H).$$

In polar coordinates r, θ, G_H is independent of θ. One finds[9,19,20]

$$G_H(\epsilon,\mathbf{x}_H) = -\frac{i}{4}H_0^{(1)}(\sqrt{\epsilon}r), \tag{28}$$

where H_0 is the Hankel function.[14]

Once the modes ϕ_j have been computed, computing the contribution to (27) from the modal sum is straightforward. The contribution from the integral over the continuum modes must be computed numerically. This computation can be intensive. However, if one is interested in the sound field near the ground and far from the source, then it can be expected that the integral over the continuous spectrum is insignificant[10,9] so that one need only compute the modal sum.

For the sound speed profiles considered here it will be shown that it is indeed the case for most frequencies that the continuous spectrum is insignificant for low altitudes at long ranges. The exceptions are those narrow bands of transition frequencies in which the number of modes changes. As discussed in the previous section, for these frequencies the emerging mode manifests itself as a sharp peak in the continuum mode amplitudes.

The long range approximation to the Green's function G is obtained by replacing the Hankel function in (28) by its large argument asymptotic form,[9,14,20]

$$G_H(\epsilon,r)\approx -\sqrt{\frac{i}{8\pi\sqrt{\epsilon}r}}\,e^{i\sqrt{\epsilon}r},$$

and dropping the part of the integral in (27) which does not propagate horizontally, that is, that part of the integral with $\epsilon\leqslant 0$. One obtains

$$G(\omega,r,z_0,z)\approx -\sqrt{\frac{i}{8\pi r}}\left[\sum_{j=1}^{N}\eta_j^{-1/4}e^{i\sqrt{\eta_j}r}\phi_j(z_0)\phi_j(z) + \int_0^\infty \epsilon^{-1/4}e^{i\sqrt{\epsilon}r}\psi_\epsilon(z_0)\psi_\epsilon(z)d\epsilon\right]. \tag{29}$$

Note that for fixed z the continuum modes $\psi_\epsilon(z)$ go to zero exponentially in ϵ as $\epsilon\to\infty$ so that the improper integral in (29) converges rapidly. As mentioned above, performing these integrals is computationally intensive. However, it is generally expected to be unnecessary to compute the continuum integral since at long ranges the term $e^{i\sqrt{\epsilon}r}$ oscillates rapidly compared to the variations of $\epsilon^{-1/4}\psi_\epsilon(z_0)\psi_\epsilon(z)$, causing the integral to be negligible.[9]

An exception to this rule is the case in which $\psi_\epsilon(z_0)\psi_\epsilon(z)$ is sharply peaked. As discussed in the previous section this occurs near those frequencies at which the num-

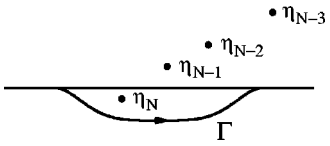


FIG. 10. The integration contour to extract the effect of poles passing near the continuous spectrum.

ber of modes changes. The contribution from these peaks can be extracted by deforming the integration contour in (29) below the real line to the contour Γ depicted in Fig. 10. The term $e^{i\sqrt{\epsilon}r}$ grows exponentially with range for Im $\epsilon<0$, however, the contour Γ need only descend below the real line a distance large compared to the $|\text{Im}\ \eta_j|$ and thus may be chosen so that this exponential growth is insignificant for ranges $r\ll 1/|\text{Im}\ \eta_N|$ (here η_N is used here since the imaginary parts of the η_j typically decrease with increasing j). In the examples considered here $|\text{Im}\ \eta_j|\ll 10^{-4}$ per meter so that for $r\leqslant 10$ km this deformation can be made. In addition, the analytic continuations of the continuum modes to complex ϵ will grow exponentially as $z\to\infty$. However, for low altitudes, for ranges up to 10 km, and for Γ sufficiently far from the poles the contribution from the resulting integral is negligible.

If any poles are encountered in making this deformation, their residues (26) must be subtracted from the integral. The long range contribution from the continuum integral is contained entirely in these residues. Ignoring the continuum integral over the deformed contour Γ one obtains as an approximation to (29)

$$G(\omega,r,z_0,z)\approx -\sqrt{\frac{i}{8\pi r}}\left[\sum_{j=1}^{N}\eta_j^{-1/4}e^{i\sqrt{\eta_j}r}\phi_j(z_0)\phi_j(z)+\sum_{j=N+1}^{M}\eta_j^{-1/4}e^{i\sqrt{\eta_j}r}\phi_j(z_0)\phi_j(z)\right], \quad (30)$$

where, for $j>N$, the quasi-modes defined by (26) are used.

To check the accurracy of the approximation (30) the continuum integral has been computed using a trapezoidal rule approximation with variable step size (a finer step size is used in the region near Re $\epsilon\sim k_0^2$ where the integrand is more rapidly varying). The integral is performed along the deformed contour Γ. In Fig. 11, ground to ground propagation, $z_0=z=0$, is studied. Ten times the logarithm of the magnitude of the Green's function times range, $|\mathbf{x}_H|G(\omega,\mathbf{x}_H,0,z)$, is displayed. This quantity is the magnitude in dB of the sound field produced by a point source relative to the field that would be produced by a point source of equal strength at a distance $|\mathbf{x}_H|$ from the source in an empty homogeneous space. The full field is compared to the individual contributions from the modal sum and the continuum integral. Frequencies 50 Hz, which is not a transition frequency, and 54 Hz, which is a transition frequency, are shown in Figs. 11(a) and 11(b), respectively. The small fluctuations in the computation of the continuum integral at short ranges is the result of numerical noise associated with the vertical discretization of the continuum modes as $\epsilon\downarrow 0$. The sum of modes and quasi-modes alone, (30), is seen to be an excellent approximation to the full field.

VI. CONCLUSIONS

A vertical modal expansion has been produced for the model commonly used to describe outdoor sound propagation on calm nights over flat ground. In this model the sound field is refracted downwards near the ground, upwards at high altitudes and is attenuated by its interaction with the ground. The full sound field is given by a sum over ducted modes plus an integral over continuum modes as given in (15). The ducted modes decrease exponentially at high altitudes while the continuum modes oscillate with very slowly decreasing amplitude at high altitudes.

Except for a discrete set of narrow frequency bands in which the number of ducted modes changes, at long ranges the sound field near the ground is well approximated by the modal sum alone. For these exceptional frequencies the continuum integral has a long range contribution to the sound field. At low altitudes, however, the contribution from the continuum integral is well approximated by a sum over quasi-modes. The quasi-modal sum has the same form as the modal sum except that the quasi-modes increase exponentially at high altitudes.

The modes and quasi-modes, and the associated wave numbers and attenuation constants, can be obtained by perturbing about a sound speed profile which is constant at high altitudes, reducing the problem to the type solved in Ref. 10. Given the modes and quasi-modes, the result, presented in (30) for propagation from a point source, is an easily evalu-

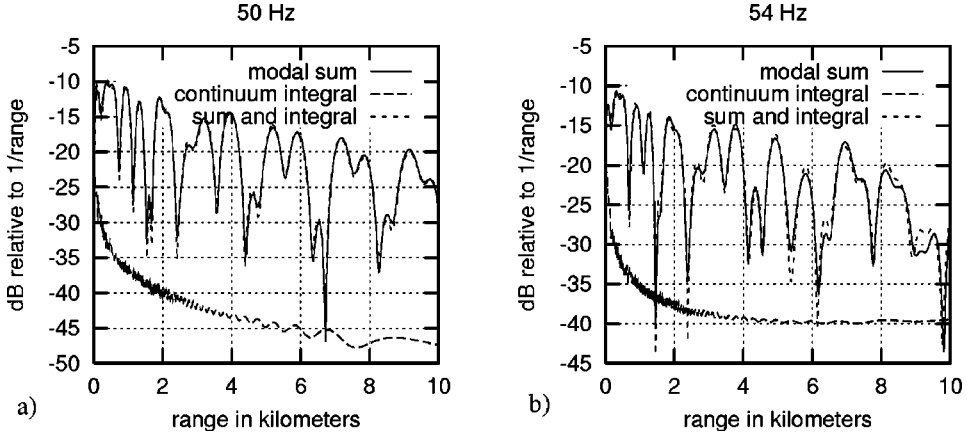


FIG. 11. A comparison of the full sound field with the individual contributions from the modal sum and the continuum integral for ground to ground propagation. The pressure field produced by a unit point source on the ground is shown as a function of range and height in dB relative to 1/range: (a) 50 Hz, no quasi-modes, (b) 54 Hz, one quasi-mode.

ated analytical expression valid at low altitudes for frequencies greater than a few Hz.

ACKNOWLEDGMENTS

It is a pleasure to thank Ken Gilbert with whom every aspect of this work has been discussed. His insight and guidance have been invaluable. In addition I'd like to thank Carrick Talmadge for many helpful discussions. This work was supported by ARDEC Contract No. DAAE30-02-D-1017.

APPENDIX: NORMALIZING THE MODES

In this appendix it is shown how, under rather general conditions, the ducted and continuum modes can be normalized. Normalizing the ducted modes by computing the normalization integral (10) can be difficult since the modes can extend very high up into the atmosphere. Instead, a form similar to that of Appendix B5 of Ref. 9 will be used. In Ref. 10 the continuum modes were normalized using the fact that for asymptotically constant sound speed profiles the continuum modes become trigonometric functions for large z so that a comparison could be made to the well known sine and cosine transforms. For the sound speed profiles considered in this paper this is no longer the case. Instead, formulas are developed which do not depend on the specific large z asymptotic form of the modes. The approach used is a generalization of a technique developed to normalize continuum modes in atomic physics.[21]

Let $\eta \in \mathbf{R}$ and consider the Green's function, $g(\mathcal{C},\eta,z_1,z_2)$, given by

$$\left(\frac{d^2}{dz_1^2}+v(z_1)-\eta\right)g(\mathcal{C},\eta,z_1,z_2)=\delta(z_1-z_2),$$

$g(\mathcal{C},\eta,z_1,z_2)=g(\mathcal{C},\eta,z_2,z_1)$ and

$$\frac{\partial}{\partial z_1}g(\mathcal{C},\eta,z_1,z_2)\bigg|_{z_1=0}=-\mathcal{C}g(\mathcal{C},\eta,0,z_2).$$

Here $v(z)=k(z)^2$ is positive. One has

$$g(\mathcal{C},\eta,z_1,z_2)=\frac{1}{\mathcal{W}(\psi_-(\mathcal{C},\eta,\cdot),\psi_+(\eta,\cdot))}\times\psi_-(\mathcal{C},\eta,z_<)\psi_+(\eta,z_>), \tag{A1}$$

where $z_{<(>)}$ is the smaller (larger) of z_1 and z_2, $\psi_-(\mathcal{C},\eta,z)$ and $\psi_+(\eta,z)$ are solutions of the differential equation

$$\left(\frac{d^2}{dz^2}+v(z)-\eta\right)\psi(z)=0,$$

$\psi_-(\mathcal{C},\eta,z)$ satisfies the initial conditions

$$\psi_-'(\mathcal{C},\eta,0)=-\mathcal{C}\psi_-(\mathcal{C},\eta,0), \tag{A2}$$

$$\psi_-(\mathcal{C},\eta,0)=1, \tag{A3}$$

$\psi_+(\eta,z)$, defined only for $\eta\in\mathbf{R}$, is the exponentially decreasing solution, and $\mathcal{W}$ denotes the Wronskian. Note that it follows from (A2) and (A3) that

$$\mathcal{W}(\psi_-(\mathcal{C},\eta,\cdot),\psi_+(\eta,\cdot))=-(\psi_+'(\eta,0)+\mathcal{C}\psi_+(\eta,0)). \tag{A4}$$

Further, using the asymptotic form (9), one finds that

$$\mathcal{W}(\psi_+(\epsilon+i0^+,\cdot),\psi_+(\epsilon-i0^+,\cdot))=-2i. \tag{A5}$$

In addition to (A1) one has the eigenfunction expansion,

$$g(\mathcal{C},\eta,z_1,z_2)=\sum_{j=1}^{N}\frac{1}{\eta_j-\eta}\phi_j(z_1)\phi_j(z_2)+\int_{-\infty}^{\infty}\frac{1}{\epsilon-\eta}\psi_\epsilon(z_1)\psi_\epsilon(z_2)\,d\epsilon, \tag{A6}$$

which follows directly from the definitions of ϕ_j and ψ_ϵ and from the completeness relation (12). Using the fact that when $\eta=\eta_j$

$$\psi_+(\eta_j,0)\psi_-(\mathcal{C},\eta_j,z)=\psi_+(\eta_j,z) \tag{A7}$$

and comparing the residues of $g(\mathcal{C},\eta,z_1,z_2)$ at the η_j computed from (A1) and (A6), one finds that

$$\phi_j(z_1)\phi_j(z_2)=\frac{\psi_+(\eta_j,0)}{(\partial/\partial\eta)\mathcal{W}(\psi_-(\mathcal{C},\eta,\cdot),\psi_+(\eta,\cdot))\big|_{\eta=\eta_j}}\times\psi_-(\mathcal{C},\eta_j,z_1)\psi_-(\mathcal{C},\eta_j,z_2).$$

Using (A7) the equality

$$\int_0^\infty\psi_+(\eta_j,z)^2dz=-\psi_+(\eta_j,0)\times\frac{\partial}{\partial\eta}(\psi_+'(\eta,0)+\mathcal{C}\psi_+(\eta,0))\bigg|_{\eta=\eta_j} \tag{A8}$$

for the normalization of the eigenfunctions in the point spectrum follows.

As regards the normalization of the eigenfunctions in the continuous spectrum, note that

$$g(\mathcal{C},\epsilon_0+i0^+,z_1,z_2)-g(\mathcal{C},\epsilon_0-i0^+,z_1,z_2)=\lim_{\delta\downarrow 0}\int_{-\infty}^{\infty}\frac{2i\delta}{(\epsilon-\epsilon_0)^2+\delta^2}\psi_\epsilon(z_1)\psi_\epsilon(z_2)\,d\epsilon=2\pi i\psi_{\epsilon_0}(z_1)\psi_{\epsilon_0}(z_2).$$

Let $\epsilon\in\mathbf{R}$ and note that

$$
\begin{aligned}
2\pi i\psi_\epsilon(z_1)\psi_\epsilon(z_2) &= \int_0^\infty \Bigg[g(\mathcal{C},\epsilon+i0^+,z_1,w)\left(\frac{d^2}{dw^2}+v(w)-\epsilon\right)g(\mathcal{C},\epsilon-i0^+,w,z_2) \\
&\quad - g(\mathcal{C},\epsilon-i0^+,w,z_2)\left(\frac{d^2}{dw^2}+v(w)-\epsilon\right)g(\mathcal{C},\epsilon+i0^+,z_1,w)\Bigg]\,dw \\
&= \left[g(\mathcal{C},\epsilon+i0^+,z_1,w)\frac{d}{dw}g(\mathcal{C},\epsilon-i0^+,w,z_2) - g(\mathcal{C},\epsilon-i0^+,w,z_2)\frac{d}{dw}g(\mathcal{C},\epsilon+i0^+,z_1,w)\right]\Bigg|_{w=0}^{\infty} \\
&= \frac{\mathcal{W}(\psi_+(\epsilon+i0^+,\cdot),\psi_+(\epsilon-i0^+,\cdot))}{\mathcal{W}(\psi_-(\mathcal{C},\epsilon,\cdot),\psi_+(\epsilon+i0^+,\cdot))\mathcal{W}(\psi_-(\mathcal{C},\epsilon,\cdot),\psi_+(\epsilon-i0^+,\cdot))}\,\psi_-(\mathcal{C},\epsilon,z_1)\psi_-(\mathcal{C},\epsilon,z_2). \qquad \text{(A9)}
\end{aligned}
$$

Here the fact that $\psi_-(\mathcal{C},\eta,z)$ is continuous in η as η crosses the real line while $\psi_+(\eta,z)$ is not has been used. Equations (14) and (13) now follow from (A4) and (A5).

[1] R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer Academic, Dordrecht, 2001).
[2] A. D. Pierce, "Propagation of acoustic-gravity waves in a temperature and wind stratified atmosphere," J. Acoust. Soc. Am. **37**, 218–227 (1965).
[3] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN Spon, London, 1997).
[4] E. Salomons, *Computational Atmospheric Acoustics* (Kluwer, Dordrecht, 2001).
[5] D. Keith Wilson, J. M. Noble, and M. A. Coleman, "Sound propagation in the nocturnal boundary layer," J. Atmos. Sci. **60**(20), 2473–2486 (2003).
[6] K. Attenborough, "Acoustical impedance models for outdoor ground surfaces," J. Sound Vib. **99**, 521–544 (1985).
[7] T. F. W. Embleton, "Tutorial on sound propagation outdoors," J. Acoust. Soc. Am. **100**, 31–48 (1996).
[8] M. Garstang, D. Larom, R. Raspet, and M. Lindeque, "Atmospheric controls on elephant communication," J. Exp. Biol. **198**, 939–951 (1995).
[9] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[10] R. Waxler, "A vertical eigenfunction expansion for the propagation of sound in a downward-refracting atmosphere over a complex impedance plane," J. Acoust. Soc. Am. **112**, 2540–2552 (2002).
[11] E. C. Titchmarsh, "Some theorems on perturbation theory," Proc. R. Soc. London, Ser. A **210**, 30–44 (1951).
[12] F. S. Henyey, D. Tang, and S. A. Reynolds, "Vertical modes with sediment absorption in generalized Pekeris waveguides," J. Acoust. Soc. Am. **113**, 2187 (2003).
[13] N. Dunford and J. T. Schwartz, *Linear Operators, Part III* (Wiley, New York, 1971).
[14] F. W. J. Olver, *Asymptotics and Special Functions* (Academic, New York, 1974).
[15] In the mathematical literature such eigenfunction expansions have been shown to exist for sound speed profiles which are asymptotically constant (for example in Ref. 13). To this author's knowledge the extension to sound speed profiles which are asymptotically decreasing (indefinitely) has not been considered. A proof of the existence of such an extension will not be attempted here. One may view the expansion used here either as a hypothetical extension of the asymptotically constant case or as an approximate form for the case in which the sound speed profile eventually becomes constant, but at a height well above the domain of interest, and at a value well below the sound speed on the ground.
[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in C*, 2nd ed. (Cambridge U.P., New York, 1992).
[17] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed. (Springer, New York, 1991).
[18] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, 3rd ed. (Pergamon, New York, 1977), Sec. 38.
[19] A. D. Pierce, *Acoustics* (Acoustical Society of America, Woodbury, NY, 1989).
[20] P. M. Morse and K. U. Ingaard, *Theoretical Acoustics* (Princeton U.P., Princeton, NJ, 1986).
[21] R. Froese and R. Waxler, "The spectrum of a hydrogen atom in an intense magnetic field," Rev. Math. Phys. **6**, 699–832 (1995), Appendix B.

Caustic envelopes and cusp coordinates due to the reflection of a spherical wave from a layered sediment

Michael A. Ainslie[a)]
TNO Physics and Electronics Laboratory, Underwater Acoustics Group, Oude Waalsdorperweg 63, 2509 JG, The Hague, The Netherlands

Alvin J. Robins
CORDA Limited, Chester House, Farnborough Aerospace Centre, Farnborough, Hampshire, GU14 6YU, United Kingdom

Dick G. Simons
TNO Physics and Electronics Laboratory, Underwater Acoustics Group, Oude Waalsdorperweg 63, 2509 JG, The Hague, The Netherlands

(Received 23 September 2002; revised 10 December 2003; accepted 30 December 2003)

A compilation of analytical formulas is presented, describing ray properties of a family of sediment sound speed profiles. The profiles are described as ''generalized power law'' profiles because they are generalizations of a similar family of simpler power law profiles considered in earlier work [M. A. Ainslie, J. Acoust. Soc. Am. **96**, 2506–2515 (1994)]. Properties considered include ray paths, caustic envelopes, and cusp coordinates due to a point source in isovelocity water. For a few example cases the caustic shapes are illustrated by computing the pressure field using a normal mode program. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1650841]

PACS numbers: 43.20.Dk, 43.20.El, 43.30.Cq, 43.30.Ma [WLS] Pages: 1449–1459

I. INTRODUCTION

The speed of sound in ocean sediments is often characterized by a monotonically increasing function of depth[1] and this behavior results in the formation of caustics in the reflected field (see paragraph 46 of Ref. 2), caused by upward refraction in the sediment layer. For the case of isovelocity water with a linear increase in sound speed with depth in the sediment, there is a minimum range that can be reached by a sediment refracted eigenray, depending on the heights of the source and receiver above the seabed. The locus of such points of minimum range, for different receiver heights, is the caustic envelope. Stickler[3] demonstrates the importance of taking such caustics into account for the interpretation of seabed reflection loss measurements. Conversely, a measurement of the caustic position can be exploited to measure the sediment sound speed gradient.[4,5]

Caustic shapes and ray paths are usually found numerically,[3,6] since general sediment profiles do not permit analytical solutions. Orlov,[7] however, in the context of reflection of electromagnetic waves from an electron plasma, obtains analytical solutions for the caustic shapes associated with selected velocity profiles of the form

$$[c_0/c(z)]^2 = 1 + az^p, \tag{1}$$

where $c(z)$ is the wave speed as a function of depth z, and c_0, a, and p are constants. Orlov[8] also considers a parabolic profile for which the ray paths, but not the caustic envelope, can be found analytically. Ainslie[9] analyzes the properties of caustics associated with the reflection of sound from the seabed with various power law profiles (PLPs). The advantage of exact solutions is in allowing such properties as the caustic shape to be found explicitly in terms of the parameters describing the profile, and hence permitting insight into the variation of caustic properties with profile shape. Possible applications are discussed in the following.

(1) The solutions can be used to help interpret the result of an acoustic measurement such as for geoacoustic inversion. This is seen as the main application and is discussed further in Sec. V.

(2) An experiment designed to measure the position of a caustic can benefit from knowing in advance the maximum range of the caustic from the source, or its height from the seabed. Using analytical solutions, these can be estimated without the need for a lengthy sensitivity study.

(3) Differentiation of the caustic coordinates with respect to a given seabed parameter provides a measure of the accuracy and precision with which it is necessary to measure the caustic position in order to infer that parameter.

(4) The availability of exact solutions for the caustic envelope is also useful for checking the accuracy and convergence of a numerical propagation model. The most obvious application of this nature is to check the ray paths computed by a numerical ray tracing program. However, as computing power increases, wave models are used at ever higher frequencies as a realistic alternative to ray tracing. Thus the caustic envelope solutions can be used as a test of the high frequency limit of wave models, especially for cases involving steep propagation angles, for which wave equation solutions are sometimes pushed to their limit. We have in mind situations requiring a high order Padé expansion, a large number of normal modes, or a large discrete Fourier transform size.

In this paper we consider a generalization of the PLP,

a)Electronic mail: ainslie@fel.tno.nl

TABLE I. First and second derivatives of the GPL profiles.

	$c(z)$	$\frac{dc/dz}{qc_0}$	$\frac{d^2c/dz^2}{3q^2c_0}$
GPL_n[a]	Eq. (2)	$\left[\frac{c(z)}{c_0}\right]^3\left(1+\frac{nq}{A}z\right)^{-1-(2/n)}$	$\chi\left[\frac{c(z)}{c_0}\right]^2\frac{dc/dz}{qc_0[1+(nq/A)z]}\left\{1+\frac{n-1}{3\chi}\left[1-\left(1+\frac{nq}{A}z\right)^{-2/n}\right]\right\}$
GPL_{-2}	Eq. (4)	$\left[\frac{c(z)}{c_0}\right]^3$	$\left[\frac{c(z)}{c_0}\right]^2\frac{dc/dz}{qc_0}$
GPL_0	Eq. (6)	$\left[\frac{c(z)}{c_0}\right]^3 e^{-3q(1-\chi)z}$	$\chi\left[\frac{c(z)}{c_0}\right]^2\frac{dc/dz}{qc_0}\left[1+\frac{e^{-3q(1-\chi)z}-1}{3\chi}\right]$

[a] $n\neq 0,\ -2$.

referred to as a generalized power law (GPL) profile, for which analytical solutions are found for the ray paths. The main advantage of the GPL is that it permits an arbitrary choice to be made for sound speed curvature—rather than a set of discrete values associated with PLP—thus allowing a better fit to any given measured profile. Some variation in the rate of change of curvature (with depth) is also possible using the GPL. We derive caustic equations for the GPL profiles and plot examples of caustic shapes for various cases. Some exhibit special features such as cusps. Where possible the caustic equations are solved analytically for the coordinates of these features.

In Sec. II we describe the GPL profile and explain how it is related to the PLP. Ray theory predictions for caustic envelopes and cusp coordinates are presented for various GPL profiles in Sec. III, followed by a comparison with normal mode predictions in Sec. IV for selected cases. The potential application of the analysis to the problem of geoacoustic inversion is discussed in Sec. V, followed by conclusions in Sec. VI.

II. GENERALIZED POWER LAW (GPL) PROFILE

The GPL profile is expressed in terms of gradient and curvature parameters q and χ such that

$$c_0^2/c(z)^2 = A[1+(nq/A)z]^{-2/n}+1-A, \tag{2a}$$

where A is a constant that we write as

$$A=\frac{n+2}{3(1-\chi)}. \tag{2b}$$

The above choice of A in terms of n and χ is made so that the first and second derivatives of the sound speed at the top of the sediment ($z=0$) are given by qc_0 and $3\chi q^2 c_0$, respectively. (See Table I for a list of expressions for these two derivatives.) In principle, for Eq. (2), the parameter n can take any real value, but we restrict attention here to integer values only, as these are amenable to analytical solution. The term "generalized power law" is used to indicate that GPL is a generalization of the straightforward power law profile (PLP).[9] The latter is obtained by setting $A=1$ in Eq. (2a), namely

$$c(z)/c_0=(1+nqz)^{1/n}. \tag{3}$$

Other noteworthy special cases, using a subscript (i.e., GPL_n) to indicate the value of n, are as follows.

(1) The "linear k^2" profile is given by GPL_{-2}:

$$c_0^2/c^2(z)=1-2qz. \tag{4}$$

This profile is often used to represent first-order effects of the sediment sound speed gradient,[2,10] and is one of the special cases considered by Orlov.[7] Notice that $c(z)$ is independent of A in this case, because A cancels exactly from Eq. (2a) when $n=-2$. [The value of $c''(0)$, from Eq. (4), is $3q^2c_0$, corresponding to $\chi=1$, so that A is indeterminate in Eq. (2b)]. The profiles GPL_{-2} and PLP_{-2} are therefore identical. More generally, in the limit $\chi\to 1$ *all* GPL profiles reduce to PLP_{-2}.

(2) The case $n=-1$ gives the parabolic k^2 profile considered by Orlov,[8]

$$c_0^2/c^2(z)=1-2qz+3(1-\chi)(qz)^2. \tag{5}$$

(3) The case $n=0$ leads to the generalized exponential profile GPL_0:

$$3(1-\chi)c_0^2/c^2(z)=2\exp[-3q(1-\chi)z]+1-3\chi. \tag{6}$$

(4) Setting $n=+1$ gives the profile GPL_{+1} considered by Rytov and Yudkevich:[11]

$$(1-\chi)c_0^2/c^2(z)=[1+q(1-\chi)z]^{-2}-\chi. \tag{7}$$

This profile, sometimes referred to as an "inverse square" profile, is used by Robins[12] and by Ainslie *et al.*[13] to examine the effect of sound speed curvature in the sediment layer. It is also used by Buckingham[14,15] to model the formation of wave guides at the sea surface due to a near surface bubble population.

Example GPL profiles are plotted versus depth in Fig. 1. It is convenient, in these plots, to show c/c_0 as a function of the dimensionless variable $18qz$. Figures 1(a)–(d) show PLP and GPL profiles for $n=-3$, -1, $+1$, and $+3$ respectively. Each graph plots the GPL profile for $\chi=-2.0$ to $+2.0$ in steps of 1.0, together with the corresponding PLP (shown as a dashed line) for the same value of n. It can be seen that an appropriate choice of n and χ gives considerable scope for modeling a variety of profile shapes. These plots illustrate a few noteworthy features, namely the following.

(1) The GPL_{-3} curve with $\chi=+2$ [Fig. 1(a)] terminates at $18qz=2$. This represents the largest value of z for which Eq. (2) yields a real sound speed. More generally, if $n\leqslant -3$ and $0<A<1$, the sound speed $c(z)$ becomes complex for $|n|qz>A$.

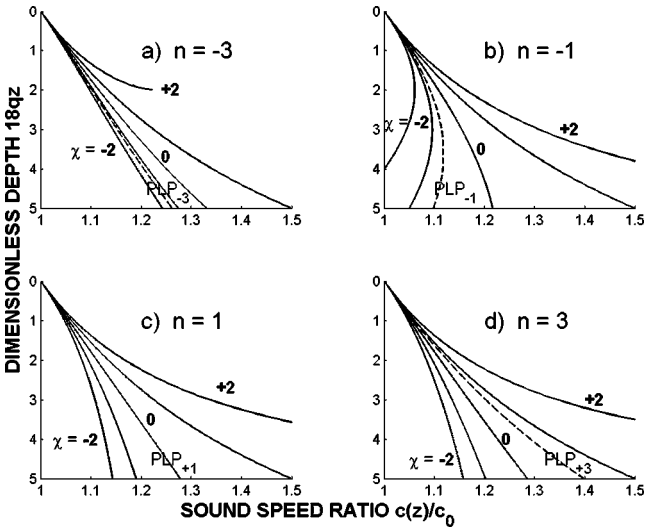


FIG. 1. GPL_n sound speed profiles (solid lines) for four values of n as marked. The curvature parameter χ increases from -2 to $+2$ in steps of 1 from left to right. The corresponding PLP profiles are shown as dashed lines, except for PLP_{+1} which is identical to GPL_{+1} with $\chi=0$.

(2) The GPL_{-1} curves with $\chi=-2$ and -1 [Fig. 1(b)] show maxima in the sound speed at $18qz=2$ and 3, respectively. It is easily shown from Eq. (2) that such maxima can occur in a GPL profile only in the case $n=-1$, and then only if $\chi<2/3\,(0<A<1)$. If both conditions are satisfied the turning depth is given by $qz=A$.
(3) The GPL_{+1} curve with $\chi=0$ [Fig. 1(c)] coincides with the PLP, and is the simple linear profile $c(z)/c_0=1+qz$. This is a special case of the more general statement that GPL_n reduces to PLP_n for $A=1$ [i.e., $\chi=(1-n)/3$].
(4) The four curves corresponding to $\chi=+1$ are identical and equal to PLP_{-2}.

Some of the profiles exhibit singularities for which the sound speed becomes infinite at finite depth. The depths at which these singularities occur and the profiles' other main features are summarized in Table II. The table excludes the special case $n=-2$ for which a singularity occurs at $qz=1/2$.

TABLE II. Summary of main features for the GPL profiles. Unless otherwise indicated the expressions quoted give values of qz for which the sound speed becomes singular.

	$A<0$		$0<A<1$	$A\geqslant 1$
$n\geqslant +1$	$\dfrac{\lvert A\rvert}{n}\left[1-\left(\dfrac{\lvert A\rvert}{1+\lvert A\rvert}\right)^{n/2}\right]$	[a]		$\dfrac{A}{n}\left[\left(\dfrac{A}{A-1}\right)^{n/2}-1\right]$
$n=0$	$\dfrac{\lvert A\rvert}{2}\ln\left(\dfrac{1+\lvert A\rvert}{\lvert A\rvert}\right)$	[a]		$\dfrac{A}{2}\ln\left(\dfrac{A}{A-1}\right)$
$n=-1$	$\lvert A\rvert\left[\left(\dfrac{1+\lvert A\rvert}{\lvert A\rvert}\right)^{1/2}-1\right]$	[b]		$A\left[1-\left(\dfrac{A-1}{A}\right)^{1/2}\right]$
$n\leqslant -3$	$\dfrac{\lvert A\rvert}{\lvert n\rvert}\left[\left(\dfrac{1+\lvert A\rvert}{\lvert A\rvert}\right)^{\lvert n\rvert/2}-1\right]$	[c]		$\dfrac{A}{\lvert n\rvert}\left[1-\left(\dfrac{A-1}{A}\right)^{\lvert n\rvert/2}\right]$

[a]Sound speed profile is real, finite and monotonic for all $qz>0$.
[b]Sound speed profile has a maximum at $qz=A$.
[c]Sound speed becomes complex for $qz>A/\lvert n\rvert$.

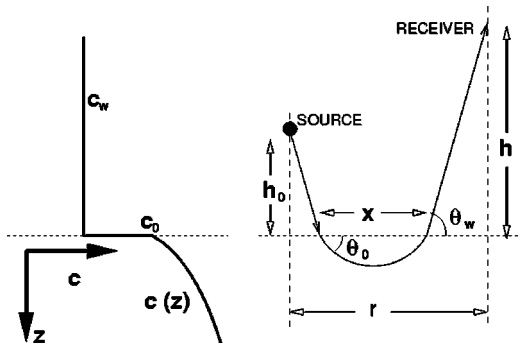


FIG. 2. Sound speed profile and ray geometry. The vertical coordinate z denotes depth increasing downwards into the sediment. The source is at height h_0 $(z=-h_0)$.

III. RAY AND CAUSTIC EQUATIONS

A. Caustic envelopes and cusp coordinates for the general case

1. Basic ray equations

We consider a point source at a height h_0 above the water–sediment boundary. From Fig. 2 it is apparent that the horizontal range r reached by an individual ray, after reflection from the sediment, is given by

$$r(\zeta,\theta_w)=x(\theta_0)+(2\zeta/q)\cot\theta_w, \tag{8}$$

where $x(\theta_0)$ is the horizontal ray displacement along the boundary and ζ is the dimensionless height

$$\zeta=q(h+h_0)/2. \tag{9}$$

With the exception of the source height h_0, the subscripts w and 0 indicate, respectively, a property (usually ray angle or sound speed) evaluated in the water and at the top of the sediment. Caustics form along the locus of paths satisfying $\partial r/\partial\theta_w=0$, that is:

$$q\,dx/d\theta_0=2\zeta\,\mathrm{cosec}^2\,\theta_w\cot\theta_w\tan\theta_0. \tag{10}$$

A caustic can exist only along a ray path, so Eqs. (8) and (10) must both be satisfied.

Defining $\zeta_c(\theta_w)$ as the height ζ at which a ray of angle θ_w grazes the caustic, it follows from Eq. (10) that

$$2\zeta_c(\theta_w)=q\sin^2\theta_w\frac{\tan\theta_w}{\tan\theta_0}\frac{dx(\theta_0)}{d\theta_0}. \tag{11}$$

Ray theory provides a simple expression for the horizontal displacement

$$x(\theta_0)=2\int_0^{z_T}\cot\theta(z)\,dz \tag{12}$$

in terms of the grazing angle determined by Snell's law,

$$c_0\cos\theta(z)=c(z)\cos\theta_0, \tag{13}$$

and the turning depth z_T such that

$$\theta(z_T)=0. \tag{14}$$

Changing the integration variable from z to θ we obtain for GPL,

TABLE III. Evaluation of integrals for the GPL ray displacement $x(\theta_0)$.

n	$(q/\cos\theta_0)x(\theta_0)$
-3	$A^{1/2}\left[\tan\psi_0\sec\psi_0+\ln\tan\left(\frac{\pi}{4}+\frac{\psi_0}{2}\right)\right]\cos^2\psi_0$
-2	$2\sin\theta_0$
-1	$2A^{1/2}\ln\tan\left(\frac{\pi}{4}+\frac{\psi_0}{2}\right)$
0	$2A^{1/2}\psi_0\sec\psi_0$
$+1$	$2A^{1/2}\tan\psi_0\sec\psi_0$
$+2$	$A^{1/2}(\psi_0+\sin\psi_0\cos\psi_0)\sec^3\psi_0$
$+3$	$2A^{1/2}\tan\psi_0\sec\psi_0\left(1+\frac{2}{3}\tan^2\psi_0\right)$

$$qx(\theta_0)=2\cos^2\theta_0\int_0^{\theta_0}\frac{\sec^2\theta\,d\theta}{[\cos^2\psi_0+(\cos^2\theta_0/A)\tan^2\theta]^{1+n/2}}, \tag{15}$$

where

$$A\sin^2\psi_0=\sin^2\theta_0. \tag{16}$$

In order to evaluate Eq. (15) we first note that it is of the form

$$a_n=\int\frac{\sec^2\alpha\,d\alpha}{(B^2+\tan^2\alpha)^{1+n/2}} \tag{17}$$

and make the substitution $\tan\alpha=B\sinh\beta$ so that (pp. 94–97 of Ref. 16)

$$a_n=\frac{1}{B^{n+1}}\int\operatorname{sech}^{n+1}\beta\,d\beta. \tag{18}$$

The results for $-3\leqslant n\leqslant+3$ are presented in Table III. The entry for $n=-1$ corrects a sign error in Ref. 8.

2. Caustic equations for arbitrary c_0/c_w

The caustic equation requires differentiation of the ray displacement $x(\theta_0)$. The derivative $q\,dx/d\theta_0$ is given in Table IV for GPL profiles satisfying $-3\leqslant n\leqslant 3$. The corresponding PLP results[17] can be obtained from Tables III and IV [or Eq. (15)] by substituting $A=1$, so that $\psi_0=\theta_0$.

TABLE IV. Evaluation of $dx/d\theta_0$ for GPL profiles.

n	$(q/2)dx/d\theta_0$
-3	$\cos^2\theta_0\sec^2\psi_0-\frac{1}{2}\sin\psi_0[3\cos^2\theta_0+A-1]\times\left[\tan\psi_0\sec\psi_0+\ln\tan\left(\frac{\pi}{4}+\frac{\psi_0}{2}\right)\right]$
-2	$\cos 2\theta_0$
-1	$\cos^2\theta_0\sec^2\psi_0-A\sin\psi_0\ln\tan\left(\frac{\pi}{4}+\frac{\psi_0}{2}\right)$
0	$\sec^2\psi_0[\cos^2\theta_0+(1-A)\psi_0\tan\psi_0]$
$+1$	$\sec^2\psi_0[1+2(1-A)\tan^2\psi_0]$
$+2$	$\sec^2\psi_0\{1+\psi_0\tan\psi_0+\frac{1}{2}(1-A)[3\tan^2\psi_0+\psi_0\tan\psi_0(1+3\tan^2\psi_0)]\}$
$+3$	$\sec^2\psi_0[1+2\tan^2\psi_0+\frac{2}{3}(1-A)\tan^2\psi_0(3+4\tan^2\psi_0)]$

Equation (11) for the caustic height can be used, with Eq. (8) and Tables III and IV, to determine the caustic envelope for GPL profiles with arbitrary values of the sound speed ratio c_0/c_w. Figure 3 shows the caustics thus calculated for GPL_{+1}. Each subplot is for a fixed curvature value ($\chi=-1/4$ to $+1/2$ as marked) and the different colors show the effect of varying the sound speed ratio between 0.95 and 1.10. Figure 4 shows the effect of varying n between -3 and $+3$, in steps of 2, for a fixed curvature parameter $\chi=+1/4$. The GPL_0 and GPL_{+2} caustic shapes (not shown) are similar, respectively, to those for GPL_{-1} and GPL_{+3}.

An important general point to bear in mind for the interpretation of Figs. 3 and 4 is that ray paths are always tangential to the caustic envelope. The angle of a caustic-grazing ray at any point on the caustic can therefore be inferred directly from the caustic shape. To understand the caustics' structure seen in Figs. 3 and 4 we first consider the behavior of Eqs. (8) and (11) when θ_0 is small,

$$qr\approx 2\theta_0+2\zeta\cot\theta_w, \tag{19a}$$

$$\zeta_c\approx\sin^2\theta_w\tan\theta_w\frac{1+\frac{1}{2}(1-6\chi)\theta_0^2}{\theta_0}. \tag{19b}$$

Denoting $(c_w/c_0)^2$ by η we use Snell's law in the form

$$\sin^2\theta_w\approx 1-\eta+\eta\theta_0^2 \tag{20}$$

to eliminate θ_w from Eq. (19) and thus solve for θ_0,

$$\theta_0\approx\frac{qr[1\pm\mu(qr)]}{2(3+\eta-\nu)}, \tag{21}$$

where

$$\nu=6\chi(1-\eta) \tag{22}$$

and

$$\mu(\xi)=\sqrt{1-\frac{8(1-\eta)(3+\eta-\nu)}{\xi^2}}. \tag{23}$$

It follows from Eq. (20) that

$$\sin^2\theta_w\approx\frac{1}{8}\left[\frac{qr}{(3+\eta-\nu)}\right]^2[1\pm\mu(qr)]\times[3+3\eta-\nu\mp(3-\eta-\nu)\mu(qr)] \tag{24}$$

and hence

$$\zeta_c(qr)\approx\left[\frac{qr}{2(3+\eta-\nu)}\right]^2\left[\frac{1\pm\mu(qr)}{\eta}\right]^{1/2}\times\left[\frac{3+3\eta-\nu\mp(3-\eta-\nu)\mu(qr)}{2}\right]^{3/2}. \tag{25}$$

The simplest behavior occurs for $c_0/c_w=1$. For this case η and μ are both equal to 1 and the caustic intersects the range axis at the height origin ($\zeta=0$). Equation (25) then simplifies to the usual result for a linear sediment profile[9]

$$\zeta_c(qr)=(qr/4)^2. \tag{26}$$

Increasing the sound speed ratio c_0/c_w (so that $\eta<1$), Eq. (23) with $\xi=qr$ implies a minimum range for which $\mu(\xi)$ can be real. At this minimum range μ vanishes and the two

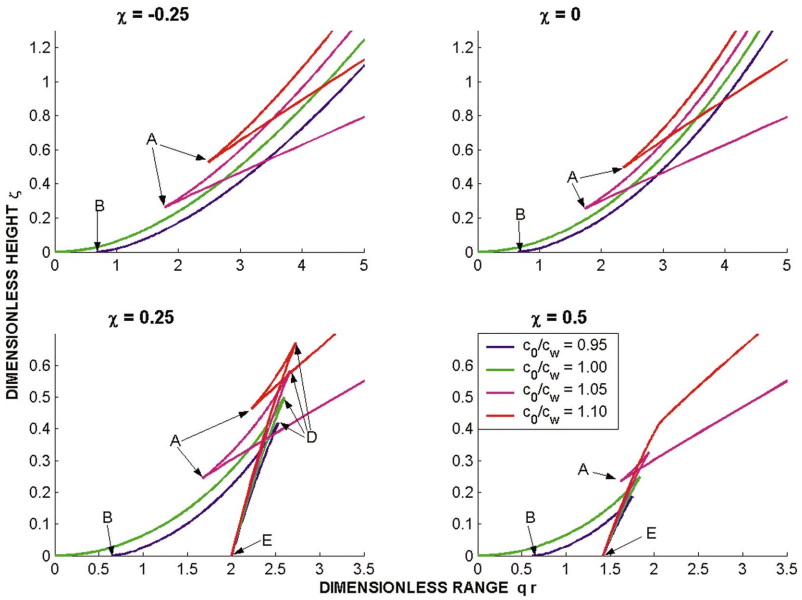


FIG. 3. Caustic shapes for GPL_{+1} profiles for $\chi=-\frac{1}{4}$ to $+\frac{1}{2}$ as marked.

branches of Eq. (25) coalesce into a single cusp (labeled "A" in Fig. 3) at the coordinates

$$qr_{\mathrm{A}}\approx[8(1-\eta)(3+\eta-\nu)]^{1/2}, \tag{27a}$$

$$\zeta_{\mathrm{A}}\approx\frac{1-\eta}{3+\eta-\nu}\left[\frac{(3+3\eta-\nu)^3}{2\eta}\right]^{1/2}. \tag{27b}$$

Similarly, if $c_0/c_w<1$, there exists a minimum range (labeled "B") for the caustic to be visible above the seabed. This occurs when the third factor of Eq. (25) vanishes, the condition for which is

$$qr_{\mathrm{B}}\approx(3-\eta-\nu)\left(\frac{\eta-1}{\eta}\right)^{1/2}, \tag{28a}$$

$$\zeta_{\mathrm{B}}=0. \tag{28b}$$

3. Approximations for $c_0/c_w\approx 1$

It is often the case that the sediment sound speed is such that $1-c_0/c_w$ is small, and the above-mentioned equations then simplify. The results derived in the following, valid in the limit $c_0/c_w\to 1$, are used in Sec. V to illustrate the potential application of this work to geoacoustic inversion. Specifically if we replace c_0/c_w by 1 everywhere except in terms proportional to $1-c_0/c_w$, Eqs. (25), (27), and (28a) become

$$\zeta_c(qr)\approx\left[\frac{qr}{8}\right]^2[1\pm\mu(qr)]^{1/2}[3\mp\mu(qr)]^{3/2}, \tag{29}$$

where

$$\mu(\xi)\approx\sqrt{1-\frac{64(c_0/c_w-1)}{\xi^2}}, \tag{30}$$

$$qr_{\mathrm{A}}\approx 8(c_0/c_w-1)^{1/2}, \tag{31a}$$

$$\zeta_{\mathrm{A}}\approx\sqrt{27}(c_0/c_w-1), \tag{31b}$$

and

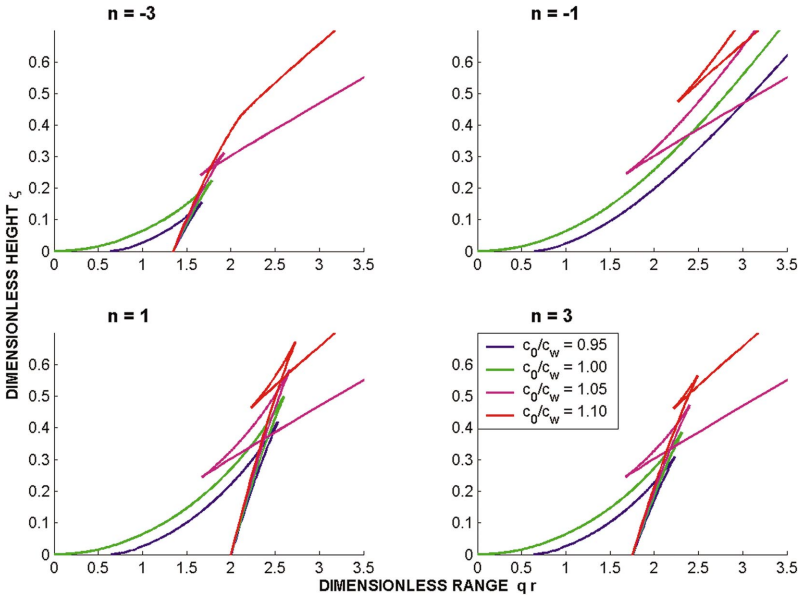


FIG. 4. Caustic shapes for GPL_n profiles for $\chi=\frac{1}{4}$ and $n=-3$ to $+3$ as marked.

$$qr_B \approx [8(1-c_0/c_w)]^{1/2}. \tag{32}$$

Features "A" and "B" are associated with shallow angle rays and are therefore influenced by near-surface properties of the sediment profile (the sound speed ratio c_0/c_w and the gradient at $z=0$). Varying the curvature has little effect on the position of these features because the parameter χ only appears in the above Eq. (25) (through ν) multiplied by $(1-\eta)$ which, by assumption, is small. As a consequence, Eqs. (31) and (32) are independent of χ. The insensitivity of "A" and "B" to curvature is confirmed by inspection of Fig. 3.

At steep angles the effect of varying the curvature becomes more obvious. Increasing χ from 0 to $\frac{1}{4}$ a second cusp (labeled "D") can be seen in Fig. 3, and this in turn results in a second axis intersection ("E"). These features are not predicted by Eq. (25) because of the small angle assumption ($\theta_0 \ll 1$) made in its derivation. They are considered further in Secs. III B and III C below.

B. Caustic envelopes and cusp coordinates for $c_0 = c_w$

The cusps observed in Figs. 3 and 4 arise wherever $\partial r/\partial \theta_w$ is stationary along a caustic. We now derive equations for their coordinates for a few special cases. Differentiating Eq. (8) twice we obtain

$$q\partial r/\partial \theta_w = 2\ \mathrm{cosec}^2\ \theta_w[\zeta_c(\theta_w)-\zeta] \tag{33}$$

and

$$q\partial^2 r/\partial \theta_w^2 = 2\ \mathrm{cosec}^2\ \theta_w\{\zeta_c'(\theta_w) - 2\ \cot\theta_w[\zeta_c(\theta_w)-\zeta]\}. \tag{34}$$

By definition $\zeta_c-\zeta$ vanishes along a caustic, so an alternative condition for a cusp is just $\zeta_c'=0$. Making the simplifying assumption that c_0 and c_w (and hence θ_0 and θ_w) are equal we now analyze the GPL profiles for $n=-2$, $+1$, and $+3$, the equations for which have straightforward solutions for the caustic envelope. From Eq. (12) and Table IV we obtain for the caustic height:

$$\zeta_c(\theta_0) = \sin^2\theta_0 \cos 2\theta_0, \quad n=-2, \tag{35a}$$

$$\zeta_c(\theta_0) = A\tan^2\psi_0[1+2(1-A)\tan^2\psi_0], \quad n=+1, \tag{35b}$$

$$\zeta_c(\theta_0) = A\tan^2\psi_0[1+2(2-A)\tan^2\psi_0 + \tfrac{8}{3}(1-A)\tan^4\psi_0], \quad n=+3, \tag{35c}$$

where ψ_0 is related to θ_0 according to Eq. (16). All three of these equations are soluble analytically for the caustic angle θ_c (the angle θ_0 of the ray that grazes the caustic at height ζ). The first two solutions are

$$\tan^2\theta_c(\zeta) = \begin{cases} \dfrac{1\pm(1-8\zeta)^{1/2}}{3\mp(1-8\zeta)^{1/2}} & (n=-2) \\ \dfrac{(1-\chi)[1\pm(1-8\chi\zeta)^{1/2}]}{4\chi} & (n=+1) \end{cases}. \tag{36}$$

TABLE V. Coordinates of features "D" and "E" for soluble GPL profiles $n=+1$ and $+3$. For both profiles the case $\chi=1$ reduces to the PLP_{-2} coordinates given by Eqs. (39) and (43).

	$n=+1$		$n=+3$	
χ	(qr_D,ζ_D) for $c_0/c_w=1$	qr_E	(qr_D,ζ_D) for $c_0/c_w=1$	qr_E
$-\frac{2}{3}$	⋯	⋯	∞	∞
$-\frac{1}{2}$	⋯	⋯	(22.22,15.13)	15.11
$-\frac{1}{4}$	⋯	⋯	(6.141,2.134)	4.404
0	∞	∞	(3.350,0.762)	2.489
$\frac{1}{4}$	(2.598,0.500)	2.000	(2.312,0.386)	1.754
$\frac{1}{2}$	(1.837,0.250)	1.414	(1.797,0.238)	1.376
$\frac{3}{4}$	(1.500,0.167)	1.155	(1.495,0.165)	1.150
1	(1.299,0.125)	1.000	(1.299,0.125)	1.000
$\frac{5}{4}$	(1.162,0.100)	0.894	(1.160,0.100)	0.893
$\frac{3}{2}$	(1.061,0.083)	0.816	(1.057,0.083)	0.812

The third equation, for $n=+3$, is soluble as a cubic equation in $\tan^2\psi_0$.[18] The corresponding caustic range follows from

$$qr_c(\theta_0) = qx(\theta_0) + 2\zeta_c(\theta_0)\cot\theta_0. \tag{37}$$

Equations for the cusp angle θ_D (the caustic angle at the cusp, labeled "D"), follow by requiring the derivative of Eq. (35) to vanish. The result for $n=-2$ and $+1$ is

$$\tan^2\theta_D = \begin{cases} \dfrac{1}{3}, & n=-2 \quad (\text{i.e., } \theta_D=\pi/6) \\ \dfrac{1}{3\chi}, & n=+1 \end{cases}. \tag{38}$$

The cusp coordinates are therefore

$$(qr_D,\zeta_D) = \begin{cases} \left(\dfrac{27^{1/2}}{4},\dfrac{1}{8}\right), & n=-2 \\ \left(\dfrac{27^{1/2}}{4\chi^{1/2}},\dfrac{1}{8\chi}\right), & n=+1 \end{cases}. \tag{39}$$

For $n=+3$ it is convenient to parametrize the coordinates in terms of A and ψ_D instead of χ and θ_D. Differentiating Eq. (35c) we obtain

$$\tan^2\psi_D = \frac{A-2\pm\sqrt{A^2-2A+2}}{4(1-A)}. \tag{40}$$

The cusp coordinates for this case are then

$$qr_D = \tfrac{4}{3}A^{1/2}\tan\psi_D(3+4\tan^2\psi_D) \times[1+(1-A)\tan^2\psi_D]^{3/2}, \tag{41a}$$

$$\zeta_D = A\tan^2\psi_D[1+2(2-A)\tan^2\psi_D+\tfrac{8}{3}(1-A)\tan^4\psi_D]. \tag{41b}$$

The coordinates of feature "D" are evaluated using Eqs. (39) and (41) and are presented in Table V.

The above-given results for $n=-2$ [i.e., $p=1$ in Eq. (1)] are consistent with those from Refs. 7 and 9. Orlov's calculations[7] are unique in that they also include the caustic

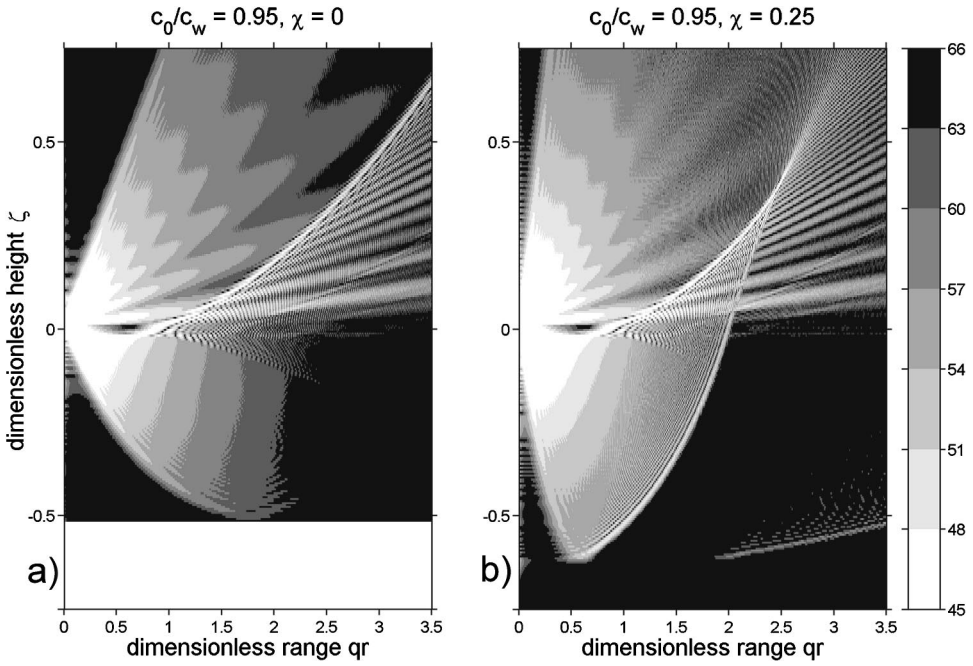


FIG. 5. Propagation loss PL vs dimensionless range qr and height ζ for the GPL_{+1} profile with a sound speed ratio $c_0/c_w=0.95$ and curvature parameter (a) $\chi=0$ and (b) $\chi=1/4$. The caustic envelope can be compared with the blue curves from Fig. 3 for corresponding χ values.

shape *inside the layered medium*. Other profiles analyzed by Orlov, not considered here, are the cases $p=2$ and $p=1/2$.

C. Calculation of range axis intersection (qr_E) for arbitrary c_0/c_w

The condition for r_E is simply $\zeta_c=0$ (and $\theta_w\neq 0$) and it follows from Eq. (11) that $dx/d\theta_0$ must also be zero. This condition results in soluble equations for the same three values of n as above: $n=-2$, $+1$, and $+3$, but this time without the additional requirement that c_0 and c_w be equal. First we consider $n=-2$ and $+1$. The caustic angle (θ_E) satisfying the above-given condition is given by

$$\tan^2\theta_E=\begin{cases}1, & n=-2 \quad (\text{i.e.,} \quad \theta_E=\pi/4)\\ \chi^{-1}, & n=+1\end{cases}. \tag{42}$$

Substituting for this angle in Eq. (8) (with $\zeta_E=0$) for the caustic range then gives

$$qr_E=\begin{cases}1, & n=-2\\ \chi^{-1/2}, & n=+1 \quad (\chi>0)\end{cases}. \tag{43}$$

Applying the same method to $n=+3$ we obtain

$$qr_E=2A^{1/2}\tan\psi_E\sec\psi_E[1+(2/3)\tan^2\psi_E]\cos\theta_E, \tag{44}$$

where

$$\tan^2\psi_E=\frac{3(A-2)\pm\sqrt{9A^2+12(1-A)}}{8(1-A)}. \tag{45}$$

The range r_E is evaluated using Eqs. (43) and (44) and this too is presented in Table V. A direct comparison is possible between the coordinates of Table V and the features ''D'' and ''E'' in Fig. 3 (also visible in Fig. 4, but not marked). No approximation was made in the derivation of these coordinates and the agreement is therefore exact.

Notice that the position of ''E'' is independent of c_0/c_w for both $n=+1$ and $+3$. This is a consequence of the proportionality between ζ_c and $dx/d\theta_0$ [Eq. (11)] and the observation therefore holds not just for all GPL profiles, but completely generally. In other words, assuming that this feature exists at all for a given profile shape $c(z)/c_0$, then its position is independent of the ratio c_0/c_w *for all profiles of that shape*.

IV. COMPARISON WITH NORMAL MODE PREDICTIONS

We now calculate the pressure field in the vicinity of the caustic for selected cases. For these calculations we use a source height h_0 of 10 m and a frequency f of 500 Hz. The sediment profile GPL_{+1} is used, with a sound speed gradient $qc_0=3\ s^{-1}$ so that $qh_0=0.02\,c_w/c_0$. The sediment absorption coefficient is $\alpha=0.01$ dB/λ. In all cases the water depth is 2000 m and the sound speed in water is 1500 m/s. For simplicity, a uniform density of 1.0 g/cm^3 is assumed everywhere.

The value used for sediment attenuation seems low compared with Hamilton's data,[19] but is in line with recent measurements for silt and clay.[6,20,21] A higher value of α would suppress some of the features at steeper angles. Reflection loss for sediment refracted paths is determined primarily by the absorption along the ray path in the sediment which is approximately proportional to the sediment grazing angle θ_0. The constant of proportionality, $2f\alpha/qc_0$, is equal to 10/3 dB rad^{-1}.

Results are presented in the form of propagation loss[22] (PL) versus dimensionless range qr and height ζ. Figures 5–7 show the result of this calculation for three different sound speed ratios between 0.95 and 1.05, and two curvature values ($\chi=0$ and $\frac{1}{4}$) as shown. The shape of the caustics in these figures is a function of the dimensionless variables qr and ζ only. The contour levels depend on the dimensionless frequency f/qc_0, which for this example is equal to 500/3. Black indicates highest loss (PL>63 dB re 1 m^2) and white

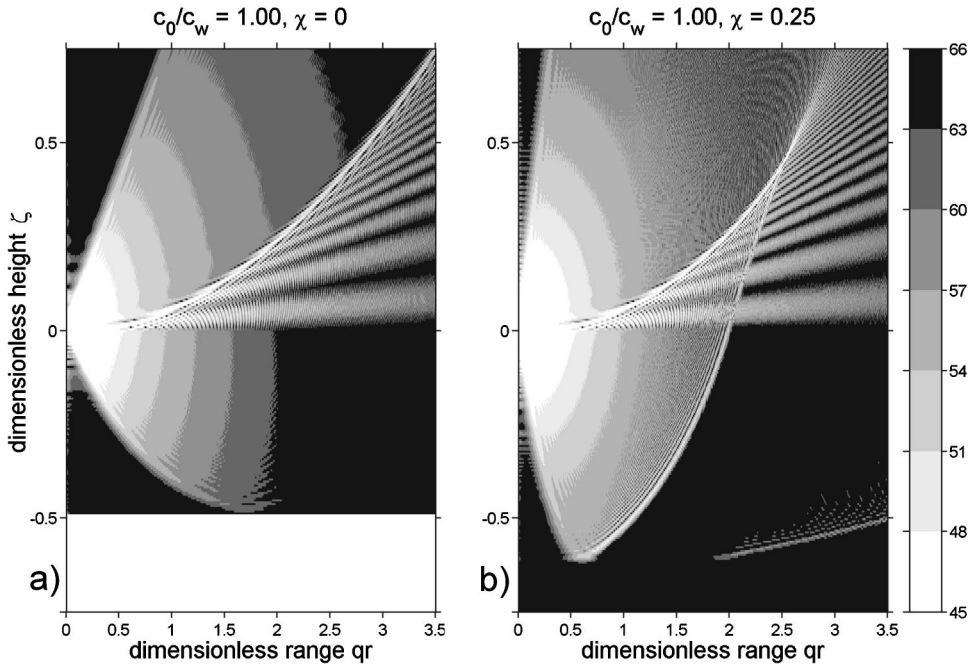


FIG. 6. As in Fig. 5 for $c_0/c_w=1.00$. The caustic envelope can be compared with the green curves from Fig. 3 for corresponding χ values.

indicates lowest (PL<48 dB re 1 m^2). Intermediate contours are at 3 dB intervals. Approximately 1500 modes are used to produce each of these figures.

The simplest behavior is for $c_0/c_w=1$ and $\chi=0$ and for this reason we discuss this case first [Fig. 6(a)]. The main feature of interest is the caustic itself, extending from the origin on the left to $\zeta=0.75$ at $qr=3.5$ on the right, consistent with the corresponding curve from Fig. 3 (green line in top-right graph). Above the caustic there is a region of spherical spreading arising from the direct path only. Beneath it there is a distinct fringe pattern caused by interference between the direct path and the shallower of the two bottom-refracted paths. Superimposed on the main (approximately horizontal) fringes is a fine-scale interference pattern due to the influence of the second (steeper) bottom-refracted path.

The main effect of increasing the curvature to $\chi=\frac{1}{4}$ [Fig. 6(b)] is the appearance of the cusp and range axis intersections (features ''D'' and ''E'' from Fig. 3). The coordinates of these features are included in Table V. Beyond ''E'' the caustic can be seen to extend deep into the sediment. Our ray path analysis is inapplicable in this region because we assume that source and receiver heights are both positive, but no such restriction applies to the normal mode calculation. The classic Airy function behavior through a caustic is illustrated most clearly here: to the left of the caustic is a fringe pattern caused by interference between the two caustic-grazing rays; to the right the field decays exponentially into

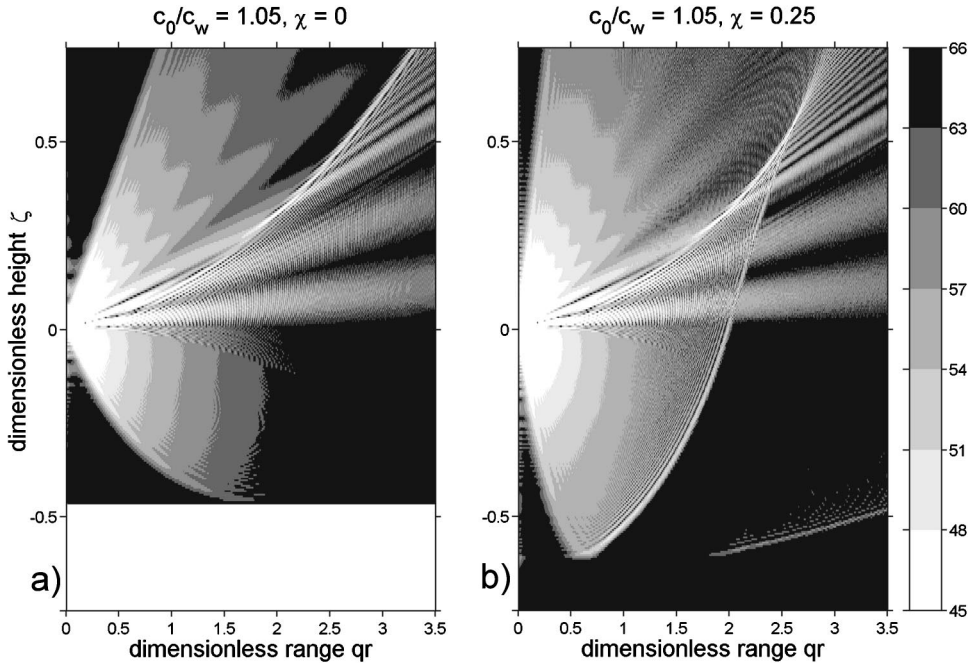


FIG. 7. As in Fig. 5 for $c_0/c_w=1.05$. The caustic envelope can be compared with the magenta curves from Fig. 3 for corresponding χ values.

TABLE VI. Measured coordinates of the caustic features from Fig. 4 ($n=+3$).

$(c_0/c_w)_T$	$r_A(q_T^{-1})$	$h_A(q_T^{-1})$	$r_B(q_T^{-1})$	$r_P(q_T^{-1})$
0.95	⋯	⋯	0.639	1.491
1.00	⋯	⋯	0.000	1.248
1.05	1.672	0.244	⋯	⋯
1.10	2.212	0.459	⋯	⋯

the shadow where there are no classical ray paths.

Now consider the case with $c_0/c_w=0.95$ (Fig. 5). The main effect on the caustic of reducing the sediment sound speed is that the left branch now intersects the range axis at "B" instead of at the origin. This is a direct result of the downward refraction at the boundary, causing the shadow for $r<r_B$. Notice the fringe pattern in the direct path region, caused by interference from a partial reflection at the water–sediment boundary. Notice also the multiples of the main refracted path in the sediment, caused by partial internal reflection inside the sediment layer. See Ref. 23 for a discussion of the reflected field due to these multipaths.

Finally for the $c_0/c_w=1.05$ case (Fig. 7) the main new feature is the appearance of a second cusp, resulting in a third caustic branch extending from "A" to plus infinity. Thus the energy is more thinly spread and this explains why this third branch is much weaker than the other two. The details of this third branch will in any case be sensitive to the sound speed profile in the water, not considered here, because of the low grazing angles involved. Notice the extension of the Lloyd mirror fringes almost to the origin for this case, because of the total internal reflection that occurs for grazing angles less than the water–sediment critical angle. The caustic nevertheless remains clearly visible.

V. POTENTIAL APPLICATION TO INVERSION

We show above that the position and shape of caustics induced by sediment refraction are sensitive to the sediment sound speed profile parameters. It follows in principle that, given a sufficiently accurate measurement of the caustic geometry, one can invert for these sediment parameters. For the measurement we have in mind a geometry involving a bottom- (or surface-) mounted vertical line array and a towed source, similar to that used by Frisk *et al.*[6] or Siderius *et al.*[24] Although such inversion is outside the present scope, it is seen as the main application of our analysis, and it is therefore considered helpful to provide a demonstration of a simple method to extract geoacoustic information. The demonstration comprises the following three steps.

Step 1: Synthesis of a measurement of the caustic shape. First the caustic shape is calculated for an arbitrary sediment profile. This synthesized caustic represents a real-world measurement and is used as input for the inversion. For this demonstration we choose the GPL_{+3} caustics of Fig. 4 (bottom right), the coordinates of which are presented in Table VI.

Step 2: Inversion of the caustic shape for the sediment profile parameters. The next step is to find a combination of geoacoustic parameters that matches the "measured" caustic shape, using one of the GPL_n profiles. The value of n can be considered as another unknown parameter, to be inferred by finding the best match to the measured caustic shape. This can be any real (integer or fractional) value of n. The alternative is to choose a particular GPL family by fixing n. For the purpose of our demonstration we choose to fix $n=+1$ because of the simple, easily invertible, equations for the caustic coordinates associated with this profile shape. The unknown parameters to be found are thus q, c_0/c_w, and χ.

First we consider the situation $c_0/c_w>1$ such that the cusp feature "A" exists (see Fig. 3). The value of the sound speed ratio and the gradient q follow from the coordinates of this feature [Eq. (31)]:

$$q=\frac{32(h_A+h_0)}{\sqrt{27}r_A^2}, \tag{46}$$

$$\frac{c_0}{c_w}=1+\frac{16(h_A+h_0)^2}{27r_A^2}. \tag{47}$$

For a slow sediment ($c_0/c_w<1$) the feature "A" is absent. Instead one can measure (in principle) the intersection of the caustic with the ζ axis (feature "B"), but the coordinates of "B" give only one equation for two unknowns. Equation (29) can nevertheless be inverted for q by noting that μ is a function of r/r_B only:

$$q=\frac{32(h_P+h_0)}{r_P^2(1+\sqrt{1+8(r_B/r_P)^2})^{1/2}(3-\sqrt{1+8(r_B/r_P)^2})^{3/2}}, \tag{48}$$

where r_P and h_P are the coordinates of a second point "P" along the caustic. To ensure the small angle assumption is satisfied, the point "P" must be reasonably close to "B," but its precise position is not critical. Assuming a GPL_{+1} profile, it then follows from Eq. (32) that

$$\frac{c_0}{c_w}=1-\frac{(qr_B)^2}{8}. \tag{49}$$

Finally, if the feature "E" exists the curvature parameter χ can be inferred through Eq. (43) (for $n=+1$):

$$\chi=\frac{1}{(qr_E)^2}. \tag{50}$$

Feature "E" exists for all four caustics of the selected graph, so the range r_E is used to determine χ. For slow sediments the value chosen for h_P is $0.1\,q_T^{-1}$. (For the present purpose, the case involving $r_B=0$ is treated as having a slow sediment.) The range of feature "E" ($r_E=1.754\,q_T^{-1}$) is taken from Table V. The result of the inversion is shown in Table VII. The subscript "*T*" indicates the true value of a parameter. In order to keep the results as general as possible, all distances are expressed in units of q_T^{-1}.

Step 3: Reconstruction of the sediment sound speed profile. The final step in the demonstration is to reconstruct the sound speed profile using the inverted parameters from Table VII. The result is plotted in Fig. 8, together with the original GPL_{+3} profiles from which the Fig. 4 caustics were calcu-

TABLE VII. Seabed parameters found by inversion of the caustic coordinates.

$(c_0/c_w)_T$	c_0/c_w	q/q_T	χ
0.95	0.944	1.051	0.294
1.00	1.000	1.027	0.308
1.05	1.050	1.075	0.281
1.10	1.102	1.155	0.243

lated. The accuracy of the inverted profiles (to a depth of $0.8\,q_T^{-1}$) can be judged by inspection of Fig. 8. At greater depths the errors grow rapidly (not shown).

Notice that the sound speed gradient is overestimated for all four profiles, but this is compensated for by underestimating the curvature, in such a way that the average gradient to $0.8\,q_T^{-1}$ remains approximately correct. The largest sound speed error (from Fig. 8) for depths up to $0.8\,q_T^{-1}$ occurs for $c_0/c_w=1.1$, at a depth of $0.64\,q_T^{-1}$ and is approximately $0.12\,c_w$. This error is caused partly by the approximate nature of Eqs. (31) and (32) and partly by the mismatch in the assumed value of n.

A possible objection is that the approach seems to rely on measurement of particular caustic features, some of which (feature "E," for example) may not even exist. We have also assumed infinite measurement accuracy and resolution. However the above demonstration is intended not as a realistic inversion, but as an illustration of the information content of the caustic envelope, due to its strong sensitivity to the seabed parameters. We imagine that best results can be obtained by matching the pressure field in the vicinity of the caustic using broadband matched field processing.[25]

One can also question the extent to which the shape of GPL_n profiles matches measured ones. A visual comparison of our Fig. 1 with Hamilton's data[1] suggests that there is considerable scope for a good match. For example, Robins[12] fits a GPL_{+1} profile to Hamilton's Fig. 8 of sound speed versus depth beneath the sea floor for terrigenous sediments and rocks. The difference between Hamilton's curve and the GPL fit (not shown) is considerably smaller than the likely measurement error.

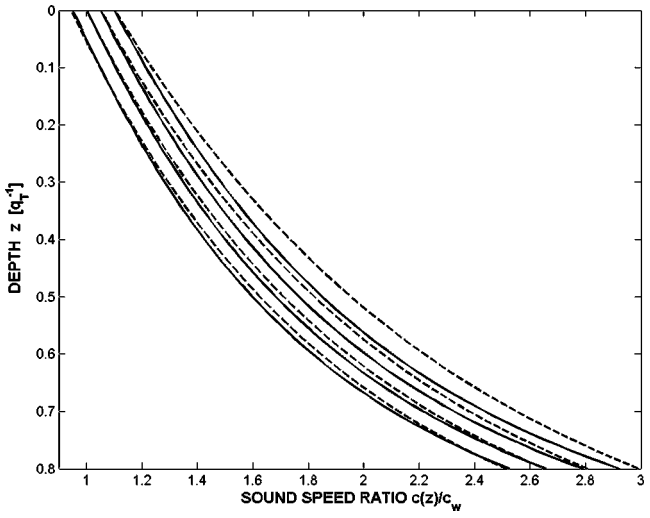


FIG. 8. True GPL_{+3} profiles used for Fig. 4 (solid lines) and reconstructed profiles obtained by inversion assuming $n=+1$ (dashed lines).

A particular functional form sometimes used to characterize realistic sediment sound speed profiles, known as the Bottom Loss Upgrade (or "BLUG") profile,[26,27] is described by McCammon.[28] The BLUG profile has similar properties to GPL_{+2}, the two profiles being identical for the special case $\chi=-1/3$. The additional degree of freedom associated with the GPL profiles (through the choice of n) means that GPL is likely to be more flexible than BLUG. Nevertheless, the caustic envelope can be calculated for the BLUG profile if desired, using the present method, from the expression for the ray cycle distance $x(\theta_0)$.[29]

VI. CONCLUSIONS

We derive ray properties for a family of sediment sound speed profiles (called "generalized power law" or GPL profiles) given by Eq. (2). The GPL profiles permit continuous variation of sound speed curvature, through a parameter denoted χ, for a given value of surface sound speed gradient. Additional control over profile shape is achieved through choice of the integer n, such that a good fit is possible to measured profiles from (e.g.) Hamilton's data. The GPL profile for $n=-2$ (denoted GPL_{-2}) is a special case that reduces trivially to the "linear k^2" profile $c(z)=c_0(1-2qz)^{-1/2}$.

Our main new results, illustrated by Figs. 3 and 4, are a set of equations for predicting caustic envelopes for arbitrary sediment sound speed curvature. The complete method is described in Sec. III A. These equations, with the caustic angle as a parameter, are confirmed in Sec. IV by comparison with a normal mode calculation of the pressure field. Also derived in Sec. III A are approximate explicit equations, valid for small angles, for the caustic envelope [Eq. (25)] and other important features [Eqs. (27) and (28)]. For a few special cases, some exact calculations are derived for cusp coordinates and range axis intersections (Secs. III B and C).

The most easily observable caustic features are those caused by rays close to grazing incidence. The shape of these features is shown to be determined primarily by the sound speed ratio c_0/c_w, and their positions are also sensitive to the gradient parameter q. The curvature parameter χ determines the existence and location of features (cusps and intersections with the range axis) associated with steeper angles. Conversely, a measurement of the position of such features can be inverted for the profile parameters.

[1] E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[2] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980).
[3] D. C. Stickler, "Negative bottom loss, critical-angle shift, and the interpretation of the bottom reflection coefficient," J. Acoust. Soc. Am. **61**, 707–710 (1977).
[4] C. W. Spofford, "Inference of geo-acoustic parameters from bottom-loss data," in *Bottom Interacting Ocean Acoustics*, edited by W. Kuperman and F. Jensen (Plenum, New York, 1980), pp. 159–171.
[5] G. V. Frisk, "Determination of sediment sound speed profiles using caustic range information," in Ref. 4, pp. 153–157.
[6] G. V. Frisk, J. A. Doutt, and E. E. Hays, "Bottom interaction of low-

frequency acoustic signals at small grazing angles in the deep ocean,'' J. Acoust. Soc. Am. **69**, 84–94 (1981).
[7] Yu. I. Orlov, ''Some features of ray theory of propagation of cylindrical and spherical waves in weakly inhomogeneous plasma,'' Sov. Radiophys. **9**, 307–311 (May–June, 1966) [Izv. VUZ. Radiofizi. **9**, 497–506 (1966)].
[8] Yu. I. Orlov, ''The problem of wave propagation through a caustic in a parabolic plasma layer,'' Izv. Vyssh. Uchebn. Zaved., Radiofiz. **9**, 1036–1038 (1966); in Russian.
[9] M. A. Ainslie, ''Caustics and beam displacements due to the reflection of spherical waves from a layered half-space,'' J. Acoust. Soc. Am. **96**, 2506–2515 (1994).
[10] I. Tolstoy and C. S. Clay, *Ocean Acoustics: Theory and Experiment in Underwater Sound* (McGraw–Hill, New York, 1966).
[11] S. M. Rytov and F. S. Yudkevich, ''The reflection of electromagnetic waves from a layer with a negative dielectric constant,'' Zh. Eksp. Teor. Fiz. **10**, 887–902 (1946); in Russian. See p. 190 of Ref. 2.
[12] A. J. Robins, ''Reflection of a plane wave from a fluid layer with continuously varying density and sound speed,'' J. Acoust. Soc. Am. **89**, 1686–1696 (1991).
[13] M. A. Ainslie, R. M. Hamson, G. D. Horsley, A. R. James, R. A. Laker, M. A. Lee, D. A. Miles, and S. D. Richards, ''Deductive multi-tone inversion of seabed parameters,'' J. Comput. Acoust. **8**, 271–284 (2000).
[14] M. J. Buckingham, ''On acoustic transmission in ocean-surface waveguides,'' Philos. Trans. R. Soc. London, Ser. A **335**, 513–555 (1991).
[15] M. J. Buckingham, ''Sound speed and void fraction profiles in the sea surface bubble layer,'' Appl. Acoust. **51**, 225–250 (1997).
[16] *Table of Integrals, Series, and Products*, edited by I. S. Gradshteyn and I. M. Ryzhik (Academic Press, New York, 1980).
[17] M. A. Ainslie, ''Stationary phase evaluation of the bottom interacting field in isovelocity water,'' J. Acoust. Soc. Am. **94**, 1496–1509 (1993); erratum **95**, 3670 (1994).
[18] J. W. Archbold, *Algebra* (Pitman, London, 1970).
[19] E. L. Hamilton, ''Compressional-wave attenuation in marine sediments,'' Geophysics **37**, 620–645 (1972).
[20] A. C. Kibblewhite, ''Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data,'' J. Acoust. Soc. Am. **86**, 716–738 (1989).
[21] C. W. Holland and P. Neumann, ''Sub-bottom scattering: A modeling approach,'' J. Acoust. Soc. Am. **104**, 1363–1373 (1998).
[22] C. L. Morfey, *Dictionary of Acoustics* (Academic, San Diego, 2001).
[23] M. A. Ainslie and D. E. Weston, ''Reflection with beam displacement from simple barriers,'' Acta Acust. **82**, 607–614 (1996).
[24] M. Siderius, M. Snellen, D. G. Simons, and R. Onken, ''An environmental assessment in the Strait of Sicily: Measurement and analysis techniques for determining bottom and oceanographic properties,'' J. Oceanic Eng. **25**, 364–386 (2000).
[25] *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. A. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht, 1995).
[26] P. C. Etter, *Underwater Acoustic Modeling: Principles, Techniques and Applications* (E & F N Spon, London, 1996).
[27] M. A. Ainslie, A. J. Robins, and M. K. Prior, ''Benchmark solutions of plane wave bottom reflection loss,'' J. Acoust. Soc. Am. **104**, 3305–3312 (1998); erratum **113**, 2180 (2003).
[28] D. F. McCammon, ''Fundamental relationships between geoacoustic parameters and predicted bottom loss using a thin layer model,'' J. Geophys. Res. [Atmos.] **93**, 2363–2369 (1988).
[29] E. K. Westwood and P. J. Vidmar, ''Eigenray finding and time series simulation in a layered-bottom ocean,'' J. Acoust. Soc. Am. **81**, 912–924 (1987).

N-shell cluster in water: Multiple scattering and splitting of resonances

Pierre-Yves Le Bas, Francine Luppé,[a] Jean-Marc Conoir, and Hervé Franklin
LAUE, UMR CNRS 6068, Université du Havre, place Robert Schuman, 76610 Le Havre, France

(Received 18 April 2003; revised 25 January 2004; accepted 1 February 2004)

Clusters of *N* thin parallel and identical shells (aligned or not) in water are considered. Assuming a harmonic plane wave is normally incident upon one cluster, the scattered-field classical expression is recalled, and then computed for different types of clusters, along with resonance spectra. The scattering *S* matrix is defined, and its unitarity property used to check the numerical results. All spectra are compared with that of a single shell, in the frequency range where resonances are due to an *A*-wave phase matching only. Whatever the cluster, each resonance of the single shell is seen to split into *M* different ones. The value of *M* depends on the number of shells, the distance between them, and the symmetries of the cluster. Apart from the very special case of aligned shells ($M=2N$), no simple law has been found to predict the value of *M*. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1689345]

PACS numbers: 43.20.Fn, 43.20.Ks [GCG] Pages: 1460–1467

I. INTRODUCTION

Analytic calculation of the acoustic field scattered by *N* infinitely long elastic cylinders in water may be performed, thanks to the Graff addition theorem,[1] and the method, indeed, is well known since Foldy's early work.[2] As the number *N* of scatterers increases, however, the numerical computation of that field becomes more and more time consuming, and other models have to be used. These models provide, under different assumptions of the statistical properties of the scatterers, the average properties of the transmitted field.[3] For identical parallel cylinders, averaging is done on all their possible positions, at a fixed density (number of cylinders per unit cross section). The only information needed, then, is the statistics of the locations of the cylinders, and the acoustic frequency response of one cylinder to an incident plane wave, either in the far field[1,3,4] only, or by means of its transition matrix.[5–7] One way to introduce symmetries in the medium, or strong correlations between scatterers, is to consider clusters of close scatterers as a single scatterer, so that the transition matrix to be known, now, is that of each cluster. As it is rather intuitive that the average medium properties should vary more strongly with frequency in the vicinity of a scatterer resonance,[6–8] this paper is dedicated to the determination of the resonances of different clusters composed of parallel cylindrical shells (aligned or not).

There are relatively few works on the subject of resonances by clusters of cylindrical or spherical shells. Huang and Gaunaurd[9] have studied the resonant behavior of two spherical shells and shown that the low-frequency resonances of a single shell are shifted downwards as the distance between the two shells is decreased. Decanini *et al.*[10,11] have developed the Korringa–Kohn–Rostoker–Berry method to determine the scattering matrix of *N* disks with various boundary conditions. Taking into account the symmetries of the system leads them to classify the resonances in distinct families associated with the irreducible representations of the symmetry group considered.

This work extends those of Kheddioui[12,13] and Lethuillier,[14,15] who studied linear arrays of thin, empty, aluminum cylindrical shells. The resonances of such shells are well known.[16–18] In the low-frequency range, the well-separated ones are due to an external wave, which means that most of its energy is in the fluid. This wave is a bending wave, usually referred to as the Scholte–Stoneley *A* wave. In the high-frequency range, the resonances are due to an S_0 type of wave, the energy of which is mostly in the solid (it is hence called an internal wave). This is the main reason for which we study the low-frequency range, as the external *A* wave may couple close enough adjacent shells,[14] while the internal S_0 wave never does.[12,13]

Throughout this paper, *N* infinitely long parallel and identical shells, of known positions in water, are considered. In Sec. II, the incident field is a harmonic plane wave that propagates in a direction normal to the axes of the cylinders. The method to determine the total scattered field is recalled, and most of the notations needed are introduced meanwhile. The scattering matrix elements of the cluster are derived in Sec. III, in a simpler way than in Ref. 10. As the *S* matrix is nondiagonal, the multichannel resonant scattering theory[19] should be used to describe the resonant behavior of the cluster. Yet, this is not done in that paper; only the frequency positions of the resonances are looked for, by use of properties of the Argand diagram in the vicinity of a resonance.[20] This is done in Sec. IV, along with a comparison of other results.[11]

II. ACOUSTIC SCATTERING BY *N* CYLINDERS

All *N* cylinders are infinitely long in the Oz direction. The incident harmonic plane wave propagates in the (xOy) plane, so that the problem is a two-dimensional one, as shown in Fig. 1, and polar coordinates are used. The observation point is $P(r,\theta)$. The hth cylinder is centered at $O_h(d_h,\chi_h)$. In the local coordinates system of cylinder

[a]Electronic-mail: luppe@univ-lehavre.fr

 0001-4966/2004/115(4)/1460/8/$20.00

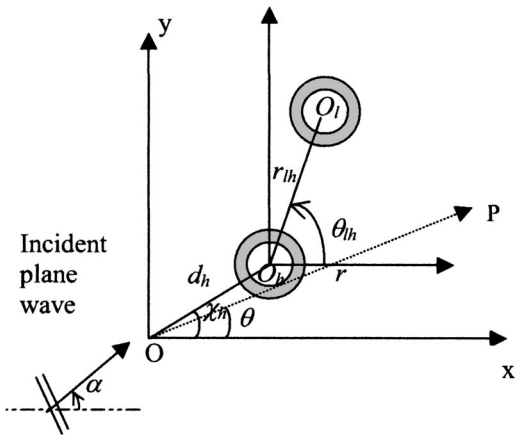


FIG. 1. Scattering of an incident plane wave by an N-shell cluster ($N=2$ here). Geometry of the problem.

number h, the observation point coordinates are (r_h, θ_h). The incident plane wave, in the main coordinates system centered at O, is

$$\phi_{\text{inc}}^{(0)}(\alpha,\theta)=e^{\mathbf{ik}\cdot\mathbf{r}}=e^{ikr\cos\alpha\cos\theta}e^{ikr\sin\alpha\sin\theta}=\sum_{n=-\infty}^{+\infty} i^n J_n(kr)e^{in(\theta-\alpha)}, \tag{1}$$

with α the incidence direction as in Fig. 1, and an assumed time dependence $e^{-i\omega t}$. In Eq. (1), $k=\omega/c$ with c the sound velocity in the fluid, and $J_n(kr)$ is the Bessel function of order n and argument kr. In the local coordinates system of cylinder h, this incident plane wave is

$$\phi_{\text{inc}}^{(h)}=e^{ikd_h\cos(\chi_h-\alpha)}\sum_{n=-\infty}^{+\infty} i^n e^{-in\alpha}J_n(kr_h)e^{in\theta_h}. \tag{2}$$

The wave scattered by cylinder h is

$$\phi_s^{(h)}=\sum_{n=-\infty}^{+\infty} C_n^{(h)}H_n^{(1)}(kr_h)e^{in\theta_h}, \tag{3}$$

with $C_n^{(h)}$ unknown coefficients that depend on the incident angle α and $H_n^{(1)}$ the Hankel function of the first kind.

The incident wave on cylinder h is the summation of the incident plane wave and of the waves scattered by the other cylinders

$$\phi^{(h)}=e^{ikd_h\cos(\chi_h-\alpha)}\sum_{n=-\infty}^{+\infty} i^n e^{-in\alpha}J_n(kr_h)e^{in\theta_h}+\sum_{\substack{l=1\\ l\neq h}}^{N}\sum_{n=-\infty}^{+\infty} C_n^l H_n^{(1)}(kr_l)e^{in\theta_l}. \tag{4}$$

The wave scattered by cylinder h may be considered as an incident wave on cylinder l, thanks to the Graff's addition theorem ($r_l<r_{lh}$ with r_{lh} the O_lO_h distance as in Fig. 1)

$$H_n^{(1)}(kr_h)e^{in\theta_h}=\sum_{m=-\infty}^{+\infty}(-1)^{m-n}e^{i(n-m)\theta_{lh}}\times H_{m-n}^{(l)}(kr_{lh})J_m(kr_l)e^{im\theta_l}=\sum_{m=-\infty}^{+\infty} G_{nm}^{hl}J_m(kr_l)e^{im\theta_l}, \tag{5}$$

with θ_{lh} defined in Fig. 1 and

$$G_{nm}^{hl}=(-1)^{m-n}e^{i(n-m)\theta_{lh}}H_{m-n}^{(l)}(kr_{lh}). \tag{6}$$

This yields the following expression of the response of cylinder h to the incident wave described in Eq. (4):

$$\phi_s^{(h)}=\sum_{m=-\infty}^{+\infty} T_{mm}^{(h)}\left[e^{ikd_h\cos(\chi_h-\alpha)}i^m e^{-im\alpha}+\sum_{\substack{l=1\\ l\neq h}}^{N}\sum_{n=-\infty}^{+\infty} G_{nm}^{lh}C_n^{(l)}\right]H_m^{(1)}(kr_h)e^{im\theta_h}, \tag{7}$$

with $T_{mm}^{(h)}$ the mth diagonal element of cylinder h diagonal transition matrix.[5] Together with Eq. (3), Eq. (7) provides the linear system of the unknown elements $C_n^{(h)}$ to be solved

$$C_m^{(h)}=T_{mm}^{(h)}\left[e^{ikd_h\cos(\chi_h-\alpha)}i^m e^{-im\alpha}+\sum_{\substack{l=1\\ l\neq h}}^{N}\sum_{n=-\infty}^{+\infty} C_n^{(l)}G_{nm}^{lh}\right]. \tag{8}$$

The total scattered field Φ_s is the sum of all scattered fields

$$\Phi_s=\sum_{h=1}^{N}\phi_s^{(h)}=\sum_{h=1}^{N}\sum_{n=-\infty}^{+\infty} C_n^{(h)}H_n^{(1)}(kr_h)e^{in\theta_h}. \tag{9}$$

Introduction of the Hankel function asymptotic development for $r\to\infty$ in Eq. (9) gives

$$\Phi_{s\infty}=\sqrt{\frac{2}{\pi kr}}e^{i(kr-(\pi/4))}\times\sum_{h=1}^{N}\sum_{n=-\infty}^{+\infty} i^{-n}C_n^{(h)}e^{-ikd_h\cos(\chi_h-\theta)}e^{in\theta}=\frac{e^{i(kr-(\pi/4))}}{\sqrt{kr}}f(\theta\leftarrow\alpha). \tag{10}$$

The angle-dependent function

$$f(\theta\leftarrow\alpha)=\sqrt{\frac{2}{\pi}}\sum_{h=1}^{N}\sum_{n=-\infty}^{+\infty} i^{-n}C_n^{(h)}e^{-ikd_h\cos(\chi_h-\theta)}e^{in\theta}, \tag{11}$$

is the far-field scattered amplitude[21] of the system of N cylinders in the θ direction, provided a given incidence angle α.

For the computations, the series has to be truncated at some order $n=n_1$, where n_1 is the lowest value of n_0 that ensures that both differences $|\Sigma_{n_0}-\Sigma_{n_0-1}|$ and $|\Sigma_{n_0-1}-\Sigma_{n_0-2}|$ are less than 0.1% of $|\Sigma_{n_0}|$. The symbol Σ_{n_0} is defined as

$$\Sigma_{n_0}=\sum_{i=1}^{N}\sum_{n=-n_0}^{n_0}C_n^{(h)}. \tag{12}$$

III. THE SCATTERING MATRIX

A. The $\hat{S}$ operator

Generalization of the resonance scattering theory (RST)[22] to the cluster case needs the definition of the *S*-scattering matrix of the cluster. Contrary to the one-shell case, mode conversions occur, as may be seen, from example, in Eq. (8), and, consequently, the infinite *S* matrix will not be diagonal. This is the reason why a resonance theory on the cluster should be based on the multichannel resonant scattering theory,[19] as mentioned already in the Introduction of that paper.

In the following, a method derived from quantum mechanics[23] is used to express the $\hat{S}$ operator associated with the scattering matrix. The *S*-matrix elements are then derived. This construction of the *S* matrix is different from that given in Ref. 10. It is obtained in a simpler way and, in addition, it allows easier computations, especially when testing the *S*-matrix unitarity.

Let the incident field $\Phi_{\text{inc}}(\theta)$ now be an angular spectrum of incident plane waves

$$\Phi_{\text{inc}}(\theta)=\frac{1}{2\pi}\int_0^{2\pi}F(\alpha)\phi_{\text{inc}}^{(0)}(\alpha,\theta)d\alpha, \tag{13}$$

with $F(\alpha)$ a 2π-periodic function of α, so that the resulting scattered far field of Eq. (10) is

$$\Phi_{\text{scat}}=\frac{1}{2\pi}\frac{e^{i(kr-(\pi/4))}}{\sqrt{kr}}\left(\int_0^{2\pi}F(\alpha)f(\theta\leftarrow\alpha)d\alpha\right). \tag{14}$$

Introduction of the $\hat{S}$ scattering operator follows from the decomposition of the incident field into ingoing and outgoing waves. The far-field expression of the incident plane wave $\phi_{\text{inc}}^{(0)}(\alpha,\theta)$ of Eq. (1) is

$$\phi_{\text{inc}}^{(0)}=\frac{1}{\sqrt{2\pi}}\left[\sum_{n=-\infty}^{+\infty}e^{in(\theta-\alpha)}\frac{e^{i(kr-(\pi/4))}}{\sqrt{kr}}+\sum_{n=-\infty}^{+\infty}(-1)^n e^{in(\theta-\alpha)}\frac{e^{-i(kr-(\pi/4))}}{\sqrt{kr}}\right]. \tag{15}$$

Introduction of the Fourier coefficients F_n defined by

$$F_n=\langle F(\alpha)|e^{-in\alpha}\rangle=\frac{1}{2\pi}\int_0^{2\pi}F(\alpha)e^{-in\alpha}d\alpha, \tag{16}$$

with

$$F(\alpha)=\sum_{n=-\infty}^{+\infty}F_n e^{in\alpha}, \tag{17}$$

leads to the following expression of the angular spectrum introduced in Eq. (13):

$$\Phi_{\text{inc}}(\theta)=\frac{1}{\sqrt{2\pi}}\left[\sum_{n=-\infty}^{+\infty}F_n e^{in\theta}\frac{e^{i(kr-(\pi/4))}}{\sqrt{kr}}+\sum_{n=-\infty}^{+\infty}F_n(-1)^n e^{in\theta}\frac{e^{-i(kr-(\pi/4))}}{\sqrt{kr}}\right], \tag{18}$$

and then to

$$\Phi_{\text{inc}}(\theta)=\frac{1}{\sqrt{2\pi}}\left[F(\theta)\frac{e^{i(kr-(\pi/4))}}{\sqrt{kr}}+F(\theta-\pi)\frac{e^{-i(kr-(\pi/4))}}{\sqrt{kr}}\right]. \tag{19}$$

The total field, $\Phi_{\text{inc}}+\Phi_{\text{scat}}$, is then

$$\Phi_{\text{tot}}(\theta)=\frac{1}{\sqrt{2\pi}}F(\theta-\pi)\frac{e^{-i(kr-(\pi/4))}}{\sqrt{kr}}+\frac{1}{\sqrt{2\pi}}\times\left[F(\theta)+\frac{1}{\sqrt{2\pi}}\int_0^{2\pi}F(\alpha)f(\theta\leftarrow\alpha)d\alpha\right]\times\frac{e^{i(kr-(\pi/4))}}{\sqrt{kr}}. \tag{20}$$

From this equation, the transition operator $\hat{T}$ acting on the 2π-periodic F function is defined as

$$\hat{T}F(\theta)=\frac{1}{2i\sqrt{2\pi}}\int_0^{2\pi}f(\theta\leftarrow\alpha)F(\alpha)d\alpha, \tag{21}$$

and the $\hat{S}$ operator as

$$\hat{\mathrm{S}}=\mathbb{1}+2i\hat{T}, \tag{22}$$

with $\mathbb{1}$, the identity operator. The total field may then be written as

$$\Phi_{\text{tot}}=\frac{1}{\sqrt{2\pi}}F(\theta-\pi)\frac{e^{-i(kr-(\pi/4))}}{\sqrt{kr}}+\frac{1}{\sqrt{2\pi}}\hat{\mathrm{S}}F(\theta)\frac{e^{i(kr-(\pi/4))}}{\sqrt{kr}}. \tag{23}$$

The first part of the right-hand side in Eq. (23) represents ingoing waves, while the second one represents outgoing waves.

The $\hat{S}$ operator characterizes the scattering by the *N* cylindrical shells; it is independent of the conditions of observation and of the incident wave features, except for the frequency. For computation needs, the *S*-matrix elements in the *N* cylindrical shell case are derived in the next section.

B. The *S* matrix

In order to characterize the *S* matrix in the frame of modal theories, a projection of the $\hat{S}$ operator on the orthonormal basis of the 2π-periodic functions

$$F_p(\alpha)=e^{ip\alpha},\quad p\in\mathbb{Z}, \tag{24}$$

must be done. Equation (25) gives the S_{pq} elements of this matrix

$$\hat{S}F_p(\theta)=\sum_{q=-\infty}^{+\infty} S_{pq}e^{iq\theta}. \tag{25}$$

Equations (22), (21), (11), and (16) yield

$$\begin{aligned}\hat{S}F_p(\theta)&=e^{ip\theta}+2\sum_{h=1}^{N}\sum_{n=-\infty}^{+\infty} e^{-ikd_h\cos(\chi_h-\theta)}(-i)^n\\&\quad\times\langle C_n^{(h)}|e^{ip\alpha}\rangle e^{in\theta}\\&=e^{ip\theta}+2\sum_{h=1}^{N}\sum_{n=-\infty}^{+\infty}\sum_{q=-\infty}^{+\infty} J_{q-n}(kd_h)\\&\quad\times e^{-i(q-n)\chi_h}i^{-q}\langle C_n^{(h)}|e^{ip\alpha}\rangle e^{iq\theta},\end{aligned} \tag{26}$$

so that the *S*-matrix elements are

$$\begin{aligned}S_{pq}&=\delta_{pq}\\&\quad+2\sum_{h=1}^{N}\sum_{n=-\infty}^{+\infty}\langle C_n^{(h)}|e^{ip\alpha}\rangle i^{-q}e^{-i(q-n)\chi_h}J_{q-n}(kd_h),\end{aligned} \tag{27}$$

and those of the *T* matrix are

$$T_{pq}=\sum_{h=1}^{N}\sum_{n=-\infty}^{+\infty}\langle C_n^{(h)}|e^{ip\alpha}\rangle i^{-q-1}e^{-i(q-n)\chi_h}J_{q-n}(kd_h). \tag{28}$$

The scalar products $\langle C_n^{(h)}|e^{ip\alpha}\rangle$, already defined in Eq. (16) with F in place of $C_n^{(h)}$, are solutions of the following linear system, similar to system (8):

$$\begin{aligned}\langle C_m^{(h)}|e^{ip\alpha}\rangle&=T_{mm}^{(h)}\Bigg[i^{2m-p}e^{-i(m-p)\chi_h}J_{m-p}(kd_h)\\&\quad+\sum_{\substack{l=1\\l\neq h}}^{N}\sum_{n=-\infty}^{+\infty}G_{nm}^{lh}\langle C_n^{(l)}|e^{ip\alpha}\rangle\Bigg].\end{aligned} \tag{29}$$

Verification of the *S*-matrix unitarity has been done to check the validity of all the numerical results presented in the next section. The actual considered size of the *S* matrix is determined from the value of that order n_1 at which the infinite summation, for $p=0$ in Eq. (27), can be truncated, the same way as explained in Sec. II. The *S*-matrix unitarity is then obtained with a 10^{-4} % precision for all cases under study.

As already stated in the Introduction, the *T*-matrix elements, given by Eq. (28), are also necessary if one has to use an effective medium theory such as the ones developed in Refs. 5–7.

IV. NUMERICAL RESULTS

The results correspond to clusters composed of identical air-filled aluminum shells in water. Aluminum is characterized by a density $\rho=2790$ kg/m^3, a longitudinal velocity $c_l=6120$ m/s, and a shear velocity $c_s=3020$ m/s. The b/a ratio, of the inner radius to the outer one, is equal to 0.9.

The form function F_∞ of a single shell is

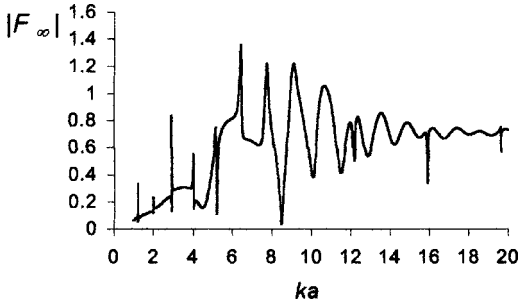


FIG. 2. Modulus of the form function of one single shell in the backscattering geometry $\theta=\alpha+\pi$.

$$F_\infty=\sqrt{\frac{2}{\pi ka}}\,f(\alpha+\pi\leftarrow\alpha). \tag{30}$$

Its modulus is plotted in Fig. 2. The sharp variations in this figure are due to resonances, i.e., circumferential waves that circumnavigate the shell with such a phase velocity that an integer number of wavelengths fits over its circumference.[24] The exact frequency resonances are found from the observation of the resonance spectrum plotted in Fig. 3. It is a plot, versus ka, of the frequency derivative ds/dx of the curvilinear abscissa s of the form-function Argand diagram (parametric plot in frequency of the form-function real part versus its imaginary part)[20]

$$\frac{ds}{dx}=\sqrt{\left(\frac{dF'_\infty}{dx}\right)^2+\left(\frac{dF''_\infty}{dx}\right)^2}, \tag{31}$$

with $x=ka$, F'_∞ the real part of F_∞, and F''_∞ its imaginary part.

The low-frequency resonances in Fig. 3 are due to an *A* wave, while the high-frequency ones are due to an S_0 wave.[16,17] The *A* wave is an external wave,[16–18] which means that its energy is mostly carried in the fluid. This property allows it to couple close shells more easily than the internal S_0 wave,[12,13] so that each resonance in Fig. 3 will split into new ones in the case of a cluster of close enough shells. This split is shown in the next subsections.

A. Aligned shells

Kheddioui and Lethuillier[12–15] have studied equally spaced (finite) linear arrays of shells in the eclipse configuration. In this configuration, all the centers of the shells are

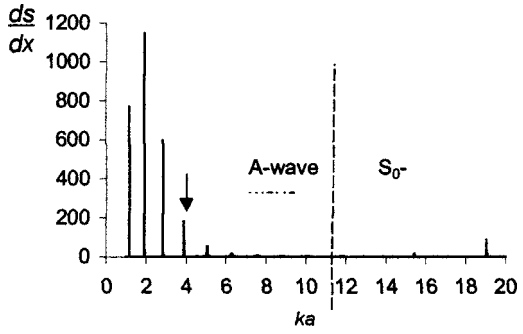


FIG. 3. Resonance spectrum of a single shell. The low-frequency resonances are due to the *A* wave, the high-frequency ones to the S_0 wave. The arrow points to that particular resonance that will be studied all along.

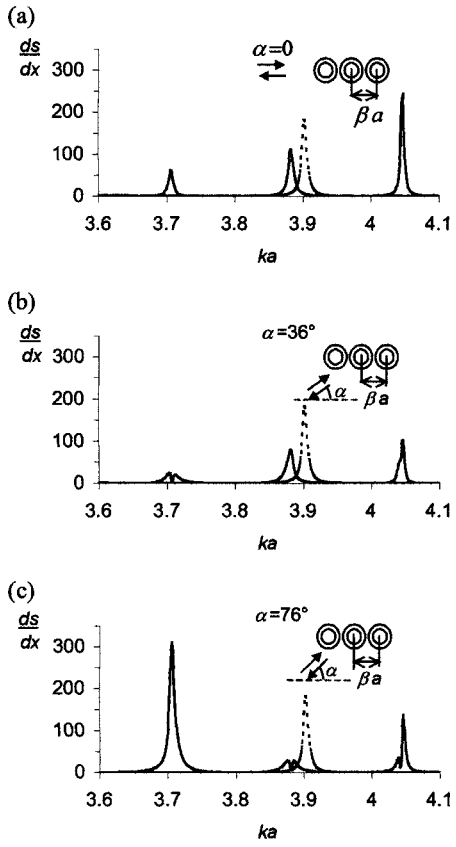


FIG. 4. Comparison of the resonance spectrum of one single shell (dashed line) and three aligned shells with $\beta=2.06$ for different values of the incidence angle α (solid line). (a) $\alpha=0°$; (b) $\alpha=36°$; (c) $\alpha=76°$.

aligned in the incidence direction, which means that $\chi_h=\alpha$ for each h. Let βa be the distance between the centers of neighboring shells. For large values of β, Kheddioui showed[12,13] that the sharp variations of the form function, that are all due to interference phenomena between scatterers and not to actual resonances, are not visible on a ds/dx versus x resonance spectrum. Lethuillier showed that for small β values, i.e., $\beta<2.3$, each resonance of a single shell splits into N resonances, with N the number of cylinders. This phenomenon occurs only for the A-wave resonances, not for the S_0-wave resonances. This is the reason why this study focuses only on the A-wave resonances. The following results show how the resonance that is pointed out by an arrow in Fig. 3 does split. The same phenomena, however, do occur for each A-wave resonance.

Figures 4(a)–(c) show resonance spectra obtained for three aligned shells, with $\beta=2.06$ in the 3.6–4.1-ka frequency domain. Figure 4(a) corresponds to $\alpha=0°$, i.e., to the eclipse configuration, as the centers of the shells are aligned on the x axis of Fig. 1. The single-shell resonance, plotted with a dashed line, is actually split into three different ones. Figures 4(b) and (c) correspond, respectively, to incidence angles α equal to 36° and 76°. It may be seen that each resonance observed in the eclipse configuration [$\alpha=0$, Fig. 4(a)] may also be split into two close ones, depending on the incidence direction. Each resonance of the single shell, then, may split into $2N$ different ones and not to N only, contrary

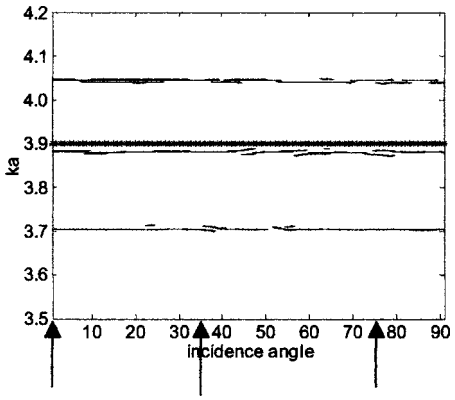


FIG. 5. Three aligned shells, $\beta=2.06$. Evolution of the resonance spectrum with the incidence angle and the reduced frequency. The stars describe the evolution of the one single-shell resonance under study. The arrows point on the incidence angle values corresponding to Figs. 4(a)–(c).

to what Lethuillier had found after a study limited to the eclipse configuration.

Figure 5 shows the evolution of the resonance spectrum with the incidence angle, for the same spatial configuration as in Figs. 4(a)–(c). Lines are drawn, in the (incidence angle, reduced frequency ka) plane, connecting those points for which there is a maximum of the resonance spectrum of the cluster. The three arrows point to the incidence angle values corresponding to Figs. 4(a)–(c). The stars represent the frequency position of the single-shell resonance under study ($ka\simeq3.9$). Its split into N ($N=3$) well-separated resonances ($ka\simeq3.7$, $ka\simeq3.87$, $ka\simeq4.04$) is clearly exhibited for most of the incidence angles, as well as the second split of each of them that occurs only at a few angles.

Similar results have been found, whatever the number N of aligned shells: each single-shell resonance gives rise to N well-separated ones, and each one of them may, in turn, be split into two overlapping resonances, depending on the incidence angle.

Figure 6 shows the evolution of the resonance spectrum with β, at a fixed incidence angle ($\alpha=18°$). The three well-separated resonances at low-β values merge into the single-shell resonance as β increases. The critical value of β corresponding to the achievement of the merging process depends on the incidence angle; whatever the incidence angle value, however, no split may be observed for β values greater than 7. This indicates that, at ka$\simeq$3.9, the distance between adja-

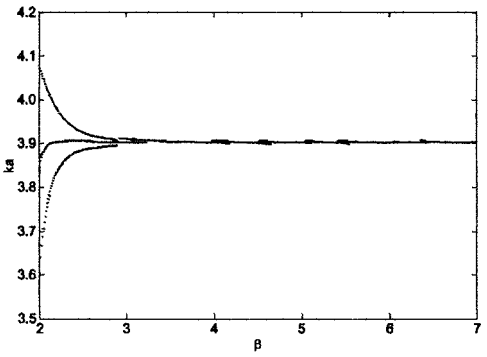


FIG. 6. Three aligned shells, incidence angle $\alpha=20°$. Evolution of the resonance spectrum with β and the reduced frequency.

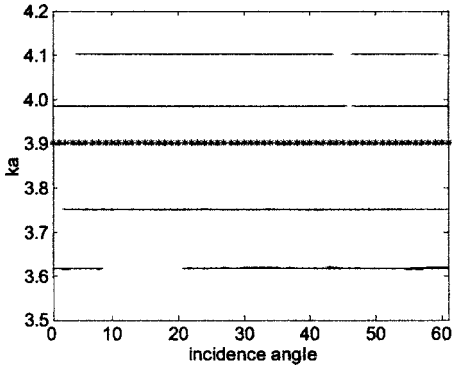


FIG. 7. Three shells in an equilateral triangle configuration, $\beta=2.06$. Evolution of the resonance spectrum with the incidence angle and the reduced frequency. The stars describe the evolution of the one single-shell resonance under study.

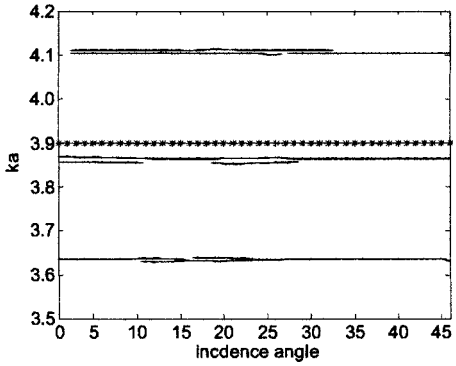


FIG. 8. Four shells in a square configuration, $\beta=2.06$. Evolution of the resonance spectrum with the incidence angle and the reduced frequency. The stars describe the evolution of the one single-shell resonance under study.

cent shells, when greater than $7a$, is too large, compared to the depth penetration in water of the *A* wave, for this wave to couple the shells.

The question now is: does the relation found between the number of shells and the number of resonances still hold in the case of nonaligned shells? The next section answers it.

B. Nonaligned shells

In this section, we study clusters of shells such that their centers are not aligned in the (x,y) plane, while their axes are still all parallel to the *z* direction (cf. Fig. 1).

Let us consider first three shells in an isosceles right-angled triangle arrangement, with a right-angle side $2.06a$ long. The evolution of the resonance spectrum with the incidence angle is not shown here, as it is quite similar to that of the three aligned shells (Fig. 5). This is not, however, a characteristic of a cluster of three shells, as may be seen in Fig. 7, which shows the evolution of the resonance spectrum when the three shells are located at the apexes of an equilateral triangle, with a side $2.06a$ long. $\alpha=0°$ corresponds to an incidence on an apex of the triangle, and $\alpha=60°$ to an incidence perpendicular to one side. Once again, stars indicate the frequency position of the single-shell resonance under study. This resonance is seen to split into four well-separated ones. Contrary to the aligned configuration, none of these well-separated resonances ever splits into two overlapping ones.

The relation found in the previous section, between the number of shells and the maximum number of resonances in which each single-shell resonance may be split into, is then no longer valid in the case of nonaligned shells.

The evolution of the resonances is studied now for a cluster composed of four shells in a square arrangement, with a square side equal to $2.06a$. Figure 8 shows the evolution of the resonance spectrum with the incidence angle. $\alpha=0°$ corresponds to an incidence perpendicular to one side of the square, and $\alpha=45°$ to incidence on an apex. One can notice three well-separated resonances. As in the case of three aligned shells, each one of those three resonances may split, at a few incidence angle values, into two close ones.

Does the introduction of a fifth scatterer at the center of the square enhance the coupling between the shells, or stop it? Figure 9 shows that none of that occurs. This figure corresponds to a small rigid cylinder in the center. All resonances of the square are shifted to lower frequencies, in such a way that the six of them are now well separated. Remember, the rigid shell is nonresonant. Its introduction has then changed the resonance spectrum of the square only slightly, as if an additional mass effect took place.

Contrary to the case of aligned shells, it is then not possible, at this point, to relate simply the number of shells to the number of resonances that each resonance of the single shell splits into. An attempt to find such a relation may be done using Ref. 10's assumption that the resonances lie in distinct families associated with the irreducible representations of the symmetry group of the cluster. The authors of Ref. 10 have shown, indeed, that there are four irreducible representations for two (aligned) shells, four for three shells in an equilateral triangle configuration, and six for four shells in square. In all three cases, the number of irreducible representations is actually equal to the number of resonances. It is formally possible to decompose the form function of a given cluster into a number of subsums equal to the number of irreducible representations. The question is now: can each subsum be associated with one (and only one) resonance?

In order to answer this question, Fig. 10(a) shows the resonance spectrum of the square case at $\alpha=21°$. At this incidence angle value, indeed, each one of the three well-separated resonances splits into two close ones, as could be

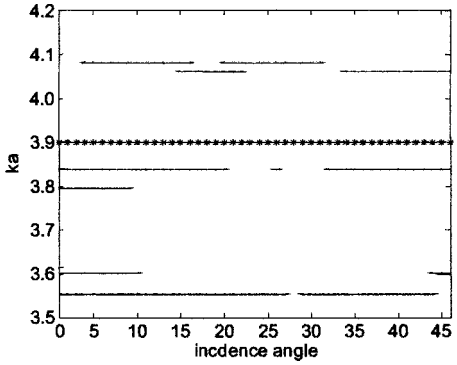


FIG. 9. Introduction of a small rigid cylinder at the center of the square configuration of Fig. 8. Evolution of the resonance spectrum with the incidence angle and the reduced frequency. The stars describe the evolution of the one single-shell resonance under study.

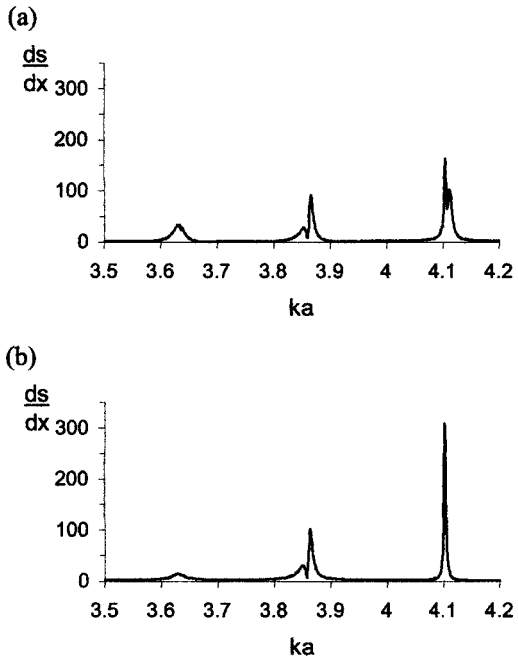


FIG. 10. (a) Four shells in a square configuration, $\beta=2.06$, $\alpha=21°$. Resonance spectrum. (b) Four shells in a square configuration, $\beta=2.06$, $\alpha=21°$. The form function is decomposed into subsums, according to the symmetry group of the cluster. Resonance spectrum associated to the first subsum.

seen already in Fig. 8. Figure 10(b) shows the resonance spectrum obtained from the first subsum that appears in the decomposition of the form function [Eq. (142) of Ref. 10 has been used]. This spectrum is a plot of the frequency derivative ds/dx, as in Eq. (31), with F_∞ replaced by the first subsum of its decomposition. Figure 10(b) shows that such a decomposition of the form function does not allow isolation of the resonances, as four resonances appear in this figure. Moreover, the number of subsums may be different from the number of resonances. For example, the N aligned shells system is invariant under the same symmetry group as the two-shell system, whatever the value of N. The form function of the N aligned shells should then be decomposed into the same number of subsums as the two-shell system, while the number of resonances for aligned shells depends on N, as shown in Sec. IV A. The case of nonaligned shells is then much more complex than the case of aligned ones. We could find no simple way to predict the number of resonances each one of the single shell will split into when placing N close shells in a given spatial configuration.

V. CONCLUSIONS

This study was on the scattering of an incident plane wave in a fluid by N cylindrical shells, in the case of normal incidence (the axes of the shells are perpendicular to the incidence direction). The S matrix of the N-shell cluster has been defined, and its unitarity property verified in order to check the validity of the numerical results. The resonance spectrum of one single shell has been compared with that of different types of clusters. The study was done in the low-frequency domain, so that all resonances of the single shell were due to the external A wave, which can couple close enough adjacent shells. Coupling between shells induces the split of each A-wave resonance of the single shell into M other ones.

Considering different spatial configurations of clusters, a simple relation between N, the number of symmetries of the cluster, and the value of M, has been sought.

When all N shells are aligned, each resonance of the single shell splits into $M=2N$ resonances, corresponding to N well-separated peaks on the resonance spectrum, with each peak composed of two very close (possibly overlapping) resonances.

Nonaligned configurations of the shells are much more complex, and we could not find any simple law to predict the value of M, except for both the equilateral triangle configuration ($M=4$) and the square one ($M=6$). In those configurations, indeed, M is equal to the number of irreducible representations of the symmetry group of the cluster, as it had been suggested by Decanini *et al.*[10]

The time needed for the computation of the resonance spectrum of a given cluster increases with N^2 at a given β value, and with $1/(\beta-2)^2$ at fixed N. For very close shells, this time may be prohibitive, even for not so large values of N. Typically, on a Pentium III, 850-MHz computer, the FORTRAN program that plots the resonance spectrum at a fixed incidence angle and for ka running from 1 to 7 (step 10^{-3}) takes about 220 min for 6 aligned shells with $\beta=2.02$, but the computation time may be reduced using a recursive T-matrix algorithm such as that presented by Wang.[25] The study of ultrasound propagation in dense media, therefore, needs the use of effective theories, with one scatterer equal to an N-scatterer cluster. Anisotropy may be introduced that way by considering identical clusters. Understanding of the frequency evolution of the effective medium properties should be easier from the knowledge of each cluster resonant behavior.

[1] M. Abramovitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[2] L. L. Foldy, "The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers," Phys. Rev. **67**(3,4), 107–119 (1945).
[3] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, San Francisco, London, 1978).
[4] P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**(4), 512–537 (1961).
[5] V. K. Varadan, V. V. Varadan, and Y. H. Pao, "Multiple scattering of elastic waves by cylinders of arbitrary cross section. I. SH waves," J. Acoust. Soc. Am. **63**(5), 1310–1319 (1978).
[6] G. C. Gaunaurd and H. Überall, "Resonance effects and the ultrasonic effective properties of particulate composites," J. Acoust. Soc. Am. **74**(1), 305–313 (1983).
[7] X. Jing, P. Sheng, and M. Zhou, "Acoustic and electromagnetic quasimodes in dispersed random media," Phys. Rev. A **46**(10), 6513–6534 (1992).
[8] A. Derode, A. Tourin, and M. Fink, "Random multiple scattering of ultrasound. I. Coherent and ballistic waves," Phys. Rev. E **64**(036605), 1–7 (2001).
[9] H. Huang and G. L. Gaunaurd, "Acoustic scattering of a plane wave by two spherical elastic shells," J. Acoust. Soc. Am. **98**(4), 2149–2156 (1995).
[10] Y. Decanini, A. Folacci, E. Fournier, and P. Gabrielli, "Exact S matrix for N-disc systems and various boundary conditions. I. Generalization of the Korringa–Kohn–Rostoker–Berry method," J. Phys. A **31**, 7865–7889 (1998).
[11] Y. Decanini, A. Folacci, E. Fournier, and P. Gabrielli, "Exact S matrix for N-disc systems and various boundary conditions. II. Determination and

partial classification of resonances,'' J. Phys. A **31**, 7891–7900 (1998).
[12] E. K. Kheddioui, J. M. Conoir, P. Pareige, and J. L. Izbicki, ''Resonant scattering by two elastic cylindrical shells,'' Acta Acust. **84**, 980–986 (1998).
[13] E. K. Kheddioui, P. Pareige, and J. L. Izbicki, ''Experimental resonant scattering by two elastic cylindrical shells in an eclipsed configuration,'' Acoust. Lett. **16**, 157–162 (1993).
[14] S. Lethuillier, P. Pareige, J. M. Conoir, and J. L. Izbicki, ''Scattering by two very close immersed shells: Numerical results,'' Proceedings of the 1999 IEEE International Ultrasonics Symposium, 731–734 (1999).
[15] S. Lethuillier, J. M. Conoir, P. Pareige, and J. L. Izbicki, ''Resonant acoustic scattering by a finite linear grating of elastic shells,'' Ultrasonics **41**, 655–662 (2003).
[16] J. L. Izbicki, J. L. Rousselot, A. Gérard, G. Maze, and J. Ripoche, ''Analysis of resonances related to Scholte–Stoneley waves around circular cylindrical shells,'' J. Acoust. Soc. Am. **90**, 2602–2608 (1991).
[17] M. Talmant, J. L. Izbicki, G. Maze, G. Quentin, and J. Ripoche, ''External wave resonances on thin cylindrical shells,'' J. Acoust. **4**, 509–523 (1991).
[18] N. D. Veksler, J. L. Izbicki, and J. M. Conoir, ''Bending *A* wave in the scattering by a circular cylindrical shell: Relation with bending eigenmodes,'' J. Acoust. Soc. Am. **96**, 287–293 (1994).
[19] E. B. Danila, J. M. Conoir, P. Pareige, and J. L. Izbicki, ''Multichannel resonant scattering theory applied to the acoustic scattering by an eccentric elastic cylindrical shell immersed in a fluid,'' Wave Motion **28**, 297–318 (1998).
[20] S. Derible, P. Rembert, and J. L. Izbicki, ''Experimental determination of acoustic resonance width via the Argand diagram,'' Acustica **84**, 270–279 (1998).
[21] M. Lax, ''Multiple scattering of waves,'' Rev. Mod. Phys. **23**(4), 287–310 (1951).
[22] *Acoustic Resonance Scattering*, edited by Herbert Uberall (Gordon and Breach Science, New York, 1992).
[23] L. D. Landau and E. M. Lifchitz, *Quantum Mechanics: Non-relativistic Theory* (Pergamon, Oxford, 1966).
[24] L. Flax, G. C. Gaunaurd, and H. Uberall, ''Theory of resonance scattering,'' in *Physical Acoustics XV* (Academic, New York, 1981), pp. 191–293.
[25] Y. M. Wang and W. C. Chew, ''A recursive *T*-matrix approach for the solution of electromagnetic scattering by many spheres,'' IEEE Trans. Antennas Propag. **41**(12), 1633–1639 (1993).

A note on the concept of acoustic center[a)]

Finn Jacobsen,[b)] Salvador Barrera Figueroa,[c)] and Knud Rasmussen[d)]
Acoustic Technology, Ørsted•DTU, Technical University of Denmark, Building 352, DK-2800, Kgs. Lyngby, Denmark

(Received 1 August 2003; revised 2 January 2004; accepted 13 January 2004)

The acoustic center of a reciprocal transducer is defined as the point from which spherical waves seem to be diverging when the transducer is acting as a source. This paper examines various ways of determining the acoustic center of a source, including methods based on deviations from the inverse distance law and methods based on the phase response. The considerations are illustrated by experimental results for condenser microphones. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1652036]

PACS numbers: 43.20.Rz, 43.38.Kb, 43.58.Vb [AJZ] Pages: 1468–1473

I. INTRODUCTION

Any finite source of sound generates a sound pressure that, in any direction, sufficiently far away under free-field conditions varies in inverse proportion to the distance from the source.[1] Very far away the exact position from which this inverse distance law applies is obviously not very important, but closer to the source the position of this ''acoustic center'' may be of some concern.

The concept of acoustic center is frequently referred to in the literature. It is defined in Refs. 2 and 3 as the position of the point from which spherical wavefronts appear to diverge, and in Refs. 4 and 5 as the position from which the sound pressure varies inversely as the distance. Knowledge of the acoustic center is of concern whenever a well-defined distance to a source is needed, for example, in testing anechoic rooms by measuring deviations from the inverse distance law.[6,7] It can also be important to know the acoustic center of a transducer that is used primarily as a receiver. Most measurement microphones are used under free-field conditions, and free-field reciprocity calibration is the most accurate method of determining the free-field sensitivity of microphones. This method involves measuring the transfer function between pairs of microphones.[3,8,9] The ''acoustic distance'' between the transmitter and receiver microphone must be known, and since the two transducers cannot be far from each other because of the extremely poor signal-to-noise ratio of the measurement, their acoustic centers must be known with great accuracy. The acoustic center of a reciprocal transducer does not depend on whether it is used as receiver or transmitter.[10–12]

In general the acoustic center of a source varies with the frequency, with the direction of the observer, and with the distance from the source,[2] as demonstrated theoretically in Refs. 13 and 14. However, Rasmussen showed that the dependence of the distance can be ignored in reciprocity calibration of microphones.[14]

Most results presented in the literature have been determined from deviations between the amplitude (or rms value) of the sound pressure and the inverse distance law. Cox measured the acoustic centers of various transducers using a dismantled lathe bed for positioning the scanning microphone and a 10-turn potentiometer for compensating for the $1/r$ dependence.[13] Rasmussen measured the acoustic centers of microphones of type LS1 (laboratory standard ''1 in.'') with and without protection grid.[14] The results, which were found to agree reasonably well with approximate theoretical considerations, are the basis of the values given in the IEC standard from 1974[15] and the slightly modified values given in the newer one.[3] Rasmussen's data were deduced from the decay of the sound pressure generated by a microphone with the distance, measured with a probe microphone and recorded with a level recorder. In a paper primarily concerned with measuring the air attenuation Hruska and Koidan described a procedure based on fitting a second-order polynomial in the distance between two transducers to an expression based on the distance and the magnitude of the transfer function between them.[16] More recently Rasmussen and Sandermann Olsen summarized the results of a comparison of acoustic center values among several European laboratories.[17] Juhl has determined the acoustic centers of microphones of types LS1 and LS2 (laboratory standard ''0.5 in.'') by calculating the sound pressure amplitude at positions on the microphone axes using the boundary value method and assuming a parabolic movement of the diaphragms.[18] And finally Wagner and Nedzelnitsky have determined the acoustic center of microphones of type LS2 using measured values of the magnitude of the transfer function between two microphones.[19]

A few examples of methods of determining the acoustic center from phase measurements have been found in the literature. Ando determined the acoustic center of a ''pipe horn'' loudspeaker from measurements of the phase shift between two positions.[20] Rasmussen also attempted to determine spherical wavefronts and thus acoustic centers from phase measurements.[14] However, this method was found to fail completely because of imperfections of the anechoic

[a)]Portions of this work were presented in ''A note on the acoustic centre of condenser microphones,'' Proceedings of Tenth International Congress on Sound and Vibration, Stockholm, Sweden, July 2003.

[b)]Author to whom correspondence should be addressed; Electronic mail: fja@oersted.dtu.dk
[c)]Electronic mail: sbf@oersted.dtu.dk
[d)]Electronic mail: kr@oersted.dtu.dk

 0001-4966/2004/115(4)/1468/6/$20.00

room—although the anechoic room used in these measurements is a very good one.[4] Trott proposed a definition of the acoustic center based on phase considerations and presented some theoretical and numerical results,[21] and Vorländer and Bietz described a method of determining the acoustic center from the group delay of the complex sensitivity of a microphone and presented some experimental results.[22]

The purpose of this paper is to examine and discuss the concept of acoustic center and to present some experimental results for condenser microphones.

II. THE CONCEPT OF ACOUSTIC CENTER

Both definitions of the acoustic center, "the position from which outgoing wavefronts appear to diverge in the far field"[2,3] and "the position from which the sound pressure varies inversely as distance,"[4,5] seem to imply the existence of an equivalent point source. The idea of replacing a real, extended source by an equivalent point source seems to be straightforward, and this apparent simplicity perhaps explains why most authors have taken the concept for granted. In fact the only fundamental considerations found in the literature are the Cox and Rasmussen theoretical analyses[13,14] and Trott's "redefinition."[21]

For simplicity, the following discussion is restricted to axisymmetric sources observed from positions on the axis. In this case the acoustic center must be somewhere on the axis. Let r indicate the physical distance from the observation point to a point on the real source, and let x indicate the position of the acoustic center. Several approaches are now possible.

(i) If an analytical expression of the sound field generated by the source is available then one can adjust the parameter x in the expression of the amplitude of the sound pressure generated by the equivalent source,

$$|p_{\mathrm{eq}}(r)| = \frac{\rho c k Q_{\mathrm{eq}}}{4\pi(r-x)}, \tag{1}$$

to the amplitude of the pressure generated by the source under investigation $p(r)$ until the two expressions agree. (Here k is the wave number, ρc is the characteristic impedance of air, and Q_{eq} is the volume velocity of the equivalent source, determined from far field considerations in the direction of concern.) The result is in general a function of k and r,

$$x(k,r) = r - \frac{\rho c k Q_{\mathrm{eq}}}{4\pi|p(r)|}. \tag{2}$$

Cox and Rasmussen used this method in the theoretical part of their studies.[13,14]

(ii) One can plot the reciprocal sound pressure amplitude as a function of r, fit a straight line over the region of concern, and determine its intersection with the r axis, as specified in Ref. 23. This corresponds to the following expression for the position of the acoustic center,

$$x(k,r) = r - \frac{1}{|p(r)|} \Bigg/ \frac{\partial(1/|p(r)|)}{\partial r} = r + |p(r)| \Bigg/ \frac{\partial|p(r)|}{\partial r}, \tag{3}$$

where the rate of change in practice (where no analytical expression is available) must be estimated over a suitable interval. This method was used by Cox and Rasmussen in their experimental work,[13,14] and by Juhl in his numerical study.[18] Similar methods based on measured transfer functions between pairs of microphones have been used in the work reported in Refs. 17, 19, and 24.

(iii) One can estimate the position of the acoustic center from the phase response corrected for the phase shift associated with the distance to the observation point.[21] This corresponds to equating the phase of the sound pressure generated by the equivalent point source with the phase of the sound pressure generated by the source under investigation. Evidently, a phase reference, for example, the velocity at some position on the source U, is needed. The result is the expression

$$x(k,r) = \frac{1}{k}\mathrm{Arctan}\left(\frac{\mathrm{Im}\{(p(r)e^{jkr}/jUe^{j\omega t})\}}{\mathrm{Re}\{(p(r)e^{jkr}/jUe^{j\omega t})\}}\right). \tag{4}$$

Note that the $e^{j\omega t}$ sign convention is used in this paper.

(iv) Alternatively, one can use the corresponding group delay multiplied by the speed of sound,[22]

$$x(k,r) = \frac{\partial}{\partial k}\left(\mathrm{Arctan}\left(\frac{\mathrm{Im}\{(p(r)e^{jkr}/jUe^{j\omega t})\}}{\mathrm{Re}\{(p(r)e^{jkr}/jUe^{j\omega t})\}}\right)\right). \tag{5}$$

(v) One can estimate the position of the center from the curvature of wavefronts determined from phase measurements.[14,20]

A. Some simple examples

A few examples will be presented to demonstrate how the methods work. The first source to be studied is a pulsating sphere. The sound pressure a distance r from the center of a pulsating sphere with radius a and vibrational velocity U is[1]

$$p(r) = \frac{j\rho c k a^2 U}{r(1+jka)} e^{j(\omega t - k(r-a))} = \frac{j\rho c k a^2 U}{r\sqrt{1+(ka)^2}} e^{j(\omega t - k(r-a) - \mathrm{Arctan}(ka))}. \tag{6}$$

It is apparent that the volume velocity of the equivalent monopole that gives the same sound pressure in the far field is

$$Q_{\mathrm{eq}} = \frac{4\pi a^2 U}{\sqrt{1+(ka)^2}}. \tag{7}$$

The two definitions based on amplitude considerations [Eqs. (2) and (3)] give $x(k,r)=0$. In other words, amplitude considerations place the acoustic center of a pulsating sphere in the center of the sphere. By contrast, the two definitions based on phase considerations give

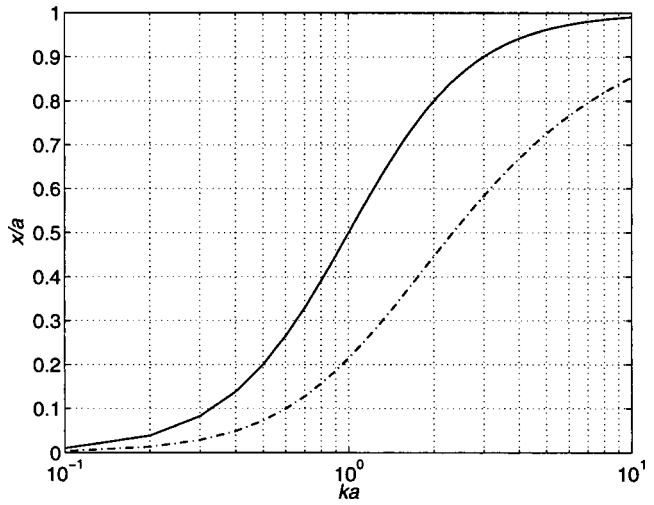


FIG. 1. The acoustic center of a pulsating sphere of radius a calculated using the phase delay [Eq. (8)] (–•–) and the group delay [Eq. (9)] (—).

$$x(k,r)=a-\frac{1}{k}\operatorname{Arctan} ka \tag{8}$$

and

$$x(k,r)=\frac{\partial}{\partial k}(ka-\operatorname{Arctan} ka)=a-\frac{a}{1+(ka)^2}, \tag{9}$$

respectively. [Equation (8) agrees with Trott's results.[21]] As Fig. 1 shows, the two methods based on phase considerations place the acoustic center of a pulsating sphere in the center of the sphere at low frequencies, but as ka increases it moves towards the surface. This makes sense since this is the point where the phase reference is taken. Note that none of the acoustic centers of the pulsating sphere depends on the distance r.

The second example is a baffled circular piston. The sound pressure generated by a vibrating piston with radius a and vibrational velocity U at a position on the axis a distance r from the piston is[1]

$$\begin{aligned} p(r)&=\rho c U e^{j\omega t}(e^{-jkr}-e^{-jk\sqrt{r^2+a^2}}) \\ &=2j\rho c U \sin(k\Delta)e^{j(\omega t-k(r+\Delta))}, \end{aligned} \tag{10}$$

where

$$\Delta=\tfrac{1}{2}(\sqrt{r^2+a^2}-r). \tag{11}$$

Inserting into Eq. (2) using a volume velocity of

$$Q_{\mathrm{eq}}=2U\pi a^2 \tag{12}$$

gives

$$x(k,r)=r-\frac{ka^2}{4\sin(k\Delta)}. \tag{13}$$

The second amplitude-based method, Eq. (3), gives a different expression,

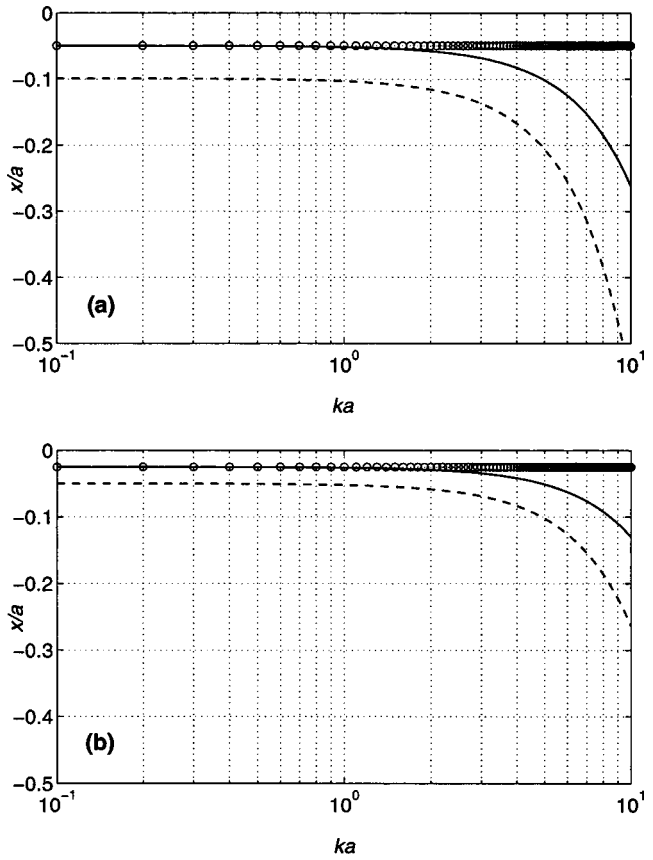


FIG. 2. The acoustic center of a circular piston in a baffle of radius a seen from a distance of (a) five radii and (b) 10 radii. —, First amplitude-based method [Eq. (13)]; ––, second amplitude-based method [Eq. (14)]; ○, both phase-based methods [Eq. (15)].

$$x(k,r)=r-\frac{\tan(k\Delta)}{k\Delta}\sqrt{r^2+a^2}. \tag{14}$$

The phase approach based on Eq. (4) gives yet another expression,

$$x(k,r)=-\Delta=-\frac{1}{2}(\sqrt{r^2+a^2}-r)\simeq-\frac{a^2}{4r}, \tag{15}$$

and since the phase of Eq. (10) is a linear function of the frequency the group delay method, Eq. (5), gives the same value in this case.

As can be seen from Fig. 2, the acoustic center given by Eq. (13) is negative (although it can be positive when the normalized frequency ka is very high), its numerical value increases with the frequency (although very weakly if $ka<1$), and it is essentially inversely proportional to r. (The last mentioned property disagrees with results presented in Rasmussen's report.[14]) Equation (14) gives values with the same tendency but approximately twice as large. Equation (15) agrees fairly well with Eq. (13) only when $ka<1$. According to all the methods the acoustic center of a baffled circular piston is placed behind the piston. Unfortunately the position of the center depends not only on the frequency but also on the distance, unlike the acoustic center of a pulsating sphere.

The third example is that of a point source on a rigid sphere. The sound pressure generated by a monopole with a volume velocity of Q placed on a rigid sphere with radius a is[25]

$$p(r)=-\frac{j\rho cQe^{j\omega t}}{4\pi a^2}\sum_{m=0}^{\infty}\frac{(2m+1)^2h_m^{(2)}(kr)}{mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka)}, \tag{16}$$

at a position in front of the monopole a distance r from the center of the sphere. In this expression $h_m^{(2)}(x)$ is the spherical Hankel function of the second kind and order m. The equivalent monopole can be shown to have a volume velocity of

$$Q_{\mathrm{eq}}=-\frac{Q}{(ka)^2}\sum_{m=0}^{\infty}\frac{(2m+1)^2j^{m+1}}{mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka)}. \tag{17}$$

(This expression approaches Q asymptotically at low frequencies.) Inserting into Eq. (2) gives

$$x(k,r)=r-\frac{1}{k}\frac{\left|\Sigma_{m=0}^{\infty}\frac{(2m+1)^2j^{m+1}}{mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka)}\right|}{\left|\Sigma_{m=0}^{\infty}\frac{(2m+1)^2h_m^{(2)}(kr)}{mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka)}\right|}. \tag{18}$$

Equation (18) agrees with equations derived by Cox and Rasmussen.[13,14] The amplitude method based on Eq. (3) gives a somewhat more complicated expression,

$$x(k,r)=r+2\frac{\Sigma_{m=0}^{\infty}\Sigma_{n=0}^{\infty}\frac{(2m+1)^2h_m^{(2)}(kr)(2n+1)^2h_n^{(1)}(kr)}{(mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka))(nh_{n-1}^{(1)}(ka)-(n+1)h_{n+1}^{(1)}(ka))}}{\Sigma_{m=0}^{\infty}\Sigma_{n=0}^{\infty}\frac{(2m+1)^2(2n+1)^2(h_m^{(2)}(kr)\partial h_n^{(1)}(kr)/\partial r+h_n^{(1)}(kr)\partial h_m^{(2)}(kr)/\partial r)}{(mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka))(nh_{n-1}^{(1)}(ka)-(n+1)h_{n+1}^{(1)}(ka))}}, \tag{19}$$

in which $h_m^{(1)}(x)$ is the spherical Hankel function of the first kind and order m, and

$$\frac{\partial h_m(kr)}{\partial r}=\frac{k}{2m+1}(mh_{m-1}(kr)-(m+1)h_{m+1}(kr)) \tag{20}$$

for Hankel functions of either kind.[26] Finally the method based on the phase delay gives

$$x(k,r)=\frac{1}{k}\mathrm{Arctan}\left(\frac{\mathrm{Im}\left\{\Sigma_{m=0}^{\infty}\frac{(2m+1)^2h_m^{(2)}(kr)e^{jkr}}{mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka)}\right\}}{\mathrm{Re}\left\{\Sigma_{m=0}^{\infty}\frac{(2m+1)^2h_m^{(2)}(kr)e^{jkr}}{mh_{m-1}^{(2)}(ka)-(m+1)h_{m+1}^{(2)}(ka)}\right\}}\right). \tag{21}$$

Figure 3 shows the acoustic center of the point source on the sphere calculated using Eqs. (18), (19), and (21). It is apparent that the acoustic center based on amplitude considerations is placed in front of the physical source at low frequencies and moves towards the source as the frequency is increased. In fact, Eq. (18) approaches 1.5 times the radius of the sphere as ka goes to zero and kr goes to infinity.[13,14] By contrast, the phase-based acoustic center is placed behind the entire sphere at low frequencies.

B. Discussion

In practice it may be difficult to determine the volume velocity of the equivalent source, so the first method, which at first glance seems quite reasonable, may be difficult to use in experimental or numerical work. More importantly, it is not completely obvious how the resulting acoustic center should be interpreted, although it is clear that it reflects deviations from the inverse distance law of a point source adjusted to give the same far field. The second method, which does not require any knowledge of the source and gives the point from which the inverse distance law appears to apply seen from a certain distance, would seem to be more useful, and this is clearly the relevant method for free-field reciprocity calibration of microphones and for testing anechoic rooms. The method based on the phase delay may well be useful for other applications. It may, for example, be useful to know the position of the phase-related acoustic centers of loudspeaker units. On the other hand it is not clear from Ref. 22 why the group delay, which is a quantity associated with the speed with which the energy of a wave packet travels in a dispersive medium,[27,28] should be relevant. The method based on the curvature of wavefronts is unlikely to be reliable in practice because of imperfections of the anechoic room.

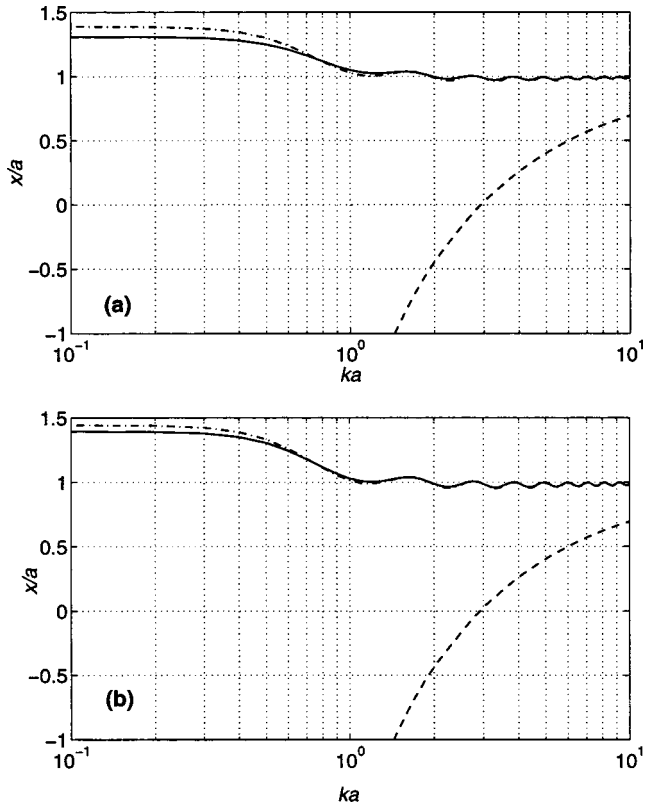


FIG. 3. The acoustic center of a point source on a rigid sphere of radius a seen from a distance of (a) five radii and (b) 10 radii. −•−, First amplitude-based method [Eq. (18)]; —, second amplitude-based method [Eq. (19)]; −−, phase delay method [Eq. (21)].

One might have hoped that the various methods would lead to the same result, but that seems to be the case *only if the source is a point monopole* (in free space or on a rigid baffle). The analysis of a few simple cases has demonstrated that the concept of an acoustic center is more complicated than one might have expected.

It is apparent that the acoustic center of a source of sound can be placed behind the physical vibrating surface or in front of it. That amplitude considerations lead to the acoustic center of a baffled piston being placed behind the piston can be explained by the fact that the sound pressure assumes a finite value in front of the piston unless the piston is infinitely small (a point source). In the latter case the center coincides with the position of the source. However, if the baffle is ''folded back to form a sphere''[14] the center moves forward at low frequencies, presumably because of interference between the direct wave and the one that has traveled around the sphere. Another explanation is that the sound pressure in all the outgoing waves [the terms of Eq. (16)] except that of zero order decays faster than $1/r$ near the sphere. Such considerations lead Rasmussen to conclude that the acoustic center of a condenser microphone could be expected to be placed at a position about half the radius in front of the diaphragm at low frequencies, and closer to the diaphragm or even behind it at high frequencies.[14]

III. EXPERIMENTAL RESULTS

To supplement the theoretical examples some experiments have been carried out. The electrical transfer functions

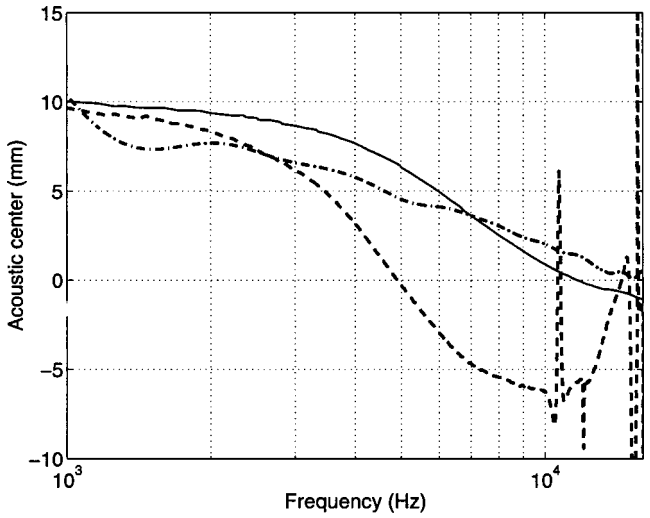


FIG. 4. Acoustic centers of microphones of type LS1 determined from the rate of change of the modulus of the transfer function (−•−), and determined from the free-field phase response using the phase delay (—) and using the group delay (−−). The free-field phase response has been corrected for the phase of the pressure sensitivity.

between three pairs of condenser microphones of type LS1 have been measured at four different distances (from 200 to 500 mm) in a small anechoic room, using a special time-selective technique for removing the influence of cross-talk, reflections from the walls, and standing waves between the microphones.[24,29] The acoustic centers were determined from the rate of change of the amplitude of these functions (corrected for the absorption of air), using the procedure based on Eq. (3). Figure 4 shows the results, which are in good agreement with values given in the IEC standard[3] up to about 18 kHz. For comparison the results of determining the acoustic center from phase measurements are also shown. In this case the physical distance between the two microphone diaphragms has been used in calculating the free-field sensitivities from measured transfer functions. The resulting complex free-field sensitivities have been corrected for the phase of the pressure sensitivities and finally processed as indicated by Eqs. (4) and (5). (The phase of the pressure sensitivity is related to the internal mechanism of the microphone and has nothing to do with the phase resulting from the geometry and the vibrational pattern of the diaphragm.) The spikes in the curve determined from the group delays are due to the differentiation. It is apparent that neither the results deduced from the phase delay nor the corresponding values determined from the group delay are in agreement with the acoustic center determined from amplitude measurements. However, it is clear from the analysis above that no such agreement could be expected.

Using the physical distance between the microphones in determining their free-field sensitivities from measured transfer functions has the effect of making the sensitivities slightly dependent on the distance. When the distance between the amplitude-based acoustic centers is used in the calculations the resulting free-field sensitivities become essentially independent of the distance, and it is obviously these centers that should be used. The amplitude-based acoustic centers do not depend appreciably on the distance

unless the microphones are much closer to each other than the minimum distance used in these measurements.

IV. CONCLUSIONS

The idea of replacing a real, extended source by an equivalent point source from which outgoing wavefronts appear to diverge is *deceptively simple*, and various procedures for determining the position of such a source give in general different results. The most useful approach, in reciprocity calibration of transducers as well as in testing the quality of anechoic rooms, would seem to be the one that gives the position from which the inverse distance law applies, as seen from positions in the region of concern.

[1] A. D. Pierce, *Acoustics. An Introduction to Its Physical Principles and Applications* (The American Heritage of Physics, New York, 1989). See Secs. 4.1, 4.5, and 5.7.

[2] C. L. Morfey, *Dictionary of Acoustics* (Academic, San Diego, 2001).

[3] IEC International Standard 61094-3, ''Measurement microphones, Part 3: Primary methods for free-field calibration of laboratory standard microphones by the reciprocity technique,'' 1995.

[4] IEC International Standard 50(801), ''International electrotechnical vocabulary,'' 1994.

[5] American National Standard ANSI S1.1, ''Acoustical Terminology,'' 1994.

[6] F. Ingerslev, O. J. Pedersen, and P. K. Møller, ''New rooms for acoustic measurements at the Danish Technical University,'' Acustica **19**, 185–199 (1967/68).

[7] K. A. Cunefare, V. B. Biesel, J. Tran, R. Rye, A. Graf, M. Holdhusen, and A.-M. Albanese, ''Anechoic chamber qualification: Traverse method, inverse square law analysis method, and nature of test signal,'' J. Acoust. Soc. Am. **113**, 881–892 (2003).

[8] W. R. MacLean, ''Absolute measurement of sound without a primary standard,'' J. Acoust. Soc. Am. **12**, 140–146 (1940).

[9] W. Wathen-Dunn, ''On the reciprocity free-field calibration of microphones,'' J. Acoust. Soc. Am. **21**, 542–546 (1949).

[10] L. L. Foldy and H. Primakoff, ''A general theory of passive linear electroacoustic transducers and the electroacoustic reciprocity theorem. I,'' J. Acoust. Soc. Am. **17**, 109–120 (1945).

[11] H. Primakoff and L. L. Foldy, ''A general theory of passive linear electroacoustic transducers and the electroacoustic reciprocity theorem. II,'' J. Acoust. Soc. Am. **19**, 50–58 (1947).

[12] F. M. Wiener, ''On the relation between the sound fields radiated and diffracted by plane obstacles,'' J. Acoust. Soc. Am. **23**, 697–700 (1951).

[13] J. R. Cox, Jr., ''Physical limitations on free-field microphone calibration,'' Massachusetts Institute of Technology, Ph.D. thesis, 1954.

[14] K. Rasmussen, Acoustic centre of condenser microphones, The Acoustics Laboratory, Technical University of Denmark, Report No. 5, 1973.

[15] IEC Publication 486, ''Precision method for free-field calibration of one-inch condenser microphones by the reciprocity technique,'' 1974.

[16] G. R. Hruska and W. Koidan, ''Free-field method for sound-attenuation measurement,'' J. Acoust. Soc. Am. **58**, 507–509 (1975).

[17] K. Rasmussen and E. Sandermann Olsen, ''Intercomparison on free-field calibration of microphones,'' The Acoustics Laboratory, Technical University of Denmark, Report PL-07, 1993.

[18] P. M. Juhl, ''A numerical investigation of standard condenser microphones,'' J. Sound Vib. **177**, 433–446 (1994).

[19] R. P. Wagner and V. Nedzelnitsky, ''Determination of acoustic center correction values for type LS2aP microphones at normal incidence,'' J. Acoust. Soc. Am. **104**, 192–203 (1998).

[20] Y. Ando, ''Experimental study of the pressure directivity and the acoustic centre of the 'circular pipe horn loud speaker,' '' Acustica **20**, 366–369 (1968).

[21] W. James Trott, ''Effective acoustic center redefined,'' J. Acoust. Soc. Am. **62**, 468–469 (1977).

[22] M. Vorländer and H. Bietz, ''Novel broad-band reciprocity technique for simultaneous free-field and diffuse-field microphone calibration,'' Acustica **80**, 365–377 (1994).

[23] American National Standard ANSI S1.10-1966, ''Method for calibration of microphones,'' 1966.

[24] S. Barrera Figueroa, New methods for transducer calibration: Free-field reciprocity calibration of condenser microphones, Acoustic Technology, Ørsted•DTU, Technical University of Denmark, Ph.D. thesis, 2003.

[25] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968/1984). See Sec. 7.2.

[26] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, DC, 1965).

[27] A. Papoulis, *Signal Analysis* (McGraw-Hill, New York, 1977). See Sec. 4-2.

[28] J. L. Leander, ''On the relation between the wavefront speed and the group velocity concept,'' J. Acoust. Soc. Am. **100**, 3503–3507 (1996).

[29] S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, ''A time selective technique for free-field reciprocity calibration of condenser microphones,'' J. Acoust. Soc. Am. **114**, 1467–1476 (2003).

Computing fluid-coupled resonance frequencies, mode shapes, and damping loss factors using the singular value decomposition

John B. Fahnline[a)]
The Applied Research Laboratory, Applied Science Building, University Park, Pennsylvania 16802

(Received 13 September 2002; revised 2 January 2004; accepted 13 January 2004)

In many acoustic design problems, it would be useful to be able to compute fluid-coupled resonance frequencies, mode shapes, and their associated damping levels. Unfortunately, conventional eigenvalue solution procedures are either computationally inefficient, unreliable, or have limited applicability. Sophisticated methods for identifying modal parameters using the singular value decomposition have recently emerged in the area of experimental modal analysis, where the available data typically consists of velocity-to-force transfer functions for several drive point locations. In this paper, we show that these techniques can be applied to numerically generated frequency domain data and are even more effective because full matrices of transfer function data are available. This typically allows the modes to be completely separated from each other, such that the modal parameters can be identified using analytical formulas. Several benchmark example problems are solved numerically, including a rectangular cantilever plate, a baffled circular plate, and a baffled circular plate covered by an open-ended rigid-walled pipe. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1652034]

PACS numbers: 43.20.Rz, 43.20.Tb, 43.30.Jx, 43.30.Rj [EGW] Pages: 1474–1482

I. INTRODUCTION

For real eigenvalue extraction problems in structural dynamics, the symmetric input matrix is divided into two parts: one with constant coefficients and the other with a simple squared dependence on angular frequency. Shifted block Lanczos algorithms[1] for determining the resulting resonance frequencies and real mode shapes for these types of problems are very efficient and reliable, especially when applied to sparse matrices. For more general eigenvalue problems, the input matrices can be complex-valued and unsymmetrical with arbitrary frequency dependence. Numerical methods for solving problems of this type are typically much slower and less reliable. Discussions of the algorithms can be found in the MSC/NASTRAN User's Guide[1] and in papers by Giordano and Koopmann,[2] Cunefare and De Rosa,[3] Schram and Pilkey,[4] and Zhang and Raveendra.[5]

In this paper, we will suggest an alternative method for numerically solving general eigenvalue problems using the singular value decomposition. Although the algorithm was specifically intended for determining modal parameters of structural-acoustic systems derived from coupled finite element/boundary element analyses, it should be applicable to any problem with a similar frequency domain formulation. We will only give a brief description of the derivation of the input matrices for structural-acoustic systems because numerous articles on the subject have been previously published. Instead, we will focus on the algorithm for numerically determining the modal parameters using the singular value decomposition. Although similar algorithms have been previously applied to problems in experimental modal analysis, their application to purely numerical computations is somewhat different and unique and represents a significant advance in both efficiency and reliability in comparison to the current state of the art.

II. FLUID-COUPLING FORMULATION

To begin the analysis, we will briefly document our coupled finite element/boundary element formulation for structural-acoustic problems. The analysis follows that of Lax,[6] Wilton,[7] and Schenk and Benthien.[8,9] Similar analyses have also been performed by Lualagnet and Guyader,[10] and Berot and Peseux,[11] among others, although they have generally restricted themselves to problems with analytical solutions for the acoustic pressure fields. We will briefly list some of the steps in the derivation of the coupled equations of motion mainly as a means of establishing notation and emphasizing differences.

The basic idea of the formulation is to use *in vacuo* modes from a finite element analysis as basis functions for fluid-coupling computations. The analysis begins by assuming the displacement can be written in terms of the first N modes of the uncoupled problem as

$$\mathbf{d}=\mathbf{\Phi a}, \quad \mathbf{\Phi}=\{\phi_1,\phi_2,\ldots,\phi_N\}, \tag{1}$$

where the modes are mass normalized and are solutions of the basic eigenvector equation

$$(\mathbf{K}-\omega_\mu^2\mathbf{M})\,\phi_\mu=0. \tag{2}$$

The real-valued matrices $\mathbf{K}$ and $\mathbf{M}$ are the structural stiffness and mass matrices, respectively. Because both $\mathbf{K}$ and $\mathbf{M}$ are typically symmetric and sparse, shifted block Lanczos algorithms can be used to calculate the *in vacuo* mode shapes and resonance frequencies very efficiently. When fluid coupling

a)Electronic mail: jbf@wt.arl.psu.edu

 0001-4966/2004/115(4)/1474/9/$20.00

is added to the problem, the structural displacements satisfy

$$[\mathbf{K}-\omega^2\mathbf{M}+\mathbf{A}(\omega)]\mathbf{d}=\mathbf{f}, \tag{3}$$

where **A** relates the generalized forces due to the acoustic pressure field to the nodal displacements and is generated in our formulation using a boundary element analysis. The matrix **A** is complex-valued, unsymmetrical, and dense, which contrasts with the real-valued, symmetric, sparse, structural stiffness and mass matrices. In our computations, the matrix **A** is forced to be symmetric by averaging it with its transpose. Otherwise, the input matrices violate modal reciprocity, resulting in various inconsistencies, including negative loss factors and/or different loss factors for degenerate mode pairs. Substituting for the displacement vector in terms of the *in vacuo* modes and premultiplying by $\mathbf{\Phi}^T$ yields

$$[\mathbf{\Phi}^T(\mathbf{K}-\omega^2\mathbf{M})\mathbf{\Phi}+\mathbf{\Phi}^T\mathbf{A}(\omega)\mathbf{\Phi}]\mathbf{a}=\mathbf{\Phi}^T\mathbf{f}. \tag{4}$$

Because the *in vacuo* modes are orthogonal and have been assumed to be mass normalized, the $\mathbf{\Phi}^T$ $(\mathbf{K}-\omega^2\mathbf{M})\mathbf{\Phi}$ term reduces to a diagonal matrix $\mathbf{\Omega}$, which, after including proportional modal damping, has terms of the form $\omega_\mu^2-\omega^2-i\eta_\mu\omega_\mu^2$, where ω_μ^2 and η_μ are the eigenvalue and structural damping loss factor for mode μ, respectively. The negative sign for the loss factor is a consequence of assuming our time dependence in the form $e^{-i\omega t}$. If experimental data for the structural loss factors is available, it can be included using the modal loss factors. Otherwise, damping can be included using the imaginary component of the stiffness matrix in modal space from the finite element analysis.[12] We have written a DMAP ALTER for a modal frequency response analysis in NASTRAN such that both the *in vacuo* modes and the modal stiffness matrix are output from a single analysis.

Without modal coupling through the pressure field, the resulting matrix system can be solved trivially because the coefficient matrix is diagonal. After fluid coupling is added, a full, complex system of equations has to be solved at each analysis frequency to determine the modal amplitudes. However, since a comparatively small number of modes are typically used as basis functions, the matrix solution times no longer dominate the overall solution times. Knowing the solution for **a**, the displacement amplitude can be determined from Eq. (1). The commercial finite element program MSC/NASTRAN uses a similar formulation to compute displacement for structures with material-dependent loss factors.[12]

We will not go into the details of computing the acoustic coupling matrices except to list some of the unique characteristics of our formulation. First, the acoustic equations are written in terms of the elemental variables volume velocity and average pressure. This forces the solution to converge as a function of acoustic element mesh density.[13] Second, so-called ''tripole sources'' are used to eliminate nonuniqueness difficulties, following the ideas of Hwang and Chang.[14] Also, thin plate-like elements that deform only in bending are represented using dipole sources in the same manner as in the papers by Martinez[15] and Wu.[16] Fourth, great care has been taken to ensure that the singular fields are integrated correctly over the boundary surface. To do this, we make extensive use of the semianalytical integration scheme originally described by Engblom and Nelson[17] and developed further by a number of researchers.[18–20] Fifth, for frequencies below coincidence, we construct acoustic element meshes by grouping structural elements together into larger acoustic elements. This allows us to adjust the structural and acoustic meshes to properly resolve both types of waves without an overly refined acoustic mesh. The idea of using different structural and acoustic meshes has been developed independently by several researchers,[21–23] over the past few years. In our formulation, the acoustic elements are constructed by simply grouping three-noded triangular and four-noded quadrilateral structural elements together into larger acoustic elements, providing a simple and elegant method for superimposing the two meshes. As input, the user supplies a table listing the acoustic element numbers for each of the structural elements.

Even after reducing the number of acoustic elements, denoted as NA, most of the computation time is spent in generating the acoustic coupling matrix, $\mathbf{\Phi}^T\mathbf{A}(\omega)\mathbf{\Phi}$ in Eq. (4). In calculating **A**, two square matrices of size NA$\times$NA are generated, one of the matrices is inverted and multiplied by the other. The time required to generate the matrices is proportional to the number of structural elements NS squared, while the times required to invert the matrix and perform the subsequent multiplication are proportional to NA.[3] Pre- and post-multiplying by the basis function matrix then yields $\mathbf{\Phi}^T\mathbf{A}(\omega)\mathbf{\Phi}$, the computation time for which is proportional to the number of basis functions NM multiplied by NA.[2] Fortunately, the resulting matrix is a slowly varying function of frequency and can be interpolated from matrices calculated at sparse intervals over the frequency band of interest.[24] The idea of interpolating frequency-dependent matrices as a means of accelerating boundary element calculations was originally proposed by Schenk and Benthien[9] and their ideas were further developed by Zhang and Raveendra.[5] However, they were concerned with interpolating the input matrices, the computational expense for which is proportional to the square of the matrix size. It is much more efficient to interpolate the output matrix in modal space because the costly matrix inversion and multiplications are avoided and the data storage requirements are reduced immensely.

Using an interpolation scheme that provides an estimate of its accuracy allows the process of generating the matrices to be automated, with the user specifying only an overall desired accuracy level. It is then possible to automatically refine the frequency spacing in regions where the matrices vary more rapidly, as they do near acoustic resonances. Through trial and error it was found that fourth-order polynomial interpolation provides a good compromise between accuracy and speed. Higher-order polynomials tend to be slow and to have difficulties mapping resonance peaks. They also require more data to be read into the program at any one time, thus making the programs more memory intensive. It would also be possible to use rational polynomials to perform the interpolation, but for the functions of interest, numerical experiments have shown that more points are required for the same accuracy level.

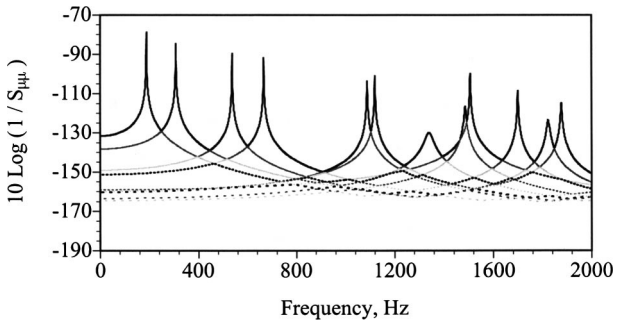


FIG. 1. Reciprocal of the singular values as a function of frequency for an unbaffled rectangular plate in water.

III. COMPUTING MODAL PARAMETERS

In the modal formulation for fluid coupling, a full system of equations with complex coefficients and arbitrary frequency dependence is solved at a series of discrete frequencies to determine a forced response displacement vector. In this paper, we suggest that it is easiest to compute the associated modal parameters using techniques borrowed from experimental modal analyses because the numerical calculations produce similar data. This represents a departure from the standard numerical formulations where computationally expensive complex eigenvalue solution routines or unreliable determinant search routines are commonly used. Of the various methods for computing modal parameters from experimental data, the complex mode indicator function (CMIF), as discussed originally by Shih *et al.*[25] provides a useful starting point for our analysis.

In the algorithm for generating the complex mode indicator function, the singular value decomposition is used with transfer function data for multiple drive points to separate the modal contributions from each other. The potential of this method for numerical calculations can be demonstrated by applying the singular value decomposition to the coefficient matrix, i.e., the matrix in square brackets in Eq. (4), at each analysis frequency and plotting the resulting singular values as a function of frequency for an example problem. To illustrate, Fig. 1 shows the reciprocal of the singular values for an unbaffled rectangular plate in water as a function of frequency.

Each curve in the figure with the same linewidth and texture represents the reciprocal of one of the singular values in the order that they are output from the SVD algorithm. The peaks in the curves correspond to the resonance frequencies for the coupled system. The general idea of the technique is similar to a determinant search, except the singular value decomposition is able to track the modal contribution that causes the individual singular values to go to zero. As we can observe from Fig. 1, the singular values for a particular mode are continuous functions of frequency, but because they are output in order of descending magnitude, the modal contributions are jumbled whenever two singular values cross. Thus, the singular value decomposition clearly provides information about the system's eigenvalues and eigenvectors, but there is still some work required to force the singular values to track the individual modes.

The best method for tracking the modes was originally proposed by Phillips *et al.*[26] The basic idea is to use the singular value decomposition at one frequency to decompose the coefficient matrix at nearby frequencies. This is possible because the change in the singular vectors from frequency to frequency is primarily a phase rotation. We start by decomposing the coefficient matrix at an initial frequency as

$$\mathbf{\Omega}+\mathbf{\Phi}^T\mathbf{A}(\omega_0)\mathbf{\Phi}=\mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^H, \tag{5}$$

where $\mathbf{U}_0$ and $\mathbf{V}_0$ are unitary matrices (i.e., $\mathbf{U}_0\mathbf{U}_0^H=1$, and $\mathbf{V}_0\mathbf{V}_0^H=1$, where the superscript H denotes a complex-conjugate transpose), and the matrix $\mathbf{S}_0$ is a diagonal matrix containing the real, positive singular values. The columns of $\mathbf{U}_0$ and $\mathbf{V}_0$ are called the left-and right-singular vectors, respectively. At the peaks in the reciprocal of the singular values, the associated left-singular vectors represent mode shapes of the coupled system. Because the number of *in vacuo* modes is typically small, the time required to compute the singular value decomposition is also small. Algorithms in FORTRAN 77 are available for computing the singular value decomposition of a complex matrix in the LAPACK library,[27] and are available on the internet at www.netlib.org under the name zgesvd.f. They have been optimized for shared-memory vector and parallel processors and are quite intricate. A more condensed version can be found in the original paper by Businger and Golub.[28]

To decompose the coefficient matrix at a secondary frequency using the singular value decomposition at an initial frequency, we premultiply the coefficient matrix at angular frequency ω by $\mathbf{U}_0^H(\omega_0)$ and postmultiply by $\mathbf{V}_0(\omega_0)$ to yield

$$\overline{\mathbf{S}}(\omega)=\mathbf{U}_0^H[\mathbf{\Omega}+\mathbf{\Phi}^T\mathbf{A}(\omega)\mathbf{\Phi}]\mathbf{V}_0, \tag{6}$$

where the overbar on the quantity $\overline{\mathbf{S}}(\omega)$ indicates that it is complex valued. The diagonal elements of $\overline{\mathbf{S}}(\omega)$ now represent the contribution of the singular vectors at ω_0 to the response at ω, and are called ''enhanced frequency response functions'' by Phillips *et al.*[26] We prefer the more descriptive name ''modal transfer functions'' because they represent the response of the modes as complete entities. At the initial frequency, the modal transfer function yields $\mathbf{S}_0$. In theory, we can judge how well the decomposition separates the modes by comparing the relative magnitudes of the off-diagonal elements of $\overline{\mathbf{S}}(\omega)$ to the diagonal terms. However, there are better ways of checking the decomposition for residual effects of nearby modes, as we will demonstrate.

Using the singular value decomposition at one frequency to decompose nearby frequencies is advantageous because it forces the change in phase from frequency to frequency to be reflected in the diagonal elements of $\overline{\mathbf{S}}(\omega)$ rather than the left-singular vectors. To demonstrate, the upper graph in Fig. 2 shows a Nyquist plot of the real and imaginary components of one of the diagonal elements of $\overline{\mathbf{S}}(\omega)$ near its resonance peak.

The input data, represented by the seven nodal points, maps out a circle, similar to a velocity-to-force transfer function near a resonance peak.[29] Thus, we can use the singular value decomposition to separate the modal contributions from each other, after which the modal parameters can be

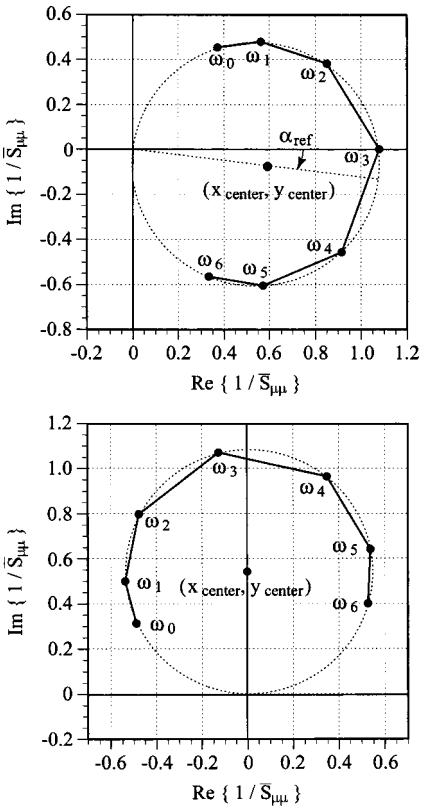


FIG. 2. Nyquist plots of a modal transfer function near a peak before and after rotating the modal circle.

identified from the appropriate diagonal element of $\overline{\mathbf{S}}(\omega)$. We will discuss the lower graph in Fig. 2 subsequently.

One of the main differences between identifying modal parameters from experimental and numerical data is the degree to which the modes can be separated from each other. Because the singular value decomposition can be performed using full matrices in a numerical implementation, there are commonly no residual effects from nearby modes, and the process of identifying the modal parameters is much simpler. To demonstrate that a modal transfer functions represents only a single mode, its Nyquist plot can be examined near the peak. Returning to the top graph in Fig. 2, the dotted line represents a least squares fit to the modal circle,[30] which yields the (x, y) location of the circle's center and it's radius. If the data contained residual effects from nearby modes, the modal circle would not intersect the origin.[29] To test this numerically, we compute the radial distance from the origin to the circle's center and compare it to its radius. If they do not agree to, within, say 1%, then the modal circle does not pass through the origin and the data does not represent a pure modal transfer function. This information can thus be used to filter out false peaks that sometimes occur near frequencies, where two singular values switch order.

Knowing the x-, y location of the modal circle's center and the x-, y locations and frequencies for each of the input data points, the resonance frequency and damping loss factor can be calculated analytically. To begin, the associated resonance frequency's location on the modal circle can be computed as $(2x_{\text{center}},2y_{\text{center}})$, and its polar angle is $\alpha_{\text{ref}} = \tan^{-1}(y_{\text{center}}/x_{\text{center}})$. We can check these results using a formula given by Ewins.[29] For two frequencies above and below the resonance, as denoted by subscripts a and b, respectively, we can write

$$\eta\omega_r^2 = \text{const} = \frac{\omega_a^2 - \omega_b^2}{\tan[(\alpha_a - \alpha_{\text{ref}})/2] - \tan[(\alpha_b - \alpha_{\text{ref}})/2]}, \tag{7}$$

where the angles are evaluated using the modal circle's center as the origin, and we have assumed proportional damping (a slightly different formula would be used if viscous damping were assumed). The minus sign is used in the denominator because we allow α_b to be negative. We can then check the accuracy of the solution for α_{ref} by substituting onto the right-hand side of Eq. (7) for various combinations of points on either side of the peak and confirming that the same constant is indeed produced for each combination.

We can also use Eq. (7) to derive analytical results for the resonance frequency and damping loss factor. Taking ω_a to be the resonance frequency ω_r and ω_b to be a point on the modal circle and applying the formula twice, we can write

$$\frac{\omega_r^2 - \omega_1^2}{-\tan[(\alpha_1 - \alpha_{\text{ref}})/2]} = \frac{\omega_r^2 - \omega_2^2}{-\tan[(\alpha_1 - \alpha_{\text{ref}})/2]}. \tag{8}$$

Solving this equation for ω_r gives

$$\omega_r^2 = \frac{-\omega_1^2 \tan[(\alpha_2 - \alpha_{\text{ref}})/2] + \omega_2^2 \tan[(\alpha_1 - \alpha_{\text{ref}})/2]}{\tan[(\alpha_1 - \alpha_{\text{ref}})/2] - \tan[(\alpha_2 - \alpha_{\text{ref}})/2]}. \tag{9}$$

After determining the resonance frequency, we can then determine the loss factor directly from Eq. (7). Thus, given an accurate circle fit passing through the origin, the modal parameters can be determined by substituting into analytical formulas.

The last remaining obstacle is to identify resonance peaks from the modal transfer functions. In advance, this would seem to be trivial because it is easy to search a string of numbers for a peak. However, because the modal transfer function data is continuously recomputed as the input data shifts from frequency to frequency, the peaks also sometimes shift. For example, let us suppose the modal transfer functions are computed for seven frequencies (numbered 0–6) using the left- and right-singular vectors for the center frequency (numbered 3). It is possible for the modal peak to switch from just above the center frequency to just below as the data sweeps past a resonance, especially for modes with high damping. Thus, a peak in the modal transfer function never occurs at the center frequency, even though a resonance is obviously present. We can eliminate this problem by searching for peaks in the functions not only at the center frequency, but also at the adjacent frequencies.

This leads to another problem though, because in most instances the same mode will be identified two or three times as the input data sweeps past a resonance peak. We thus need some method to detect and eliminate duplicate modes. Fortunately, it is relatively easy to find duplicate modes using the modal assurance criteria. To provide a common phase reference for the left-singular vectors, the modal circles are rotated so that the resonance frequency is aligned with the imaginary axis of the complex plane, and thus the center of

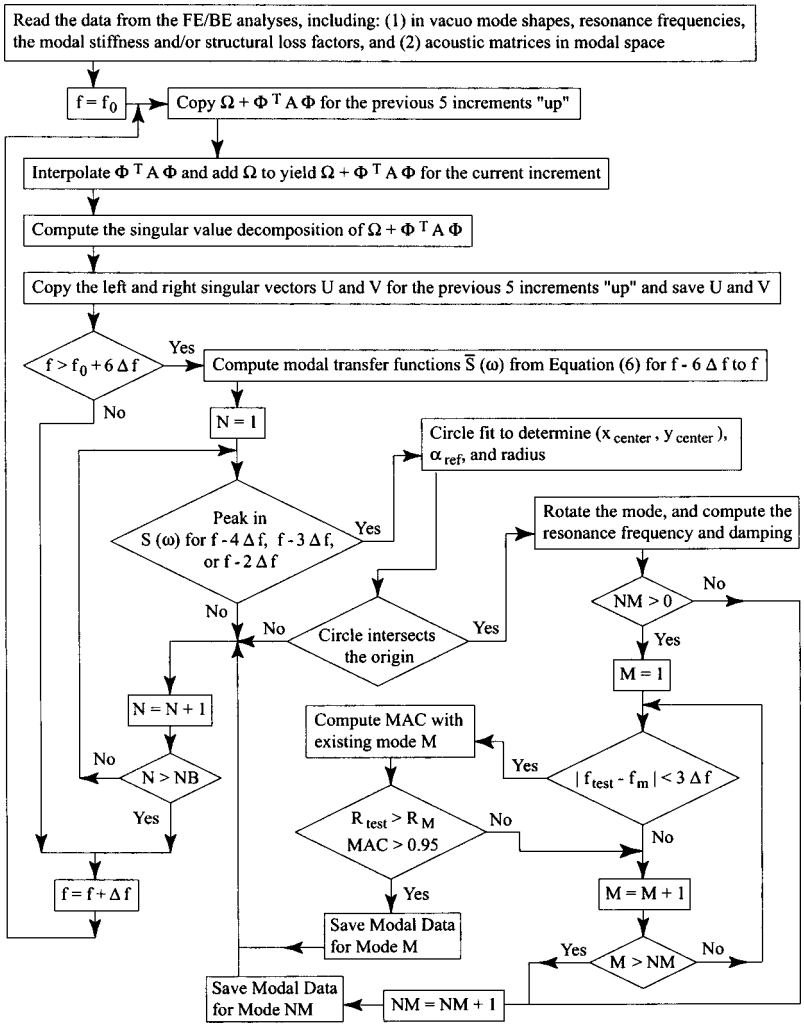


FIG. 3. Flow chart of the algorithm for computing the resonance frequencies, mode shapes, and damping loss factors.

the circle is located at $(0, r_{\text{center}})$, as illustrated in the lower graph in Fig. 2 with the rotated modal circle. The transformation is performed by subtracting the center location, rotating the circle, and then adding the circle's radius as an imaginary number. The resonance frequency is rotated to the positive y axis to be consistent with our definition for the damping loss factor. The modal assurance criteria can then be computed as usual, commonly yielding values near unity for duplicate modes. Finally, after identifying duplicate modes, we retain the one with the larger modal circle as the better representation of the modal transfer function. The end result is a stable, accurate, and robust method for identifying modes. To reduce the computational burden, the duplicate mode check is only performed for center frequencies within a few frequency increments of each other. A flow chart of the full algorithm is given in Fig. 3, where we assumed that the modal transfer functions are computed for seven frequencies, three on either side of the center frequency.

In practice, the algorithm seems to work better using 13 frequencies. All of these calculations increase the computations times, but for realistic structural-acoustic applications, the vast majority of the time is typically spent in the generation of the acoustic coupling matrices, so the overall computation time is only modestly increased.

To test the basic accuracy of the algorithm, several problems were analyzed with added stiffnesses and/or masses and the results were compared to those given by MSC/NASTRAN. In each case, the algorithm did a good job of finding the modes and predicting the resonance frequencies and damping loss factors, to within a few decimal places. It would be possible to check the solution for the eigenvalues and eigenvectors directly by substituting them back into the left-hand side of Eq. (4) and determining if the right-hand side goes to zero. It not, inverse iteration could be used to refine the predictions. However, because high precision is not really required for the modal parameters and because the acoustic matrices are not computed to extremely high precision in the first place, the extra effort needed to refine the predictions is not warranted.

Although this algorithm is usually very effective at separating the modes from each other, there are situations where it is not appropriate. A good example is the idealized problem of a spring-mounted piston driving the fluid within a rigid-walled tube. Only a single *in vacuo* basis function is required to represent the motion of the piston, but numerous acoustic modes exist in the fluid. The result is a single frequency response function with numerous peaks, making it impossible to separate the modes from each other using the singular value decomposition. This is not a major difficulty though because there are numerous methods available for

computing resonance frequencies and loss factors that would be applicable to this type of frequency response data.

As an aside, we note that in our numerical calculations the matrix solution for the basis function amplitudes in Eq. (4) is more accurately determined using the LU factorization than the singular value decomposition. This is especially true near frequencies where two of the smaller singular values cross, leading to ''dropouts'' in the basis function amplitudes. The accuracy of the computed singular values can be shown to be equal to the magnitude of the largest term in the input matrix multiplied by the relative error in the individual terms.[31] For example, if the entries in $\mathbf{\Omega}+\mathbf{\Phi}^T\mathbf{A}(\omega)\mathbf{\Phi}$ are accurate to two digits, singular values smaller than $10^{-2}\|\mathbf{\Omega}+\mathbf{\Phi}^T\mathbf{A}(\omega)\mathbf{\Phi}\|_\infty$, cannot be accurately computed. This restriction would seem to have dire consequences because the smallest singular values yield the information used to compute the modal parameters and the input matrix is only computed to limited accuracy (due to numerical integration and interpolation). However, the solution for the singular values seems to be more sensitive to errors near frequencies where two singular values cross, rather than near the actual zeros. In practice, this simply means that the matrix inverse is best computed in an extra step using the LU factorization rather than the singular value decomposition.

IV. EXAMPLE PROBLEMS

Despite the importance of damping in structural-acoustic systems, there are few papers where acoustic radiation damping loss factors have been computed numerically (some data is given for a sphere by Giordano and Koopmann[2]). Simple approximations derived for applications in statistical energy analysis are given by Cremer and Heckl,[32] but these formulas cannot be applied to fluid-coupling problems because they do not account for fluid motion. This attests to the difficulty of solving the problem using conventional numerical techniques. There is also very little experimental data available, except for a few papers by Clarkson and Brown,[33] where band-averaged acoustic loss factors have been determined for SEA applications. To help in future research and to demonstrate the convergence of our numerical solution, we will provide benchmark data for several example problems, including a rectangular cantilever plate in water, a baffled circular plate with clamped boundary conditions in water, and a baffled circular plate with clamped boundary conditions covered by an open-ended rigid-walled pipe in water. For all of the problems, the structural damping was set to zero so that the predicted loss factors can be attributed entirely to acoustic radiation damping.

TABLE I. Physical constants for the cantilever rectangular plate.

L_x	7.7 in.	ρ	0.284 lb./in.^{3}
L_y	15.4 in.	E	2.9×10^7 lb./in.^{2}
h	0.0693 in.	ν	0.29

A. Cantilever plate

We begin by examining a rectangular plate in water with cantilever boundary conditions. The plate is mounted perpendicular to an infinite baffle, and replicates one of the plates studied by Lindholm *et al.*[34] This type of problem has been studied extensively because the modes resemble those of wing-like structures, where they may possibly contribute to flutter instabilities. The plate's material properties and dimensions are listed in Table I.

This is a good test case to begin our analyses because transverse shear effects are negligible in the plate and the assumption of clamped boundary conditions is closely satisfied in the experimental measurements by Lindholm *et al.* Also, the resonance frequency shifts from air to water are relatively large, providing a good test for the numerical formulation. Our goal in the analysis will be to compute the first ten fluid-coupled resonance frequencies and the associated loss factors.

For the numerical computations, we start by solving for the *in vacuo* modes using the finite element method. Several structural meshes were used to test convergence, and it was determined that a mesh with 1152 square elements with approximately quarter inch sidelength yields the first 10 resonance frequencies to within a few Hz. With the structural mesh determined, the finite element analysis was then rerun with the upper frequency for the eigenvalue analysis set to twice the maximum frequency of interest, such that 21 modes up to 1000 Hz are included as basis functions. The conventional recommendation is to include modes up to 1.6 times the maximum frequency of interest, but we have included modes up to double the maximum frequency to account for the effects of fluid coupling. For the acoustic analysis, several meshes are constructed by grouping structural elements together into larger acoustic elements, yielding acoustic meshes with 128 (3×3 blocks), 288 (2×2 blocks),

TABLE II. Resonance frequencies and loss factors for a cantilever rectangular plate.

	In vacuo	128 Acs. elements		288 Acs. elements		1152 Acs. elements	
(x, y)	F_μ (Hz)	F_μ (Hz)	η_μ	F_μ (Hz)	η_μ	F_μ (Hz)	η_μ
(0, 0)	9.61	3.43	8.7×10^{-8}	3.36	8.6×10^{-8}	3.29	8.5×10^{-8}
(1, 0)	41.48	19.47	2.8×10^{-11}	18.98	2.7×10^{-11}	18.47	2.5×10^{-11}
(0, 1)	59.95	22.92	9.1×10^{-6}	22.35	8.8×10^{-6}	21.79	8.5×10^{-6}
(1, 1)	134.95	65.71	2.0×10^{-9}	63.87	1.8×10^{-9}	61.98	1.7×10^{-9}
(0, 2)	168.42	71.51	8.3×10^{-5}	69.23	7.9×10^{-5}	67.12	7.5×10^{-5}
(1, 2)	259.09	132.71	2.5×10^{-8}	128.30	2.3×10^{-8}	123.95	2.1×10^{-8}
(2, 0)	260.46	141.05	9.1×10^{-7}	135.90	6.6×10^{-7}	130.96	6.0×10^{-7}
(0, 3)	332.03	157.31	4.7×10^{-4}	150.96	4.4×10^{-4}	145.23	4.1×10^{-4}
(2, 1)	354.88	194.54	9.2×10^{-8}	186.95	2.6×10^{-8}	179.76	2.2×10^{-8}
(1, 3)	428.16	231.72	2.9×10^{-7}	222.62	2.6×10^{-7}	213.97	2.3×10^{-7}

and 1152 elements. Dipole sources are used in the acoustic analysis so that the boundary condition only has to be enforced over one side of the plate, thus reducing the number of acoustic elements by half. Based on the simple rule of 6 acoustic elements per wavelength, we expect the 128, 288, and 1152 acoustic element meshes to yield accurate predictions up to 10, 15, and 30 kHz, respectively. Intuitively, these frequencies seem high, and indeed, we will not claim the predicted resonance frequencies and loss factors are accurate below the limiting frequencies. Thus, it is usually better to infer the prediction's accuracy from convergence tests rather than from the simple six acoustic elements per wavelength rule.

Finally, as discussed in Sec. I, the acoustic-coupling matrices are generated over the frequency range of interest using the automated scheme for choosing the interpolation frequencies. In all of the subsequent calculations, the interpolation accuracy was set to 1%. Approximately 50 interpolation frequencies were required to map out the range from 0.1–500 Hz. Most of the interpolation points are concentrated at low frequencies, where the higher-order modes radiate sound inefficiently, thus producing acoustic coupling matrices with high-order frequency dependence.

To compute the resonance frequencies and loss factors, a frequency sweep is performed based on Eq. (4) with a 1 Hz resolution. At each frequency, the acoustic coupling matrix is determined by interpolation. The coefficient matrices and the resulting singular value decomposition are retained for a number of frequency increments to be used in computing the modal transfer functions. Once peaks have been identified in the modal functions, the resonance frequencies, mode shapes, and damping loss factors are identified using the modal circle fit algorithm along with analytical formulas and the results are stored in memory. When duplicate modes are found, they are compared, and if the new mode is determined to be more accurate, the results for the modal parameters are copied in place of the original results. Table II lists the computed resonance frequencies and loss factors for this problem.

The numbering scheme for the modes is similar to that used by Leissa,[35] where the number of nodal lines along the width and length are used to identify the mode shapes. In the table, all the loss factors are reported for reference purposes, even though loss factors of less than 1×10^{-5} would be insignificant in comparison to typical structural loss factors. For dipole source models, the convergence in the predicted resonance frequencies and loss factors is generally slower than for similar problems, where the structure has an overall net volume velocity. The results in the table show that both the predicted resonance frequencies and radiation loss factors converge downward, as the acoustic element mesh is refined. It would be preferable for the loss factors to converge upward so that the resulting radiated power predictions could be considered to be conservative estimates. Presently, a method for controlling the direction of convergence is unknown.

TABLE III. Material properties and dimension for the baffled circular plate with clamped boundary conditions.

a	0.3048 m	E	10.9×10^{10} N/m^2
h	0.005 m	ν	0.29
ρ	7840 kg/m^3		

We can check the first few fluid-coupled resonance frequencies by comparing to the measurements for plate #15 in the paper by Lindholm *et al.*[34] For the in-air measurements, the first five resonance frequencies are listed as 9.3, 42.0, 57.8, 135.0, and 162.0 Hz. In water, the first five resonance frequencies are listed as 3.1, 18.8, 21.1, 62.0, and 65.3 Hz. The agreement between their measured resonance frequencies and our predictions is excellent, thus providing some confirmation of the general accuracy of the formulation.

B. Clamped circular plate

For the second example problem, we will consider a baffled circular plate with clamped boundary conditions in water. The material properties and dimensions for the plate are listed in Table III.

For this problem, analytical solutions for the first two fluid-coupled resonance frequencies have been derived by Lamb,[36] whose results are summarized by Leissa.[37] In the original paper by Lamb, the predictions for the resonance frequencies were compared to experimental measurements for an iron plate, yielding agreement to within 1%.

In the numerical computations, it was determined that a mesh with 24 and 72 elements in the radial and circumferential directions, respectively, yields the first 10 resonance frequencies to within a half percent (as compared to the resonance frequencies for the *in-vacuo* modes listed by Leissa[38]). The basis functions include 51 modes up to 2500 Hz. The acoustic meshes are illustrated in Fig. 4.

The two acoustic meshes in the figure have 192 and 432 elements and are used along with an 1728 element mesh to test the convergence of the predictions for the fluid-coupled resonance frequencies and damping. In generating the acoustic coupling matrices, approximately 50 interpolation frequencies were required to map out the range from 2–1000 Hz. Table IV lists the resulting resonance frequencies and damping ratios for the first ten modes.

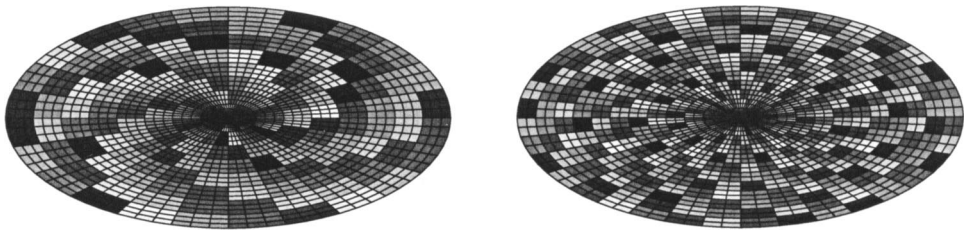


FIG. 4. Structural and acoustic element meshes for a circular plate, where acoustic elements are represented by groups of elements with a similar shade.

TABLE IV. Resonance frequencies and loss factors for a baffled circular plate with clamped boundary conditions.

	In vacuo	192 Acs. elements		432 Acs. elements		1728 Acs. elements	
(r,θ)	F_μ (Hz)	F_μ (Hz)	η_μ	F_μ (Hz)	η_μ	F_μ (Hz)	η_μ
(0, 0)	98.34	39.83	0.017 55	39.78	0.017 48	39.75	0.017 44
(0, 1)	204.89	113.21	0.000 11	112.58	0.000 11	112.21	0.000 11
(0, 2)	335.99	211.40	1.7×10^{-6}	209.02	1.6×10^{-6}	207.56	1.6×10^{-6}
(1, 0)	383.53	232.63	0.023 42	230.59	0.022 92	229.31	0.022 59
(0, 3)	491.80	337.01	6.7×10^{-8}	330.79	5.6×10^{-8}	326.90	5.2×10^{-8}
(1, 1)	587.76	398.57	0.002 13	393.30	0.002 01	389.99	0.001 94
(0, 4)	672.28	491.61	$<1\times10^{-10}$	478.76	$<1\times10^{-10}$	470.54	$<1\times10^{-10}$
(1, 2)	816.58	594.32	0.000 17	583.68	0.000 15	576.86	0.000 14
(2, 0)	862.64	619.83	0.025 40	608.51	0.024 71	601.10	0.024 20
(0, 5)	877.69	676.60	6.3×10^{-8}	653.95	9.9×10^{-8}	638.95	1.1×10^{-7}

In this case, the numbering scheme corresponds to the number of nodal lines in the radial and circumferential directions. Contrary to the previous example, both the loss factors and resonance frequencies now converge downward as the acoustic mesh is refined. The reason for the change in the convergence properties for the loss factors in comparison to the first example problem is still unknown. As a check, analytical solutions for the first two fluid-coupled modes are given for the specified material properties as 39.87 and 112.78 Hz using Lamb's formulas,[36] which compare very well to the computed results.

C. Clamped circular plate covered by a rigid-walled pipe

For the final example problem, we will repeat the analysis from the previous section with an open-ended rigid-walled pipe placed over the plate. Its length L is taken to be 1.5 m, and it is assumed to be circular with the same radius as the plate. This problem differs from the two previous problems, because acoustic resonances will occur in the fluid-filled pipe, thus providing a more general test of the algorithm. The structural mesh now consists of 3888 elements, with the normal surface velocity taken to be identically zero on the walls of the pipe. The acoustic meshes consists of 432, 972, and 3888 elements, where the condensation level for the rigid-walled pipe is the same as for the circular plate. The structural elements used to model the pipe walls are somewhat larger that those used for the plate, but this is permissible because the pipe mesh only has to resolve acoustic waves. In generating the acoustic coupling matrices, approximately 100 interpolation frequencies were required to map out the range from 2–1000 Hz. In comparison to the clamped circular plate alone, the extra interpolation frequencies are primarily concentrated near the uncoupled plane wave acoustic resonances at 201 and 665 Hz (as computed for the 3888 element acoustic mesh). As a simple check, an approximation for the lowest quarter-wavelength acoustic mode is given as $f=c/4(L+0.6133a)=225$ Hz. Table V lists the predictions for the fluid-coupled resonance frequencies and loss factors.

In the table, the modes are labeled by their circumferential Fourier component rather than the corresponding *in vacuo* mode shapes because some of the modes do not have *in vacuo* equivalents. We can check the accuracy of the resonance frequency shifts due to fluid coupling by computing an approximate analytical solution for the lowest resonance frequency by extending Lamb's formula.[36] The added mass correction to the lowest resonance frequency is then given by

$$f\approx f_0\sqrt{h\rho_s/[h\rho_s+\rho(L+0.6133a)]}=15.00\ \text{Hz}, \qquad (10)$$

where ρ_s is the density for the plate. Comparing the results in Tables IV and V, the radiation loss factors decrease for most of the modes because they occur below the cut-on frequency for acoustic waves in the pipe. The exception is the $N=0$ mode at approximately 430 Hz, which excites the fundamental acoustic plane wave resonance and is above its cut-on frequency. Despite the addition of acoustic resonances to the system, residual effects of nearby modes could not be de-

TABLE V. Resonance frequencies and loss factors for a clamped circular plate covered by an open-ended rigid-walled pipe in water.

	432 Acs. elements		972 Acs. elements		3888 Acs. elements	
Fourier component	F_μ (Hz)	η_μ	F_μ (Hz)	η_μ	F_μ (Hz)	η_μ
$N=0$	20.70	0.002 47	20.28	0.002 36	17.92	0.001 59
$N=1$	107.65	7.4×10^{-7}	106.35	1.8×10^{-7}	105.24	3.2×10^{-8}
$N=2$	207.19	5.0×10^{-8}	203.82	1.5×10^{-8}	201.35	5.6×10^{-9}
$N=0$	211.23	0.001 55	208.99	0.001 24	206.16	0.000 22
$N=3$	333.66	1.8×10^{-8}	326.20	1.6×10^{-8}	321.02	1.2×10^{-8}
$N=1$	389.17	0.000 02	382.24	4.1×10^{-6}	376.99	6.9×10^{-7}
$N=0$	448.45	0.046 14	437.32	0.042 67	415.91	0.035 10
$N=4$	488.92	$<1\times10^{-10}$	474.68	$<1\times10^{-10}$	464.85	$<1\times10^{-10}$
$N=2$	587.24	5.0×10^{-6}	574.34	1.1×10^{-6}	564.94	1.4×10^{-8}
$N=0$	607.94	0.010 18	593.47	0.007 59	582.24	0.005 27

tected in any of the modal transfer functions, thus allowing the analytical formulas to be used in calculating the modal parameters.

V. CONCLUSIONS

It has been shown that sophisticated parameter identification routines based on the singular value decomposition can be adapted for purely numerical predictions of fluid-coupled resonance frequencies, mode shapes, and loss factors. Because full matrices of modal transfer function data are available, in most instances the modes can be completely separated from each other, making the calculations more robust and reliable than similar calculations using experimental data. Also, it was shown that the overall convergence of the solution can be ensured by first testing the convergence of the finite element predictions for the *in vacuo* resonance frequencies and then testing the convergence of the predicted fluid-coupled resonance frequencies as a function of acoustic element mesh density. Although a detailed analysis of the computational cost was not given, we can conclude that the algorithm is very efficient since only a limited number of computations are required over and above those associated with nominal frequency response calculations.

[1] D. N. Herting, MSC NASTRAN *Advanced Dynamic Analysis User's Guide*, Macneal-Schwendler, LA, 1993, 2002, Appendix A, pp. 146–160.
[2] J. A. Giordano and G. H. Koopmann, ''State space boundary element–finite element for fluid-structure interaction analysis,'' J. Acoust. Soc. Am. **98**, 363–372 (1995).
[3] K. A. Cunefare and S. De Rosa, ''An improved state-space method for coupled fluid-structure interaction analysis,'' J. Acoust. Soc. Am. **105**, 206–210 (1999).
[4] U. Schramm and W. D. Pilkey, ''Reanalysis of free vibrations considering the frequency-dependency of the structural matrices,'' Int. J. Anal. Exp. Modal Anal. **10**, 53–68 (1995).
[5] Z. Zhang and S. T. Raveendra, ''Eigenfrequency extraction based on the boundary element method,'' *Internoise 2002*, Paper No. N427.
[6] M. Lax, ''The effect of radiation on the vibrations of a circular diaphragm,'' J. Acoust. Soc. Am. **16**, 5–13 (1944).
[7] D. T. Wilton, ''Acoustic radiation and scattering from elastic structures,'' Int. J. Numer. Methods Eng. **13**, 123–138 (1978).
[8] H. A. Schenk and G. W. Benthien, ''Numerical solution of acoustic-structure interaction problems,'' Naval Ocean Systems Center Technical Report 1263, San Diego, CA, 1989.
[9] H. A. Schenk and G. W. Benthien, ''Application of a coupled finite-element boundary-element technique to large-scale structural acoustic problems,'' in *Advances in Boundary Elements: Vol. 2 Field and Fluid Flow Solutions*, edited by C. A. Brebbia and J. J. Connor (Computational Mechanics Publications, Cambridge, MA, 1989), pp. 309–317.
[10] B. Laulagnet and J. L. Guyader, ''Modal analysis of a shell's acoustic radiation in light and heavy fluids,'' J. Acoust. Soc. Am. **131**, 397–415 (1989).
[11] F. Berot and B. Peseux, ''Vibro-acoustic behavior of submerged cylindrical shells: analytical formulation and numerical model,'' J. Fluids Struct. **12**, 959–1003 (1998).
[12] K. Blakely, MSC/NASTRAN *User's Guide: Basic Dynamic Analysis*, Macneal-Schwendler, LA, 1993, pp. 113–114.
[13] J. B. Fahnline, ''A lumped parameter model for the acoustic power output of a vibrating structure,'' J. Acoust. Soc. Am. **100**, 3539–3547 (1996).
[14] J. Y. Hwang and S. C. Chang, ''A retracted boundary integral equation for exterior acoustic problems with unique solution for all wave numbers,'' J. Acoust. Soc. Am. **90**, 1167–1180 (1991).
[15] R. Martinez, ''The thin-shape breakdown (TSB) of the Helmholtz integral equation,'' J. Acoust. Soc. Am. **90**, 2728–2738 (1990).
[16] T. W. Wu, ''A direct boundary element method for acoustic radiation and scattering from mixed regular and thin bodies,'' J. Acoust. Soc. Am. **97**, 84–91 (1995).
[17] J. J. Engblom and R. B. Nelson, ''Consistent formulation of sound radiation from arbitrary structures,'' J. Appl. Mech. **42**, 295–300 (1975).
[18] P. Tuong Ha Duong, ''A finite element method for the double-layer potential solutions of the Neumann exterior problem,'' Math. Methods Appl. Sci. 191–208 (1980).
[19] K. Hayami and C. A. Brebbia, ''A new coordinate transformation method for singular and nearly singular integrals over general curved boundary elements,'' *Boundary Elements IX*, edited by C. A. Brebbia, W. L. Wendland and G. Kuhn (Computational Mechanics Publications, Southampton, 1987), Vol. 1, pp. 379–399.
[20] K. Hayami and H. Matsumoto, ''A numerical quadrature for nearly singular boundary element integrals,'' Eng. Anal. Boundary Elem. **13**, 143–154 (1994).
[21] C. Vallance and N. Vlahopolous, ''Structural noise induced vibration prediction for a component of a launch vehicle,'' *Proceedings of Internoise 95*, 1995, pp. 1303–1306.
[22] J. B. Fahnline, ''Condensing structural finite-element meshes into coarser acoustic element meshes,'' J. Acoust. Soc. Am. **104**, 1802 (1998).
[23] M. Guerich and M. A. Hamdi, ''A numerical method for vibro-acoustic problems with incompatible finite element meshes using B-spline functions,'' J. Acoust. Soc. Am. **105**, 1682–1694 (1999).
[24] J. B. Fahnline, ''Fast acoustic power computations for structural optimization problems,'' J. Acoust. Soc. Am. **108**, 2587 (2000).
[25] C. Y. Shih, Y. G. Tsuei, R. J. Allemang, and D. L. Brown, ''Complex mode indicator function and its applications to spatial domain parameter estimation,'' in *Proc. of the VII Int. Modal Anal. Conf.*, 1989, pp. 533–540.
[26] A. W. Phillips, R. J. Allemang, and W. A. Fladung, ''The complex mode indicator function (CMIF) as a parameter estimation method,'' in *Proc. of the XVI Int. Modal Anal. Conf.*, 1998, pp. 705–710.
[27] E. Anderson, Z. Bai, C. Bischof, J. Demmel, and J. Dongarra, LAPACK *User's Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, 1999).
[28] P. A. Businger and G. H. Golub, ''Algorithm 351: Singular value decomposition of a complex matrix,'' Commun. ACM **12**, 564–565 (1969).
[29] D. J. Ewins, *Modal Testing: Theory and Practice* (Wiley, New York, 1984), pp. 184–192.
[30] J. M. Montalvao and N. M. M. Maia, ''Single mode identification techniques for use with small microcomputers,'' J. Sound Vib. **124**, 13–26 (1988).
[31] G. H. Golub and C. F. Van Loan, *Matrix Computations* (John Hopkins University Press, Baltimore, MD, 1989), pp. 246–247.
[32] L. Cremer and M. Heckl, *Structureborne Sound* (Springer-Verlag, New York, 1973), pp. 461–463.
[33] B. L. Clarkson and K. T. Brown, ''Acoustic radiation damping,'' *J. Vib, Acoustics, Stress, and Reliability in Design*, 1985, Vol. 107, pp. 357–360.
[34] U. S. Linholm, D. D. Kana, W. Chu, and H. N. Abramson, ''Elastic vibration characteristics of cantilever plates in water,'' J. Ship Res. **9**, 11–22 (1965).
[35] A. Leissa, *Vibrations of Plates* (Acoustical Society of America, New York, 1993), pp. 76–87.
[36] H. Lamb, ''On the vibrations of an elastic plate in contact with water,'' Proc. R. Soc. London, Ser. A **98**, 205–216 (1920).
[37] A. Leissa, in Ref. 35, pp. 299–300.
[38] A. Leissa, in Ref. 35, pp. 7–8.

Pressure and velocity fields produced by an underwater explosion[a),b)]

Kendall S. Hunter and Thomas L. Geers[c)]
Department of Mechanical Engineering, University of Colorado, Boulder, Colorado 80309

(Received 5 February 2003; revised 22 November 2003; accepted 29 December 2003)

An integrated model for a moderately deep underwater explosion bubble has recently been developed. For the initial shock-wave phase, the model employs an empirical pressure-profile expression for the far field to determine bubble radius as a function of time. For the subsequent oscillation phase, the model accounts for first-order wave effects in both the external liquid and the internal gas. In the present paper, a radiation model consistent with the integrated bubble model is formulated and evaluated. The distant pressure and velocity fields produced during the shock-wave phase derive from the empirical pressure-profile expression. Those produced during the oscillation phase derive from a matched-asymptotic-expansion solution for a translating acoustic source/dipole. Response histories and snapshots for far-field pressure and velocity are computed. Comparisons with existing experimental data exhibit good correlation. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1648680]

PACS numbers: 43.25.Vt, 43.25.Yw [WMC] Pages: 1483–1496

I. INTRODUCTION

In a recent paper [Geers and Hunter (2002)], a bubble model was developed for a moderately deep underwater explosion that integrates the shock-wave and oscillation phases of the motion. For the short shock-wave phase, a *hyperacoustic* relation between bubble volume-acceleration and far-field pressure profile made it possible to determine radial bubble motion from an empirical pressure-profile formula. The relation also provided initial conditions for the subsequent oscillation phase, in which bubble motion was determined by the integration of bubble–surface response equations that incorporated wave effects in both the internal gas and the external liquid. When specialized to bubble dilation and translation, which maintains bubble sphericity, these response equations produced bubble-motion histories that exhibited agreement with experimental data substantially better than that exhibited by previous bubble-motion equations.

In the present paper, we first enhance the empirical pressure-profile formula for the shock-wave phase in order to accommodate several explosives, not just TNT. This enhancement employs the ''late-time'' decay model of Kirkwood and Bethe (1942), as well as late-time pressure-integral data [Price (1979)]. Then, we modify the hyperacoustic relation so as to conform to the enhanced pressure profile. Next, employing the results of Leppington and Levine (1987), we obtain expressions for the distant *acoustic* fields radiated by the spherical bubble dilating and translating in the oscillation phase. In the limit $c\to\infty$, these expressions reduce to equations previously obtained for an incompressible fluid [Hicks (1970)]. Finally, we perform for selected cases, far-field pressure and velocity calculations, making comparisons, when possible, with corresponding measured data [Arons *et al.* (1948); Arons and Yennie (1948); Swift and Decius (1948)]. Agreement is found to be good.

The approach in this paper is based on an inviscid, irrotational model of fluid flow, which admits a velocity potential $\phi(\mathbf{r},t)$, with associated velocity and pressure given by

$$\mathbf{u}(\mathbf{r},t)=\nabla\phi(\mathbf{r},t), \tag{1a}$$

$$p(\mathbf{r},t)=-\rho\left[\frac{\partial\phi(\mathbf{r},t)}{\partial t}+\frac{1}{2}\nabla\phi(\mathbf{r},t)\cdot\nabla\phi(\mathbf{r},t)\right]+p_{\text{atm}}+\rho g d(\mathbf{r}), \tag{1b}$$

where $\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$, in which $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the usual Cartesian unit vectors, ρ is the density of the liquid, p_{atm} is atmospheric pressure, g is the acceleration of gravity, and $d(\mathbf{r})$ is the depth below the free surface of the liquid. Equation (1b) is, of course, Bernoulli's equation; in the far field, the quadratic term is negligible, which produces the acoustic pressure $p_{\text{ac}}(\mathbf{r},t)=-\rho\partial\phi(\mathbf{r},t)/\partial t$.

Of interest here is the far field located beyond a certain radius R, as measured from the center of the explosive charge at detonation. This radius is taken as being at least two times the maximum radius attained by the bubble, a_M, which, in turn, is typically 15 times the radius of the charge, a_c. The shock wave, which is generated at very early time, is modeled as propagating radially and hyperacoustically in the near field $a(t)\leqslant r\leqslant R$, where $r=|\mathbf{r}|$, and radially and *quasiacoustically* in the far field $r>R$. Hyperacoustic propagation is characterized by a wavefront velocity that appreciably exceeds the linear-acoustic speed of sound c; quasiacoustic propagation is characterized by a hyperacoustic profile with a wavefront velocity very close to c. Finally, wave propagation in the far field produced by bubble motion during the oscillation phase is modeled as linear acoustic, but, because of bubble translation, not purely radial.

[a)]Submitted in January 2003 to the Journal of the Acoustical Society of America.
[b)]Presented at the 73rd Shock and Vibration Symposium.
[c)]Electronic mail: geers@spot.colorado.edu

TABLE I. Material constants for shock-wave similitude equations; the units of P_c and v_c are GPa and km/s, respectively [Price (1979)].

Material	P_c	v_c	A_c	B_c
TNT (1.60 g/cc)	1.67	1.01	0.18	0.185
Pentolite (1.71 g/cc)	1.92	1.29	0.194	0.257
HBX-1 (1.72 g/cc)	1.58	1.17	0.144	0.247

TABLE II. Impulse fitting constants for Eq. (4); all five constants are dimensionless.

Material	C_P	C_1	C_2	C_A	C_B
TNT (1.60 g/cc)	0.6129	0.9833	1.838 00	0.1901	0.1521
Pentolite (1.71 g/cc)	0.9221	1.1310	0.129 40	0.2063	0.1238
HBX-1 (1.72 g/cc)	0.9222	1.0990	0.040 03	0.2166	0.1123

II. SHOCK-WAVE-PHASE FIELD

During the short shock-wave phase of bubble motion, there is too little time for buoyancy effects to produce significant translation; hence, as stated above, the shock-wave field is spherically symmetric. Elegant analytical models have been developed for this hyperacoustic field [see, e.g., Lamb (1923); Penney (1940); Penney and Dasgupta (1942); Kirkwood and Bethe (1942); Brinkley and Kirkwood (1947); Rogers (1977)] but they have not fully captured the complexity of the physics. Fortunately, empirical formulas based on extensive experimental data have proven quite robust [Coles *et al.* (1946); Farley and Snay (1978); Price (1979)].

A. Similitude relation

The following *similitude relation* for the shock-wave *pressure profile* has proven accurate over a wide range of the radial distance r [Geers and Hunter (2002)]

$$P(r,t)=P_c(a_c/r)^{1+A_c}f([a_C/r]^{B_c}v_ct/a_c), \tag{2}$$

in which P_c, v_c, A_c, and B_c are constants associated with the charge material, a_c is the equivalent spherical radius of the charge, and $f(\tau)=0$ for $\tau<0$. Some recommended values for the constants appear in Table I; previously used choices for $f(\tau)$ are [Coles *et al.* (1946); Geers and Hunter (2002)]

$$f(\tau)=e^{-\tau}, \quad \tau\leqslant 1 \tag{3}$$

$$f(\tau)=0.8251e^{-1.338\tau}+0.1749e^{-0.1805\tau}, \quad \tau\leqslant 7.$$

In the present development, we require a similitude relation that is accurate over a longer duration and applies to a variety of charge materials. Hence, we replace (2) with

$$P(r,t)=P_c(a_c/r)^{1+A(t)}f([a_c/r]^{B(t)}v_ct/a_c), \tag{4}$$

in which

$$A(t)=\frac{A_c}{1+(C_Av_ct/a_c)^5}, \tag{5a}$$

$$B(t)=\frac{B_c}{1+(C_Bv_ct/a_c)^5}, \tag{5b}$$

$$f(\tau)=C_Pe^{-C_1\tau}+(1-C_P)(1+C_2\tau)^{-4/5}, \tag{5c}$$

where C_P, C_1, C_2, C_A, and C_B are fitting constants for a particular charge material, and, again, $f(\tau)=0$ for $\tau<0$. Equations (4) and (5) provide to a hyperacoustic pulse an acoustic tail with time dependence deduced by Kirkwood and Bethe (1942).

We determined fitting constants by performing least-squares fits to data for the *integral* of the pressure profile $I_p(r,t)$ [Price (1979)]; these constants appear in Table II. The powers of five in (5a) and (b) were also part of the fitting process. Figure 1 shows pressure-integral histories produced by (4), (5), and Tables I and II. Also shown are individual data points from Price (1979); the fits are seen to be excellent over the time range of the data, $0\leqslant(a_c/r)^{B_c}v_ct/a\leqslant 10$. Shown too are pressure-integral histories produced by least-squares fitting with (2) and (3b) [Geers and Hunter (2002)]; while the fit for TNT is good, the fits for Pentolite and HBX-1 are lacking.

The use of (4) instead of (2) affects the calculation of bubble motion as described in Geers and Hunter (2002). A straightforward extension of their development yields the following relation between far-field pressure and bubble motion:

$$P(r,t)=\frac{\rho}{4\pi r}\left(\frac{a_c}{r}\right)^{A(t)}\ddot{V}([a_c/r]^{B(t)}t), \tag{6}$$

where $V(t)$ is the bubble volume. Equating the right side of this equation to the right side of (4), and employing (5c), we obtain for bubble volume acceleration

$$\ddot{V}(t)=\frac{4\pi a_c}{\rho}P_c[C_Pe^{-C_1t/T_c}+(1-C_P)(1+C_2t/T_c)^{-4/5}], \tag{7}$$

where $T_c=a_c/v_c$. Integrating this result twice with initial conditions $\dot{V}(0)=0$ and $V(0)=V_c$, where V_c is the volume of the charge, and using $V(t)=4\pi a^3(t)/3$, we find the bubble's radius $a(t)$ and surface velocity $\dot{a}(t)$ during the shock-wave phase of bubble motion. The values of these quantities at a selected transition time (see subsection C) constitute the initial conditions for the oscillation phase.

B. Far field

The wavefront velocity U of a plane shock wave can be found from the Rankine–Hugoniot conditions as [Cole (1948)]

$$U=\sqrt{\left(\frac{\rho}{\rho_{sw}}\right)\frac{P(r,0)-p}{\rho_{sw}-\rho}}, \tag{8}$$

where unsubscripted quantities refer to ambient conditions, $P(r,0)$ is the wavefront pressure, and ρ_{sw} is the wavefront density, found from the modified Tait equation of state for water [Tomita and Shima (1977)],

$$\rho_{sw}=\rho\left[\frac{P(r,0)+B}{p+B}\right]^{1/7.15}, \tag{9}$$

where $B=304.7$ MPa. For $r>2a_M\approx 30a_c$, $P(r,0)<30$ MPa, so (8) and (9) yield wavefront velocities that exceed c by less than 3%. Hence, waves in the far field propagate at essentially the linear-acoustic speed of sound.

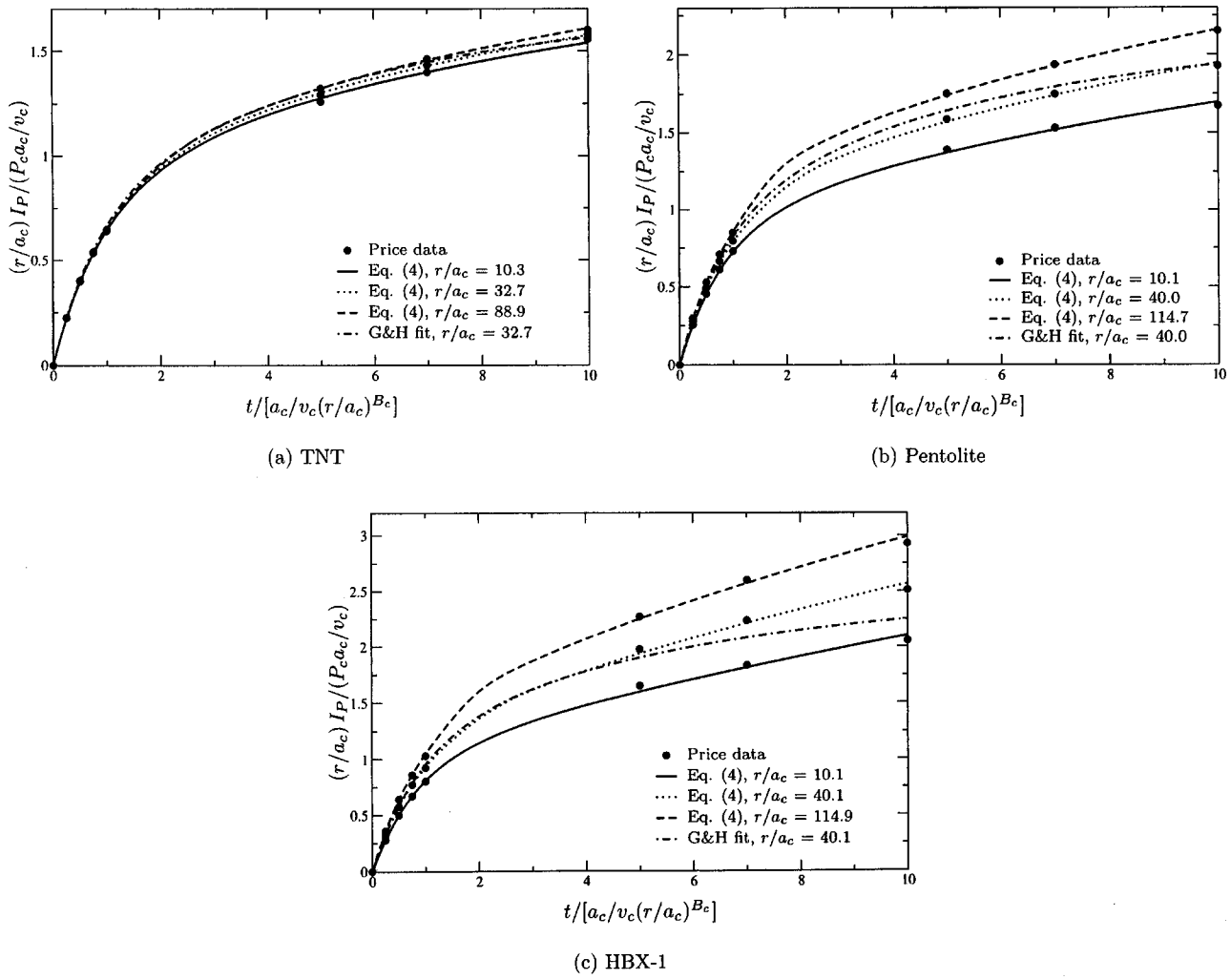


FIG. 1. The integral of (4) for (a) TNT; (b) Pentolite; (c) HBX-1, and the corresponding data from Price (1979).

Quasiacoustic propagation of the pressure profile given by (6) yields, for $r \geqslant R$

$$p(r,t) = \frac{\rho}{4\pi r} H(t^*) \left(\frac{a_c}{r}\right)^{A(t^*)} \ddot{V}([a_c/r]^{B(t^*)} t^*), \qquad (10)$$

in which $H(\,)$ is the Heaviside step function and $t^* = t - (r - R)/c$. Here, R is taken as pertaining to the far-field point of interest closest to the origin, and the finite radius of the bubble, being much smaller than R, is neglected.

Employing the acoustic relation $p_{ac}(r,t) = -\rho \partial \phi(r,t)/\partial t$, we integrate (10) and enforce the initial condition $\phi(r,[r-R]/c) = 0$ to obtain the velocity-potential field for $r \geqslant R$

$$\phi(r,t) = -\frac{1}{4\pi r} \int_0^{t^*} \left(\frac{a_c}{r}\right)^{A(\xi)} \ddot{V}([a_c/r]^{B(\xi)} \xi) d\xi. \qquad (11)$$

Hence, the velocity field may be obtained by partial differentiation of (11) with respect to r which, by Leibniz' rule, yields

$$u_r(r,t) = \frac{H(t^*)}{4\pi r c} \left(\frac{a_c}{r}\right)^{A(t^*)} \ddot{V}([a_c/r]^{B(t^*)} t^*) + \frac{1}{4\pi r^2} \times \int_0^{t^*} \left(\frac{a_c}{r}\right)^{A(\xi)} [1 + A(\xi)] \ddot{V}([a_c/r]^{B(\xi)} \xi) d\xi + \frac{1}{4\pi r^2} \int_0^{t^*} \xi B(\xi) \left(\frac{a_c}{r}\right)^{A(\xi)+B(\xi)} \times \ddot{\ddot{V}}([a_c/r]^{B(\xi)} \xi) d\xi. \qquad (12)$$

For $A(t) = B(t) = 0$, this equation and (10) yield the familiar relation for a spherical acoustic wave

$$u_r(r,t) = \frac{p(r,t)}{\rho c} + \frac{1}{\rho r} \int_0^t p(r,\xi) d\xi, \qquad (13)$$

and (11) yields the velocity-potential field for an acoustic source

$$\phi(r,t) = -\dot{V}(t^*)/4\pi r. \qquad (14)$$

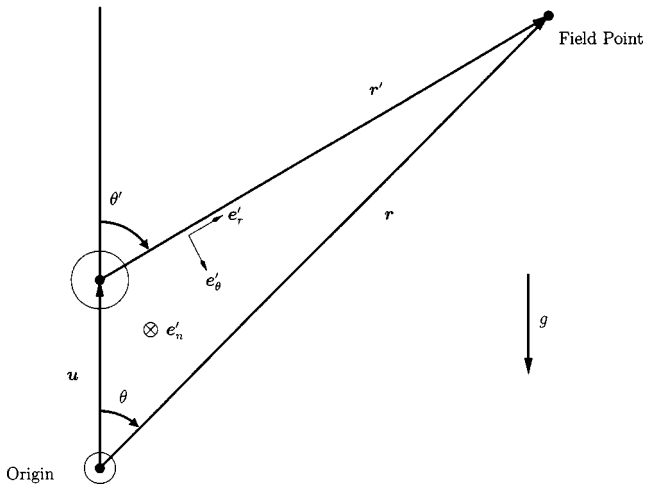


FIG. 2. Geometry in the plane with normal $\mathbf{e}'_n$.

C. Phase-transition criterion

Equations (10) and (12) define the quasiacoustic fields produced in the far field by the underwater-explosion bubble during its shock-wave phase. Still undefined, however, is the duration of that phase, which we shall call t_I. To remedy this deficiency, we observe that (4) approximates an acoustic field at times large enough that $A(t)$ and $B(t)$ are suitably small. From Tables I and II, $A_c<B_c$ and $C_B<C_A$, so from (5a) and (5b), $A(t)<B(t)$; hence, we focus on $B(t)$, establishing our smallness criterion from (4) as

$$(a_c/r)^{B(t_I)}\geqslant 1-\epsilon, \tag{15}$$

where ϵ is a small, positive number. Now, $B(t_I)$ must be small, so a Maclaurin series expansion of the left side of (15) yields $1-B(t_I)\ln(r/a_c)+\cdots$; thus, from (5b), (15) may be written

$$(C_B t_I/T_c)^5\geqslant(B_c/\epsilon)\ln(r/a_c)-1. \tag{16}$$

But $(B_c/\epsilon)\ln(r/a_c)\gg 1$ and $[\ln(r/a_c)]^{1/5}=1.325\pm 0.04$ for $30\leqslant(r/a_c)\leqslant 130$; hence, (16) yields

$$t_I/T_c\geqslant 1.325(B_c/\epsilon)^{1/5}/C_B. \tag{17}$$

For the B_c and C_B values given in Tables I and II, minimum values of t_I/T_c for $\epsilon=0.05$ are 11.3 for TNT, 14.8 for Pentolite, and 16.2 for HBX-1. Hence, a reasonable value for the duration of the shock-wave phase, for all three charge materials, is $t_I=16T_c$. This value falls within the range of the data in Price (1979).

It is shown in Geers and Hunter (2002) that bubble–surface response in the oscillation phase is insensitive to variations in t_I over the range $3T_c\leqslant t_I\leqslant 7T_c$. We have found that the upper limit of this insensitivity extends to $t_I=16T_c$.

III. OSCILLATION-PHASE FIELD

During the oscillation phase of motion, the bubble undergoes both dilation and translation, producing an acoustic far field that is no longer spherically symmetric. Figure 2 shows the geometry for the oscillation-phase field model. The bubble surface moves radially relative to its center with the known displacement $a(t)$, and the bubble center, initially at the origin, translates according to the known displacement $\mathbf{u}(t)=u(t)\mathbf{k}$. From Fig. 2, the vector from the bubble center to a field point $\mathbf{r}$ is given by

$$\mathbf{r}'(\mathbf{r},t)=x\mathbf{i}+y\mathbf{j}+[z-u(t)]\mathbf{k}. \tag{18}$$

Also appearing in Fig. 2 are the unit vectors, $\mathbf{e}'_r$, $\mathbf{e}'_n$, and $\mathbf{e}'_\theta$. The first is

$$\mathbf{e}'_r(\mathbf{r},t)=\frac{\mathbf{r}'(\mathbf{r},t)}{|\mathbf{r}'(\mathbf{r},t)|}, \tag{19}$$

and the second may be expressed as

$$\mathbf{e}'_n(\mathbf{r})=\frac{\mathbf{k}\times\mathbf{e}'_r(\mathbf{r},t)}{|\mathbf{k}\times\mathbf{e}'_r(\mathbf{r},t)|}, \tag{20}$$

which, as indicated by the $\otimes$ symbol, points into the plane of the figure. The third is simply

$$\mathbf{e}'_\theta(\mathbf{r},t)=\mathbf{e}'_n(\mathbf{r})\times\mathbf{e}'_r(\mathbf{r},t). \tag{21}$$

A. Velocity potential

Leppington and Levine (1987) found an expression for the far-field velocity potential generated by a dilating and translating compact sphere that is accurate to $O(M^3)$, where M is a Mach number characterizing the motion. Here, the qualifier "compact" means that the radius of the sphere is small relative to the characteristic pulse width in the fluid, which is the same criterion behind (6). Their potential is expressed as

$$\phi(\mathbf{r},t)=\phi_S\{\mathbf{r},t|Q\}+\phi_D\{\mathbf{r},t|\mu\}, \tag{22}$$

in which $\phi_S\{\mathbf{r},t|Q\}$ is a *translating-source potential* of strength Q, and $\phi_D\{\mathbf{r},t|\mu\}$ is a *translating-dipole potential* of strength μ. The potential strengths Q and μ are given by

$$Q(t)=a^2(t)\dot a(t)\left\{1+\frac{9}{2}\left[\frac{\dot a(t)}{c}\right]^2+5\frac{a(t)\ddot a(t)}{c^2}-\frac{1}{4}\left[\frac{\dot u(t)}{c}\right]^2\right\}+\frac{3a^4(t)\ddot a(t)-a^3(t)\dot u(t)\ddot u(t)}{6c^2}, \tag{23a}$$

$$\mu(t)=\tfrac{1}{2}a^3(t)\dot u(t), \tag{23b}$$

where $\dot a(t)$ and $\dot u(t)$ are the bubble's radial and translational velocities, respectively.

Leppington and Levine used the method of matched asymptotic expansions [see, e.g., Van Dyke (1964)] to obtain both an inner-potential representation, which is a perturbation from the incompressible solution, and the outer-potential representation given by (22) and (23). The outer-potential representation is of primary interest in this paper. However, by neglecting the translation terms, one may readily use their inner-potential representation to obtain various equations of motion (EOM) for the dilating sphere, each accurate to a different power of the dilational Mach number $M_0=\dot a/c$. These EOM, for the first three orders in M_0, correspond to the M_0^0 equation of Lamb (1923), the M_0^1 equations of Herring (1941) and of Keller and Kolodner (1956) [see also Prosperetti and Lezzi (1986) and Geers and Hunter (2002)],

and the M_0^2 equation of Tomita and Shima (1977) [see also Fujikawa and Akamatsu (1980) and Lezzi and Prosperetti (1987)].

The potentials $\phi_S\{\mathbf{r},t|Q\}$ and $\phi_D\{\mathbf{r},t|\mu\}$ define the far field generated by the bubble motion in terms of singularities translating with speed $\dot{u}(t)$ along the z axis. The first potential is the first fundamental solution of the nonhomogeneous wave equation, i.e., the solution of [Morse and Ingard (1968)]

$$\nabla^2\phi_S-\frac{1}{c^2}\frac{\partial^2\phi_S}{\partial t^2}=-4\pi Q(t)\,\delta(x)\,\delta(y)\,\delta(z-u(t)). \tag{24}$$

This solution is Landau and Lifshitz (1951), Chap. 8]

$$\phi_S\{\mathbf{r},t|Q\}=-\frac{Q(t')}{r'(\mathbf{r},t')\lambda(\mathbf{r},t')}, \tag{25}$$

in which

$$r'(\mathbf{r},t')=[x^2+y^2+z'^2(\mathbf{r},t')]^{1/2}, \tag{26a}$$

$$\lambda(\mathbf{r},t')=[1-\dot{u}(t')\gamma'(\mathbf{r},t')/c], \tag{26b}$$

where

$$z'(\mathbf{r},t')=z-u(t'), \tag{27a}$$

$$\gamma'(\mathbf{r},t')=\cos[\theta'(\mathbf{r},t')]=\mathbf{k}\cdot\mathbf{e}_r'(\mathbf{r},t'), \tag{27b}$$

and t' is given implicitly by

$$t'=t-\frac{r'(\mathbf{r},t')-R}{c}. \tag{28}$$

Note that $t'+R/c$ is the time at which the signal arriving at field point $\mathbf{r}$ at time t was generated by the source. Although (28) can yield multiple values for t' if the source is translating at supersonic speeds, the source translation considered here is always subsonic; hence, t' is a single-valued increasing function of t. In the absence of translation, $t'=t^*$.

The moving-dipole potential may be written as

$$\phi_D\{\mathbf{r},t|\mu\}=-[\partial/\partial z\,\phi_S\{\mathbf{r},t|\mu\}]_{t=t'}, \tag{29}$$

which is the negative of Leppington and Levine's expression. Performing the differentiation, we find

$$\begin{aligned}\phi_D\{\mathbf{r},t|\mu\}=&-\frac{\mu(t')\gamma'(\mathbf{r},t')}{r'^2(\mathbf{r},t')\lambda^2(\mathbf{r},t')}\left[1+\frac{\dot{u}(t')}{c\lambda(\mathbf{r},t')}\gamma'(\mathbf{r},t')\right.\\&\left.+\frac{z'(\mathbf{r},t')\ddot{u}(t')-\dot{u}^2(t')}{c^2\lambda(\mathbf{r},t')}\right]\\&+\frac{\mu(t')\dot{u}(t')}{cr'^2(\mathbf{r},t')\lambda^2(\mathbf{r},t')}-\frac{\dot{\mu}(t')\gamma(\mathbf{r},t')}{cr'(\mathbf{r},t')\lambda^2(\mathbf{r},t')},\end{aligned} \tag{30}$$

where $\dot{\mu}(t')=\frac{1}{2}[a^3(t')\ddot{u}(t')+3a^2(t')\dot{a}(t')\dot{u}(t')]$.

B. Velocity potential expressions through $O(M)$

Geers and Hunter (2002) determined bubble–surface motions with equations accurate through terms of $O(M_0)$ and $O(M_1)$, where $M_0=\dot{a}/c$ and $M_1=\dot{u}/c$. At this level of accuracy, the model results were found to be in good agreement with experimental data for two dilational bubble periods. Here, we express (25) and (30) to the same order of accuracy in these surface-velocity Mach numbers. For brevity, t' dependence is suppressed until the end of this section.

By inserting (23a) into (25), and applying the binomial theorem to λ^{-1}, we find

$$\begin{aligned}\phi_S\{\mathbf{r},t|Q\}=&-\left\{\frac{a^2\dot{a}}{r'(\mathbf{r})}\left[1+5\frac{a\ddot{a}}{c^2}\right]+\frac{3a^4\dddot{a}-a^3\dot{u}\ddot{u}}{6c^2r'(\mathbf{r})}\right\}\\&\times\left(1+\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right)+\cdots,\end{aligned} \tag{31}$$

in which the ellipsis $(\cdots)$ denotes terms of $O(M_0^l)$, $O(M_1^l)$ for $l\geqslant 2$, and $O(M_0^mM_1^n)$ for $m,n\geqslant 1$. To resolve (31), we require expressions for $\dddot{a}$, $\ddot{a}$, and $\ddot{u}$ in terms of lower-order derivatives. The appearance of c^{-2} multiplying these derivatives indicates that, in order to maintain a consistent overall error, the expressions need only be accurate to $O(M^0)$ at most. Hence, $\ddot{a}$ and $\ddot{u}$ may be obtained from the incompressible-fluid equations for dilation and translation [Hicks (1970); see also Geers and Hunter (2002)]

$$a\ddot{a}+\tfrac{3}{2}\dot{a}^2-\tfrac{1}{4}\dot{u}^2-gu=P_0/\rho, \tag{32a}$$

$$a\ddot{u}+3\dot{a}\dot{u}-2ga=0, \tag{32b}$$

in which $P_0=K_c(a_c/a)^{3\kappa}-p_{atm}-\rho g\bar{d}$, where K_c and κ are constants for the charge material and $\bar{d}$ is the original depth of the bubble center. The dilational jerk, $\dddot{a}$, may be found by differentiating (32a) with respect to time and then eliminating $\ddot{a}$ and $\ddot{u}$ that appear in the result with (32); this procedure yields

$$\dddot{a}=6\frac{\dot{a}^3}{a^2}-\frac{5\dot{u}^2\dot{a}}{2a^2}+\frac{2g\dot{u}}{a}-\frac{4\dot{a}P_0}{\rho a^2}-\frac{3\kappa K_c\dot{a}}{\rho a^2}\left(\frac{a_c}{a}\right)^{3\kappa}. \tag{33}$$

Solving (32) for $\ddot{a}$ and $\ddot{u}$ and inserting the results, along with (33), into (31), we find

$$\begin{aligned}\phi_S\{\mathbf{r},t|Q\}=&-\left\{\frac{a^2\dot{a}}{r'(\mathbf{r})}\left[1+3\frac{P_0}{\rho c^2}-\frac{3\kappa}{2}\frac{K_c}{\rho c^2}\left(\frac{a_c}{a}\right)^{3\kappa}\right]\right.\\&\left.-\frac{2a^2\dot{u}}{3r'(\mathbf{r})}\left(\frac{ga}{c^2}\right)\right\}\left[1+\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right]+\cdots.\end{aligned} \tag{34}$$

Finally, we recognize that, during the oscillation phase, P_0 and $K_c(a_c/a)^{3\kappa}$ remain small relative to the fluid's bulk modulus ρc^2, and that the Froude number c^2/ga is very large. Thus, the appropriate form of the source potential is simply

$$\phi_S\{\mathbf{r},t|Q\}=-\frac{q}{r'(\mathbf{r})}\left(1+\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right)+\cdots, \tag{35}$$

where the stationary-source strength $q=a^2\dot{a}$.

Next, we insert (23b) into (30) and again apply the binomial theorem to the negative powers of λ to obtain

$$\phi_D\{\mathbf{r},t|\mu\}=-\frac{a^3\dot{u}\gamma'(\mathbf{r})}{2r'^2(\mathbf{r})}\left\{\left[1+2\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right]\right.$$
$$\left.+\left[\frac{\dot{u}}{c}\gamma'(\mathbf{r})+\frac{z'(\mathbf{r})\ddot{u}-\dot{u}^2}{c^2}\right]\left[1+3\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right]\right\}$$
$$+\left[\frac{a^3\dot{u}^2}{2cr'^2(\mathbf{r})}-\frac{(3a^2\dot{a}\dot{u}+a^3\ddot{u})\gamma'(\mathbf{r})}{2cr'(\mathbf{r})}\right]$$
$$\times\left[1+2\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right]+\cdots. \tag{36}$$

To resolve this equation, we must replace $\ddot{u}$ with an expression written in terms of lower-order derivatives. Because $\ddot{u}$ is divided by either c or c^2 in (36), (32b) constitutes an acceptable equation from which to obtain $\ddot{u}$. Inserting $\ddot{u}$ from (32b) into (36), we find

$$\phi_D\{\mathbf{r},t|\mu\}=-\frac{a^3\dot{u}\gamma'(\mathbf{r})}{2r'^2(\mathbf{r})}\left[1+3\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right]\left[1+\frac{2gz'(\mathbf{r})}{c^2}\right]$$
$$+\frac{a^3\dot{u}^2}{2cr'^2(\mathbf{r})}-\frac{(3a^2\dot{a}\dot{u}+a^3\ddot{u})\gamma'(\mathbf{r})}{2cr'(\mathbf{r})}$$
$$-\frac{2a^2\dot{u}\gamma'^2(\mathbf{r})}{r'(\mathbf{r})}\left(\frac{ga}{c^2}\right)+\cdots. \tag{37}$$

The two Froude numbers $c^2/gz'(\mathbf{r})$ and c^2/ga that appear in this equation are very large; hence, we obtain

$$\phi_D\{\mathbf{r},t|\mu\}=-\frac{\mu}{r'^2(\mathbf{r})}\left\{\left[1+3\frac{\dot{u}}{c}\gamma'(\mathbf{r})\right]\gamma'(\mathbf{r})-\frac{\dot{u}}{c}\right\}$$
$$-\frac{\dot{\mu}\gamma'(\mathbf{r})}{cr'(\mathbf{r})}+\cdots, \tag{38}$$

where $\dot{\mu}=\frac{1}{2}(a^3\ddot{u}+3a^2\dot{a}\dot{u})$.

From (22), (35), and (38), the complete expression for the velocity potential through $O(M)$ may be written in Cartesian coordinates (x, y, z) and current time t as

$$\phi(\mathbf{r},t)=-q(t')\mathcal{Q}(\mathbf{r},t')-\mu(t')\mathcal{M}(\mathbf{r},t')$$
$$-\dot{\mu}(t')\mathcal{N}(\mathbf{r},t'), \tag{39}$$

with previously defined functions

$$t'=t-\{\sqrt{x^2+y^2+[z-u(t')]^2}-R\}/c, \tag{40a}$$

$$q(t')=a^2(t')\dot{a}(t'), \tag{40b}$$

$$\mu(t')=\tfrac{1}{2}a^3(t')\dot{u}(t'), \tag{40c}$$

$$\dot{\mu}(t')=\tfrac{1}{2}[a^3(t')\ddot{u}(t')+3a^2(t')\dot{a}(t')\dot{u}(t')], \tag{40d}$$

and newly introduced functions

$$\mathcal{Q}(\mathbf{r},t')=\frac{1}{r'(\mathbf{r},t')}+\frac{z'(\mathbf{r},t')}{r'^2(\mathbf{r},t')}\dot{u}(t')/c, \tag{41a}$$

$$\mathcal{M}(\mathbf{r},t')=\frac{z'(\mathbf{r},t')}{r'^3(\mathbf{r},t')}-\frac{\dot{u}(t')/c}{r'^2(\mathbf{r},t')}+\frac{3z'^2(\mathbf{r},t')}{r'^4(\mathbf{r},t')}\dot{u}(t')/c, \tag{41b}$$

$$\mathcal{N}(\mathbf{r},t')=c^{-1}\frac{z'(\mathbf{r},t')}{r'^2(\mathbf{r},t')}. \tag{41c}$$

We recall that $z'(\mathbf{r},t')=z-u(t')$ and $r'(\mathbf{r},t')=\sqrt{x^2+y^2+z'^2(\mathbf{r},t')^2}$.

C. Small translation relative to standoff

Now, t' is only given implicitly in (40a) for a given field point. This is unnecessary, however, when $(u/r)^2\ll 1$, so it is worthwhile to simplify the computation under this condition. To this end, we find that

$$r'(\mathbf{r},t')=r-\frac{z}{r}u(t')+\cdots. \tag{42}$$

From this and (40a), $u(t')$ may be expanded in a Taylor series as

$$u(t')=u(t)-\frac{\dot{u}(t)}{c}\left[r-R-\frac{z}{r}u(t)\right]+\cdots, \tag{43}$$

which is valid for $(\dot{u}/c)^2\ll 1$. Hence, (40a), (42), and (43) yield the explicit formula

$$t'=t-c^{-1}\left\{(r-R)\left[1+\frac{z}{r}\dot{u}(t)/c\right]-\frac{z}{r}u(t)\right\}+\cdots. \tag{44}$$

D. Pressure and velocity fields

We obtain the acoustic pressure field from $p_{\mathrm{ac}}(\mathbf{r},t')=-\rho\partial\phi(\mathbf{r},t')/\partial t$ and (39) as

$$p_{\mathrm{ac}}(\mathbf{r},t)=\rho\left[q(t')\frac{\partial}{\partial t'}\mathcal{Q}(\mathbf{r},t')+\dot{q}(t')\mathcal{Q}(\mathbf{r},t')\right.$$
$$+\mu(t')\frac{\partial}{\partial t'}\mathcal{M}(\mathbf{r},t')+\dot{\mu}(t')\mathcal{M}(\mathbf{r},t')$$
$$\left.+\dot{\mu}(t')\frac{\partial}{\partial t'}\mathcal{N}(\mathbf{r},t')+\ddot{\mu}(t')\mathcal{N}(\mathbf{r},t')\right]\frac{\partial t'}{\partial t}(\mathbf{r},t'). \tag{45}$$

Here, from (40b) and (d),

$$\dot{q}(t')=a^2(t')\ddot{a}(t')+2a(t')\dot{a}^2(t'), \tag{46a}$$

$$\ddot{\mu}(t')=\tfrac{1}{2}[a^3(t')\dddot{u}(t')+6a^2(t')\dot{a}(t')\ddot{u}(t)$$
$$+3a^2(t')\ddot{a}(t')\dot{u}(t')+6a(t')\dot{a}^2(t')\dot{u}(t')], \tag{46b}$$

and, from (41) and (40a)

$$\frac{\partial}{\partial t'}\mathcal{Q}(\mathbf{r},t')=\frac{z'(\mathbf{r},t')\dot{u}(t')}{r'^3(\mathbf{r},t')}+\frac{z'(\mathbf{r},t')\ddot{u}(t')-\dot{u}^2(t')}{cr'^2(\mathbf{r},t')}$$
$$+\frac{2z'^2(\mathbf{r},t')\dot{u}^2(t')}{cr'^4(\mathbf{r},t')}, \tag{47a}$$

$$\frac{\partial}{\partial t'}\mathcal{M}(\mathbf{r},t') = -\frac{\dot{u}(t')}{r'^3(\mathbf{r},t')} + \frac{3z'^2(\mathbf{r},t')\dot{u}(t')}{r'^5(\mathbf{r},t')} - \frac{\ddot{u}(\mathbf{r},t')}{cr'^2(\mathbf{r},t')} + z'(\mathbf{r},t') \times \frac{3z'(\mathbf{r},t')\ddot{u}(t') - 8\dot{u}^2(t')}{cr'^4(\mathbf{r},t')} + \frac{12z'^3(\mathbf{r},t')\dot{u}^2(t')}{cr'^6(\mathbf{r},t')}, \tag{47b}$$

$$\frac{\partial}{\partial t'}\mathcal{N}(\mathbf{r},t') = -\frac{\dot{u}(t')}{cr'^2(\mathbf{r},t')} + \frac{2z'(\mathbf{r},t')\dot{u}(t')}{cr'^4(\mathbf{r},t')}, \tag{47c}$$

$$\frac{\partial t'}{\partial t}(\mathbf{r},t') = \left[1 - \frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\dot{u}(t')/c\right]^{-1} = 1 + \frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\dot{u}(t')/c + \cdots. \tag{47d}$$

We determine the x component of fluid velocity from $u_x(\mathbf{r},t') = \partial\phi(\mathbf{r},t')/\partial x$ and (39) as

$$u_x(\mathbf{r},t) = -\left[q(t')\frac{\partial}{\partial x}\mathcal{Q}(\mathbf{r},t') + \dot{q}(t')\mathcal{Q}(\mathbf{r},t')\frac{\partial t'}{\partial x}(\mathbf{r},t') + \mu(t')\frac{\partial}{\partial x}\mathcal{M}(\mathbf{r},t') + \dot{\mu}(t')\mathcal{M}(\mathbf{r},t')\frac{\partial t'}{\partial x}(\mathbf{r},t) + \dot{\mu}(t')\frac{\partial}{\partial x}\mathcal{N}(\mathbf{r},t') + \ddot{\mu}(t')\mathcal{N}(\mathbf{r},t')\frac{\partial t'}{\partial x}(\mathbf{r},t')\right], \tag{48}$$

where, from (41) and (40a)

$$\frac{\partial}{\partial x}\mathcal{Q}(\mathbf{r},t') = -\frac{x}{r'^3(\mathbf{r},t')}\left[1 + 2\frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\dot{u}(t')/c\right] + \frac{\partial}{\partial t'}\mathcal{Q}(\mathbf{r},t')\cdot\frac{\partial t'}{\partial x}(\mathbf{r},t'), \tag{49a}$$

$$\frac{\partial}{\partial x}\mathcal{M}(\mathbf{r},t') = -\frac{x}{r'^4(\mathbf{r},t')}\left[3\frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')} - 2\dot{u}(t')/c + \frac{12z'^2(\mathbf{r},t')\dot{u}(t')/c}{r'^2(\mathbf{r},t')}\right] + \frac{\partial}{\partial t'}\mathcal{M}(\mathbf{r},t')\cdot\frac{\partial t'}{\partial x}(\mathbf{r},t'), \tag{49b}$$

$$\frac{\partial}{\partial x}\mathcal{N}(\mathbf{r},t') = -\frac{2xz'(\mathbf{r},t')}{cr'^4(\mathbf{r},t')} + \frac{\partial}{\partial t'}\mathcal{N}(\mathbf{r},t')\cdot\frac{\partial t'}{\partial x}(\mathbf{r},t'), \tag{49c}$$

$$\frac{\partial t'}{\partial x}(\mathbf{r},t') = -\frac{x/c}{r'(\mathbf{r},t')}\left[1 - \frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\dot{u}(t')/c\right]^{-1} = -\frac{x/c}{r'(\mathbf{r},t')}\left[1 + \frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\dot{u}(t')/c\right] + \cdots. \tag{49d}$$

The y component of velocity, $u_y(\mathbf{r},t)$, is obtained by replacing x with y in (48) and (49).

Finally, we determine the z component of fluid velocity from $u_z(\mathbf{r},t) = \partial\phi(\mathbf{r},t')/\partial z$ and (39), which yields (48) with x replaced by z, and with

$$\frac{\partial}{\partial z}\mathcal{Q}(\mathbf{r},t') = -\frac{z'(\mathbf{r},t')}{r'^3(\mathbf{r},t')} + \frac{\dot{u}(t')/c}{r'^2(\mathbf{r},t')}\left\{1 - 2\left[\frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\right]^2\right\} + \frac{\partial}{\partial t'}\mathcal{Q}(\mathbf{r},t')\cdot\frac{\partial t'}{\partial z}(\mathbf{r},t'), \tag{50a}$$

$$\frac{\partial}{\partial z}\mathcal{M}(\mathbf{r},t') = \frac{1}{r'^3(\mathbf{r},t')}\left\{1 - 3\left[\frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\right]^2\right\} + \frac{4z'(\mathbf{r},t')}{r'^4(\mathbf{r},t')}\left\{2 - 3\left[\frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\right]^2\right\}\dot{u}(t')/c + \frac{\partial}{\partial t'}\mathcal{M}(\mathbf{r},t')\cdot\frac{\partial t'}{\partial z}(\mathbf{r},t'), \tag{50b}$$

$$\frac{\partial}{\partial z}\mathcal{N}(\mathbf{r},t') = \frac{1}{cr'^2(\mathbf{r},t')}\left\{1 - 2\left[\frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\right]^2\right\} + \frac{\partial}{\partial t'}\mathcal{N}(\mathbf{r},t')\cdot\frac{\partial t'}{\partial z}(\mathbf{r},t'), \tag{50c}$$

$$\frac{\partial t'}{\partial z}(\mathbf{r},t') = -\frac{z'(\mathbf{r},t')}{cr'(\mathbf{r},t')}\left[1 - \frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\dot{u}(t')/c\right]^{-1} = -\frac{z'(\mathbf{r},t')}{cr'(\mathbf{r},t')}\left[1 + \frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\dot{u}(t')/c\right] + \cdots. \tag{50d}$$

Two remarks are appropriate at this point. First, the fields produced by an incompressible-fluid field-generation model [Taylor (1942); Shiffman and Friedman (1943); Hicks (1970)] are obtained by simply letting $c\to\infty$ in (41), (47), (49), and (50). Second, it is often more desirable to have the axisymmetric velocity fields expressed in terms of the primed coordinate system shown in Fig. 2. The expressions for $u'_r(\mathbf{r},t)$ and $u'_\theta(\mathbf{r},t)$ are given in the following section.

E. Velocity field in a moving spherical coordinate system

Expressions for the fluid–velocity components in the moving spherical-coordinate system of the bubble are obtained here through transformation of the Cartesian velocity components. The transformation matrix follows from the unit-vector definitions (19)–(21), which are written here as

$$\mathbf{e}'_r(\mathbf{r},t') = \gamma'_x(\mathbf{r},t')\mathbf{i} + \gamma'_y(\mathbf{r},t')\mathbf{j} + \gamma'_z(\mathbf{r},t')\mathbf{k}, \tag{51a}$$

$$\mathbf{e}'_\theta(\mathbf{r},t') = \frac{r'(\mathbf{r},t')}{\sqrt{x^2+y^2}}\{\gamma'_x(\mathbf{r},t')\gamma'_z(\mathbf{r},t')\mathbf{i} + \gamma'_y(\mathbf{r},t')\gamma'_z(\mathbf{r},t')\mathbf{j} - [\gamma'^2_x(\mathbf{r},t') + \gamma'^2_y(\mathbf{r},t')]\mathbf{k}\}, \tag{51b}$$

$$\mathbf{e}'_n(\mathbf{r}) = \frac{1}{\sqrt{x^2+y^2}}[-y\mathbf{i} + x\mathbf{j}], \tag{51c}$$

where the direction cosines γ_x', γ_x', and γ_z' are given by

$$\gamma_x'(\mathbf{r},t') = \frac{x}{r'(\mathbf{r},t')}, \tag{52a}$$

$$\gamma_y'(\mathbf{r},t') = \frac{y}{r'(\mathbf{r},t')}, \tag{52b}$$

$$\gamma_z'(\mathbf{r},t') = \frac{z'(\mathbf{r},t')}{r'(\mathbf{r},t')}. \tag{52c}$$

Note that (52c) is the same as (27b). Multiplying (51b) and (c) by $\sqrt{x^2+y^2}/r'(\mathbf{r},t')$ and $\sqrt{x^2+y^2}$, respectively, and substituting velocity quantities for unit-vector quantities, we obtain by inspection the transformation of velocity from the Cartesian system to the primed system as

$$\begin{bmatrix} u_r'(\mathbf{r},t) \\ u_\theta'(\mathbf{r},t)\sqrt{x^2+y^2}/r'(\mathbf{r},t) \\ u_n'(\mathbf{r},t)\sqrt{x^2+y^2} \end{bmatrix} = \begin{bmatrix} \gamma_x' & \gamma_y' & \gamma_z' \\ \gamma_x'\gamma_z' & \gamma_y'\gamma_z' & -(\gamma_x'^2+\gamma_y'^2) \\ -y & x & 0 \end{bmatrix}_{\mathbf{r},t'} \begin{bmatrix} u_x(\mathbf{r},t) \\ u_y(\mathbf{r},t) \\ u_z(\mathbf{r},t) \end{bmatrix}. \tag{53}$$

Because the velocity field is axisymmetric, $u_n'=0$.

From (48) we may write the Cartesian velocity components in the general form

$$u_{x_i}(r,t) = -\Bigg[q(t')\frac{\partial}{\partial x_i}\mathcal{Q}(\mathbf{r},t') + \dot{q}(t')\mathcal{Q}(\mathbf{r},t')\frac{\partial t'}{\partial x_i}(\mathbf{r},t') + \mu(t')\frac{\partial}{\partial x_i}\mathcal{M}(\mathbf{r},t') + \dot{\mu}(t')\mathcal{M}(\mathbf{r},t')\frac{\partial t'}{\partial x_i}(\mathbf{r},t') + \dot{\mu}(t')\frac{\partial}{\partial x_i}\mathcal{N}(\mathbf{r},t') + \ddot{\mu}(t')\mathcal{N}(\mathbf{r},t')\frac{\partial t'}{\partial x_i}(\mathbf{r},t') \Bigg], \tag{54}$$

in which $x_i=(x,y,z)$. The partial derivatives in this equation, which are given by Eqs. (49) and (50), may be written in terms of the direction cosines (52). Proceeding first with the x component of velocity, we employ (49d), (52), and (26) to find

$$\frac{\partial t'}{\partial x}(\mathbf{r},t) = -\frac{\gamma_x'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}. \tag{55}$$

The remaining three partial x derivatives may be similarly treated by introducing (52) and (55) into (49a)–(c), which yields, after some manipulation

$$\frac{\partial}{\partial x}\mathcal{Q}(\mathbf{r},t') = -\frac{\gamma_x'(\mathbf{r},t')}{r'(\mathbf{r},t')}\mathcal{Q}(\mathbf{r},t') - \frac{\gamma_x'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}\cdot\frac{\partial}{\partial t'} \times \mathcal{Q}(\mathbf{r},t') + \frac{\dot{u}(t')}{cr'^2(\mathbf{r},t')}\gamma_x'(\mathbf{r},t')\gamma_z'(\mathbf{r},t'), \tag{56a}$$

$$\frac{\partial}{\partial x}\mathcal{M}(\mathbf{r},t') = -\frac{2\gamma_x'(\mathbf{r},t')}{r'(\mathbf{r},t')}\mathcal{M}(\mathbf{r},t') - \frac{\gamma_x'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}\cdot\frac{\partial}{\partial t'}\mathcal{M}(\mathbf{r},t') + \frac{1}{r'^3(\mathbf{r},t')} \times\left[1+6\frac{\dot{u}(t')}{c}\gamma_x'(\mathbf{r},t')\right]\gamma_x'(\mathbf{r},t')\gamma_z'(\mathbf{r},t'), \tag{56b}$$

$$\frac{\partial}{\partial x}\mathcal{N}(\mathbf{r},t') = -\frac{\gamma_x'(\mathbf{r},t')}{r'(\mathbf{r},t')}\mathcal{N}(\mathbf{r},t') - \frac{\gamma_x'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}\cdot\frac{\partial}{\partial t'} \times\mathcal{N}(\mathbf{r},t') + \frac{1}{cr'(\mathbf{r},t')}\gamma_x'(\mathbf{r},t')\gamma_z'(\mathbf{r},t'). \tag{56c}$$

The corresponding partial y derivatives are given by (55) and (56) with x replaced by y. To obtain the partial z derivatives, we first introduce (52) and (26b) into (50d) to get

$$\frac{\partial t'}{\partial z}(\mathbf{r},t') = -\frac{\gamma_z'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}. \tag{57}$$

Then, we introduce this equation, along with (52), into (50a)–(c), and manipulate the resulting equations to obtain

$$\frac{\partial}{\partial z}\mathcal{Q}(\mathbf{r},t') = -\frac{\gamma_z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\mathcal{Q}(\mathbf{r},t') - \frac{\gamma_z'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}\cdot\frac{\partial}{\partial t'} \times\mathcal{Q}(\mathbf{r},t') + \frac{\dot{u}(t')}{cr'^2(\mathbf{r},t')}[1-\gamma_z'^2(\mathbf{r},t')], \tag{58a}$$

$$\frac{\partial}{\partial z}\mathcal{M}(\mathbf{r},t') = -\frac{2\gamma_z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\mathcal{M}(\mathbf{r},t') - \frac{\gamma_z'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}\cdot\frac{\partial}{\partial t'} \times\mathcal{M}(\mathbf{r},t') + \frac{1}{r'^3(\mathbf{r},t')} \times\left[1+6\frac{\dot{u}(t')}{c}\gamma_z'(\mathbf{r},t)\right][1-\gamma_z'^2(\mathbf{r},t')], \tag{58b}$$

$$\frac{\partial}{\partial z}\mathcal{N}(\mathbf{r},t') = -\frac{\gamma_z'(\mathbf{r},t')}{r'(\mathbf{r},t')}\mathcal{N}(\mathbf{r},t') - \frac{\gamma_z'(\mathbf{r},t')}{c\lambda(\mathbf{r},t')}\cdot\frac{\partial}{\partial t'} \times\mathcal{N}(\mathbf{r},t') + \frac{1}{cr'(\mathbf{r},t')}[1-\gamma_z'^2(\mathbf{r},t')]. \tag{58c}$$

The two nonzero velocity components in the primed coordinate system are now obtained by the following procedure. Equations (55) and (56) are introduced into (54) to obtain $u_x(\mathbf{r},t)$; $u_y(\mathbf{r},t)$ is obtained in similar fashion. Next, $u_z(\mathbf{r},t)$ is found by inserting (57) and (58) into (54). The resulting expressions for the three Cartesian velocities are then introduced into (53), and, through matrix multiplication and use of the identity $\gamma_x'^2(\mathbf{r},t')+\gamma_y'^2(\mathbf{r},t')+\gamma_z'^2(\mathbf{r},t')=1$, the radial velocity is found to be

$$u_r'(\mathbf{r},t)=\frac{q(t')}{r'(\mathbf{r},t')}\mathcal{Q}(\mathbf{r},t')+\frac{1}{c\lambda(\mathbf{r},t')}\left[q(t')\frac{\partial}{\partial t'}\mathcal{Q}(\mathbf{r},t')+\dot{q}(t')\mathcal{Q}(\mathbf{r},t')\right]+\frac{2\mu(t')}{r'(\mathbf{r},t')}\mathcal{M}(\mathbf{r},t')+\frac{1}{c\lambda(\mathbf{r},t')}\left[\mu(t')\frac{\partial}{\partial t'}\mathcal{M}(\mathbf{r},t')+\dot{\mu}(t')\mathcal{M}(\mathbf{r},t')\right]+\frac{\dot{\mu}(t')}{r'(\mathbf{r},t')}\mathcal{N}(\mathbf{r},t')+\frac{1}{c\lambda(\mathbf{r},t')}\left[\dot{\mu}(t')\frac{\partial}{\partial t'}\mathcal{N}(\mathbf{r},t')+\ddot{\mu}(t')\mathcal{N}(\mathbf{r},t')\right], \tag{59}$$

and the meridional velocity is found to be

$$u_\theta'(\mathbf{r},t)=\left\{\frac{q(t')\dot{u}(t')}{cr'^2(\mathbf{r},t')}+\frac{\mu(t')}{r'^3(\mathbf{r},t')}\left[1+6\frac{\dot{u}(t')}{c}\gamma_z'(\mathbf{r},t')\right]+\frac{\ddot{\mu}(t')}{cr'(\mathbf{r},t')}\right\}\sin\theta'(\mathbf{r},t'), \tag{60}$$

in which $\sin\theta'(\mathbf{r},t')=\sqrt{1-\gamma_z'^2(\mathbf{r},t')}$.

IV. COMPUTATIONAL RESULTS

In this section, we show bubble-motion histories, as well as histories and *snapshots* of far-field pressure and fluid velocity, for underwater explosions in seawater, treated as both a compressible fluid ($\rho=1025$ kg/m^3, $c=1524$ m/s) and as an incompressible fluid ($\rho=1025$ kg/m^3, $c=\infty$). The term ''snapshot'' is used here to refer to a plot of a far-field quantity as a function of the radial distance from the bubble at a specific instant in time. Pressure, displacement, and time are normalized to the quantities

$$\overline{p}=p_{\text{atm}}+\rho g\overline{d},\quad \overline{a}=a_c(K_c/\overline{p})^{1/3\kappa},\quad \overline{t}=\overline{a}/\sqrt{\overline{p}/\rho}. \tag{61}$$

A. Small, deep bubble

Comparisons are made for this case with experimental data from Arons *et al.* (1948), Arons and Yennie (1948), and Swift and Decius (1948). These data consist of maximum bubble radii, times of minimum bubble radius, and pressure histories. With m_c as charge mass, the data from the first two sources pertain to the ''reduced standoff'' $m_c/R=0.887$ kg$^{1/3}$/m (0.352 lb$^{1/3}$/ft) for 0.229-kg (0.5-lb), 1.137-kg (2.5-lb), and 5.448-kg (12.0-lb) charges of 1.5-g/cc TNT detonated at depths of 76.2 m (250 ft) and 152.4 m (500 ft). The data from Swift and Decius (1948) pertain to 19 0.30-kg (0.66-lb) shots of 1.5-g/cc TNT detonated at depths of 91.5 m (300 ft), and 167.6 m (550 ft). The bubbles in all of these tests exhibited maximum radii less than 1 meter and translated less than 1/2% of their detonation depths. Such translation caused less than a 5% change in r at any field point where measurements were taken. Hence, the effects of bubble translation were found to be negligible. The ranges of the normalization values for this case are 845 KPa (123 psi)$\leqslant\overline{p}\leqslant$1.59 MPa (231 psi), 17.2 cm (6.76 in.)$\leqslant\overline{a}\leqslant$58.0 cm (22.9 in.), and 4.3 ms$\leqslant\overline{t}\leqslant$20 ms.

Figure 3 shows bubble-radius histories for a spherical charge computed from (7) during the shock-wave phase ($0\leqslant t\leqslant 16T_c$) and, for $u(t)=0$, with three different dilation EOM during the oscillation phase ($t>16T_c$) [Geers and Hunter (2002)]. The first is Lamb's incompressible-flow EOM (1923), the second is the EOM of Keller and Kolodner (1956), denoted by K&K, and the third is the EOM of Geers and Hunter (2002), denoted by G&H. Also shown in Fig. 3 are boxes consisting of horizontal lines that mark measured bubble maxima and vertical lines that mark the times of measured bubble minima.

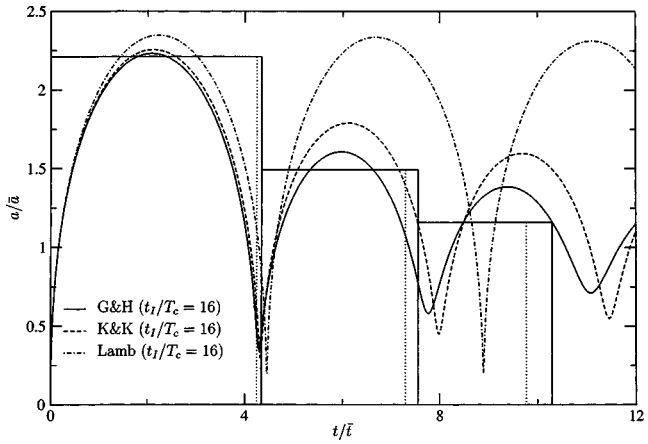


FIG. 3. Dilation bubble-displacement histories produced by three dilation EOM.

Figure 4 shows calculated pressure histories for $t>16T_c$ (oscillation phase) at a standoff $r=R=4.0\overline{a}$. The pressure histories are produced by the three bubble histories of Fig. 3 and by (45) with $\dot{u}=0$. The range of T_c is 0.51 ms$\leqslant T_c\leqslant$1.46 ms. For consistency, $c=\infty$ was used in (41) and (47) to perform the pressure calculations for the Lamb history, whereas $c=152$ m/s was used to perform the calculations for the Keller and Kolodner (K&K) and the Geers and Hunter (G&H) histories. Shown in small circles are pressure-history data from Arons and Yennie (1948), which dictated the rather small standoff being considered. Also appearing in Fig. 4 are values for the integrals of pressure over the first and second bubble pulses (I_1 and I_2, respectively), each of which extends from a zero crossing with a positive slope to the following zero crossing, which has a negative slope. It is worth remarking that, even at this small standoff, the dynamic pressure $p_{\text{ac}}-\rho(\nabla\phi\cdot\nabla\phi)/2$ is indistinguishable from the acoustic pressure.

We see in Fig. 4 that the first G&H bubble pulse agrees very well with the corresponding measured pulse. The second G&H bubble pulse also agrees well with its measured counterpart, although it occurs a little late. The first K&K bubble pulse yields an accurate pressure-integral value, but the pressure history itself is not accurate. The second K&K bubble pulse and both Lamb bubble pulses are unsatisfactory.

The deficiency of the Lamb bubble pulses in Fig. 4 is not due to the use of $c=\infty$ in (41) and (47) when generating the pressure field from the Lamb dilation history in Fig. 3. This is demonstrated in Fig. 5, which shows pressure histories generated from the G&H dilation history in Fig. 3 by means of (45), (41), and (47) with $c=1524$ m/s and with $c=\infty$. We see that the $c=1524$-m/s pressure histories are de-

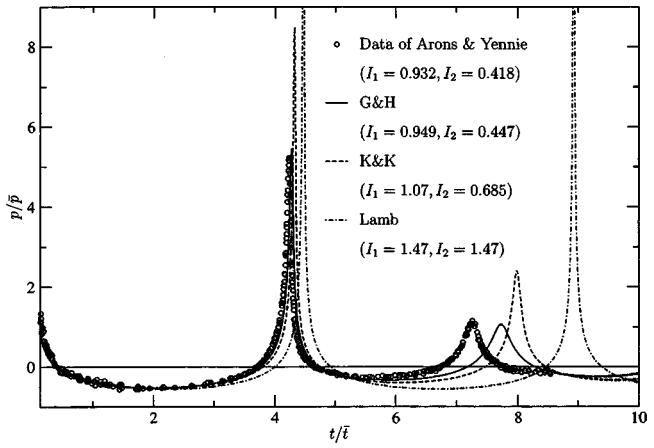


FIG. 4. Pressure histories produced by the bubble-displacement histories of Fig. 3 and (45) with $u(t)=0$; $r=R=4\bar{a}$.

layed in time by $(r-R)/c$ relative to the $c=\infty$ pressure histories, which is of little significance. Hence, we conclude that the calculation of a pressure history at a given point may be accomplished satisfactorily with the incompressible-flow generator consisting of (45), (41), and (47) with $c=\infty$.

The conclusion of the previous paragraph does not mean, however, that the calculation of a pressure *field* over an extended region may be accomplished satisfactorily with the incompressible-flow field generator. This is demonstrated in Fig. 6, which shows pressure snapshots produced by means of (45), (41), and (47) with $c=1524$ m/s and $c=\infty$. We see that the $c=1524$-m/s snapshots exhibit true wave-propagation behavior, with the pulse decaying as $1/r$ while moving across the window $5\leqslant r/\bar{a}\leqslant 15$ at the speed of sound. In contrast, the $c=\infty$ snapshots merely exhibit the instantaneous $1/r$ decay characteristic of an incompressible flow field. If one is interested in the field only in the annular region $R-\Delta/2\leqslant r\leqslant R+\Delta/2$, and if the spatial pulse width, $W=cT$ (where T is the temporal width), is much larger than Δ, the incompressible pulse generator is adequate. In Fig. 6, $R=5\bar{a}$ and $W\approx 3\bar{a}$; hence, one would expect the incompressible flow-field generator to produce satisfactory fields in the annular region with midsurface $R=5\bar{a}$ and thickness $\Delta=0.3\bar{a}$.

Figure 7 shows calculated radial-velocity histories at $r=R=4.0\bar{a}$ produced by (12) for the shock-wave phase, and by (59) with $u(t)=0$ and the three dilation histories of Fig. 3 for the oscillation phase. Here too, the fluid–velocity calculation based on the Lamb dilation history pertains to $c=\infty$, whereas those based on the K&K and G&H dilation histories pertain to $c=1524$ m/s. Unfortunately, no experimental data for fluid velocity exist for this case. We see in Fig. 7 that the K&K and G&H velocity histories are fairly close to one another, and that the Lamb history deviates quite severely from the other two after the first bubble pulse.

B. Large bubble at intermediate depth

As mentioned above, Figs. 3–7 pertain to a situation in which bubble translation is negligible; such depths are characterized by $\bar{p}/\rho g\bar{a}>65$ [Snay (1962); Snay and Tipton

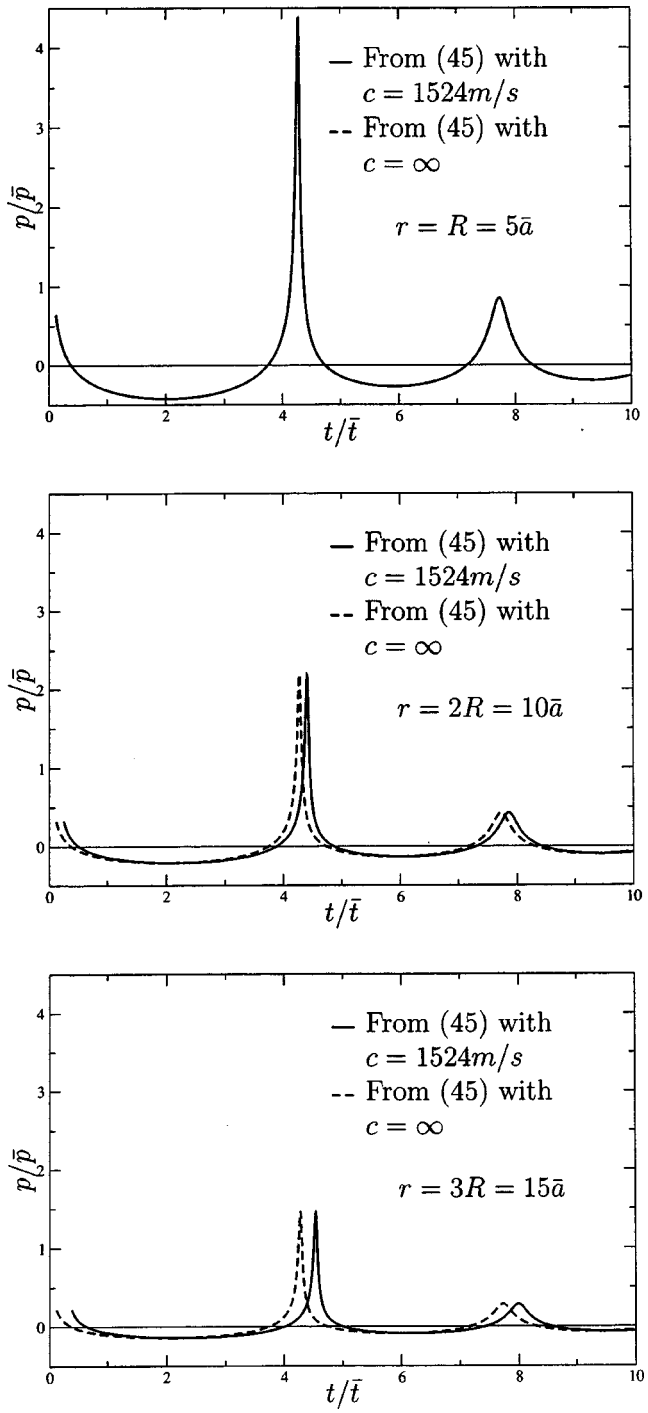


FIG. 5. Pressure histories produced by the G&H bubble-displacement history of Fig. 3 and (45) with $u(t)=0$ at three field locations; $R=5\bar{a}$.

(1962)]. At shallow depths, for which $\bar{d}<10\bar{a}$, free-surface effects are pronounced [Taylor and Davies (1942)], and must be included in the bubble model. At intermediate depths, bubble translation is significant, but the effects of the free surface on bubble motion are minor. Here, we consider an intermediate-depth situation: a 226.7-kg (500-lb) charge of 1.5-g/cc TNT detonated at a depth of 45.7 m (150 ft). This bubble has an initial-depth equilibrium radius $\bar{a}=2.25$ m (7.39 ft), and translates 3.35 m (11 ft) between the times of its first and second radial minima. Hence, $\bar{d}/\bar{a}=20$, $\bar{p}=547$ KPa (79.3 psi), $\bar{t}=96.3$ msec, and $\bar{p}/\rho g\bar{a}=24$.

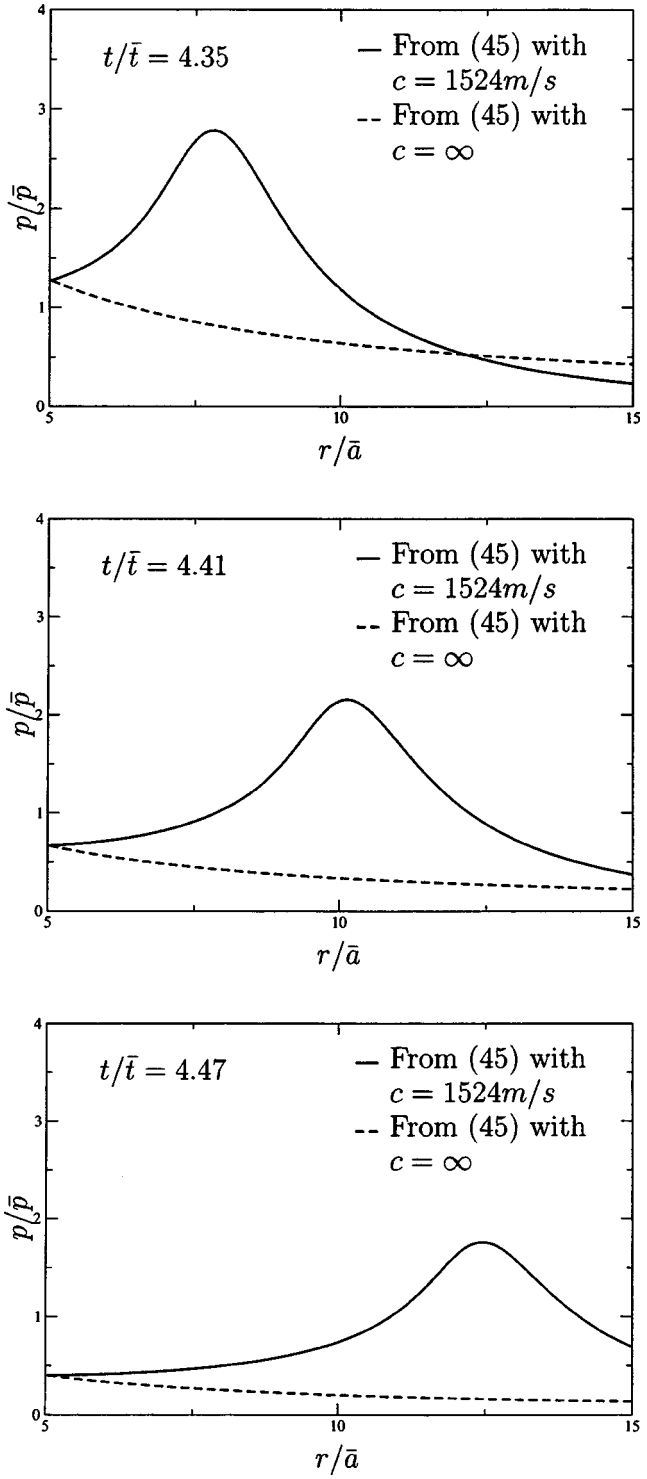


FIG. 6. Pressure snapshots produced by the G&H bubble-displacement history of Fig. 3 and (45) with $u(t)=0$ at three times; $R=5\bar{a}$.

Shown in Fig. 8 are dilation and translation histories for this bubble, as computed with (7) for the shock-wave phase, and with Eqs. (67) and (69) in Geers and Hunter (2002) for the oscillation phase. Because of bubble distortion, translation is not well predicted by a spherical-bubble model. Hence, form drag has been introduced to reduce bubble translation, parametrized by the drag coefficient C_D [Hicks (1970)]. In Fig. 8(a), $C_D=0$ yields good agreement with experimental data for dilation, but in Fig. 8(b) it yields poor

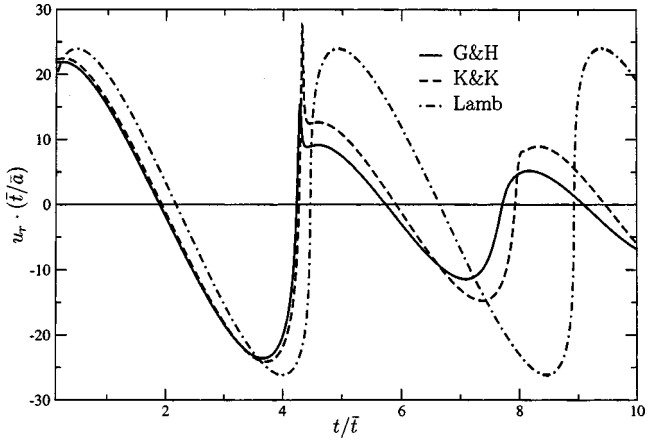


FIG. 7. Radial-velocity histories, produced by the bubble-displacement histories of Fig. 3 and (59) with $u(t)=0$; $r=R=4\bar{a}$.

agreement with experimental data for translation after the time of first bubble minimum. In Fig. 8(b), $C_D=1$ yields good agreement with experimental data for translation, but in Fig. 8(a) it yields poor agreement with experimental data for dilation after the time of first bubble minimum. There is no value of C_D that produces satisfactory results for both dilation and translation, which illustrates the deficiency of a spherical bubble model at intermediate depths.

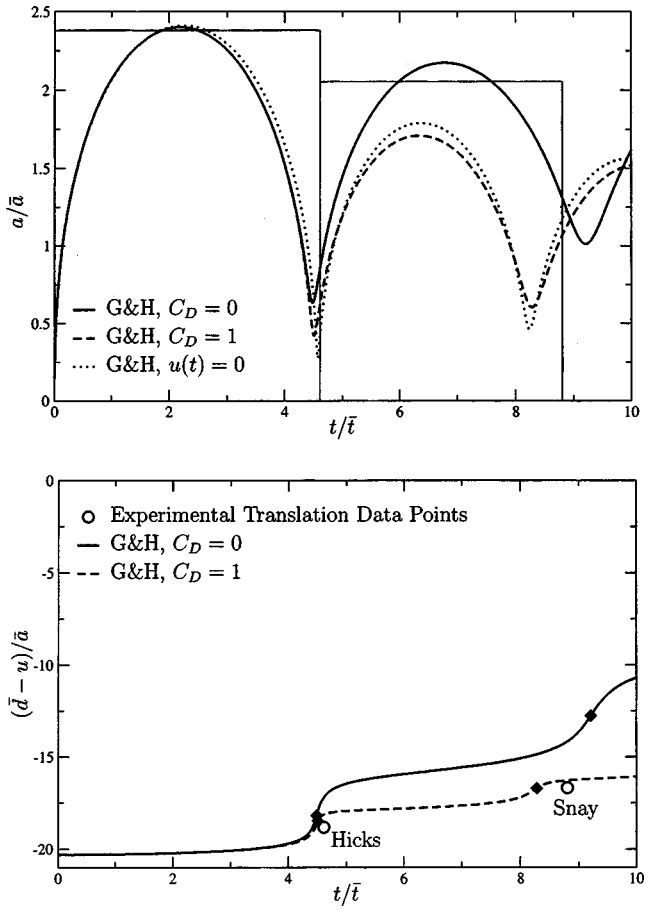


FIG. 8. Dilation and translation bubble-displacement histories from (67) and (69) in Geers and Hunter (2002), compared with experimental data from Snay (1962) and Hicks (1970).

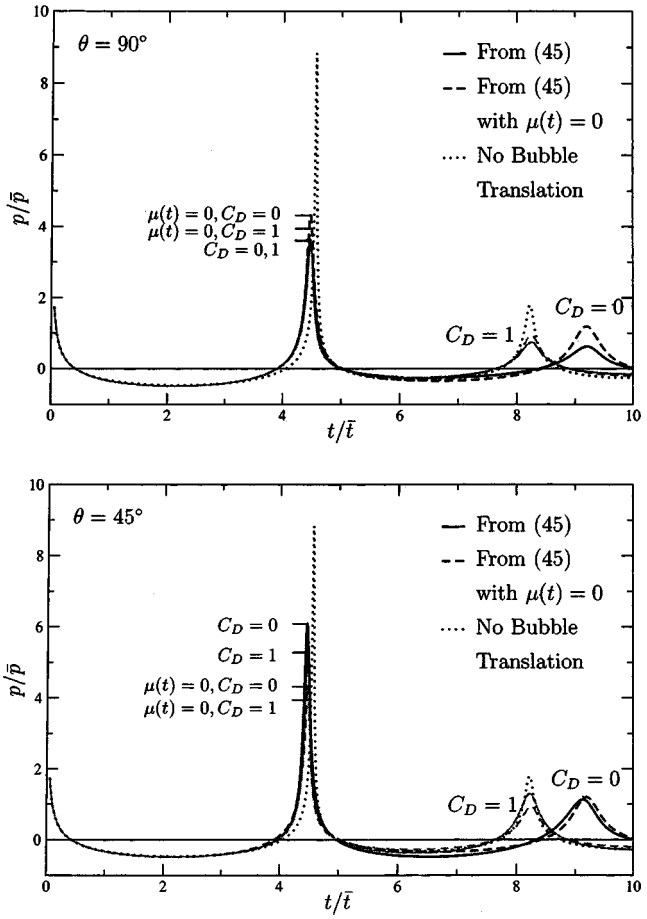


FIG. 9. Pressure histories produced with the bubble displacement histories of Fig. 8 and (45); $r=R=5\bar{a}$.

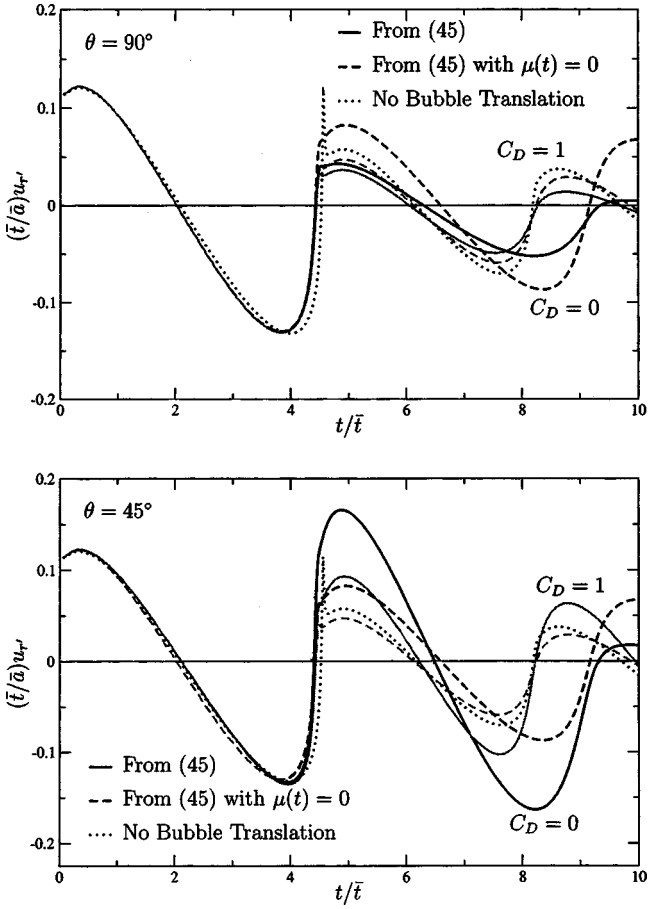


FIG. 10. Radial-velocity histories, in the moving coordinate system, produced with the bubble-displacement histories of Fig. 8 and (59); $r=R=5\bar{a}$.

Figure 9 shows pressure histories based on the $C_D=0$ and $C_D=1$ histories of Fig. 8. The solid curves are the product of (45) in their entirety, whereas the dashed curves are the product of (45) with $\mu(t)$ and its derivatives set to zero. Hence, the dashed curves pertain to computations in which translation is included in the calculation of bubble motion but not in the calculation of radiated pressure. The dotted curves pertain to computations in which translation is neglected altogether. We observe in Fig. 9 that the effects of varying C_D between zero and one (dark curves vs their light counterparts) are manifested principally in the timing of the second bubble pulse. We also observe that neglecting the dipole contribution to the radiated pressure can produce significant error in calculated pulse peaks (dashed curves vs their solid counterparts).

Looking further at Fig. 9, we see that neglecting translation in the calculation of bubble motion produces very large error (dotted curves vs their solid counterparts), as previously found by Hicks (1970). The reason for this resides in the response valleys at $t/\bar{t}=4.6$ in Fig. 8(a). We observe that the $C_D=0$ and $C_D=1$ valleys are about equally sharp, but that the valley for $u(t)=0$ is much sharper, thereby producing a higher pressure pulse. The same is true for the second set of valleys, but to a lesser degree. Finally, by comparing the solid curves in Fig. 9(a) with their counterparts in Fig. 9(b), we conclude that dipole radiation makes a significant contribution to far-field pressure.

Figure 10 shows radial velocity histories u_r' from (59) that correspond to the pressure histories in Fig. 9. At a given time t, u_r' points in the direction of the vector from the bubble center at time t' to the field point of interest. We observe that varying C_D between zero and one (dark curves vs their light counterparts) produces the same timing effect seen in Fig. 9, but at $\theta=45°$ also affects velocity magnitude after the first bubble minimum. Neglecting bubble translation, either partially (dashed curves vs their solid counterparts) or totally (dotted curves vs their solid counterparts), produces large error when $C_D=0$, but smaller error when $C_D=1$. A comparison of the solid curves in Fig. 10(a) with their counterparts in Fig. 10(b) shows that dipole radiation plays even a larger role in the velocity field than it does in the pressure field.

Finally, Fig. 11 displays again the $C_D=0$ and $C_D=1$ radial velocity histories from (59) seen in Fig. 10, but adds the corresponding meridional velocity histories u_θ' from (60). We see that u_r' consistently overshadows u_θ', which would be expected in the moving coordinate frame. However, in the fixed spherical coordinate system defined by $\mathbf{e}_r,\mathbf{e}_\theta=[\mathbf{e}_r',\mathbf{e}_\theta']_{t=0}$, this is not the case.

V. CONCLUSION

Equations (45), (48), (59), and (60) constitute closed-form expressions for the pressure and velocity fields gener-

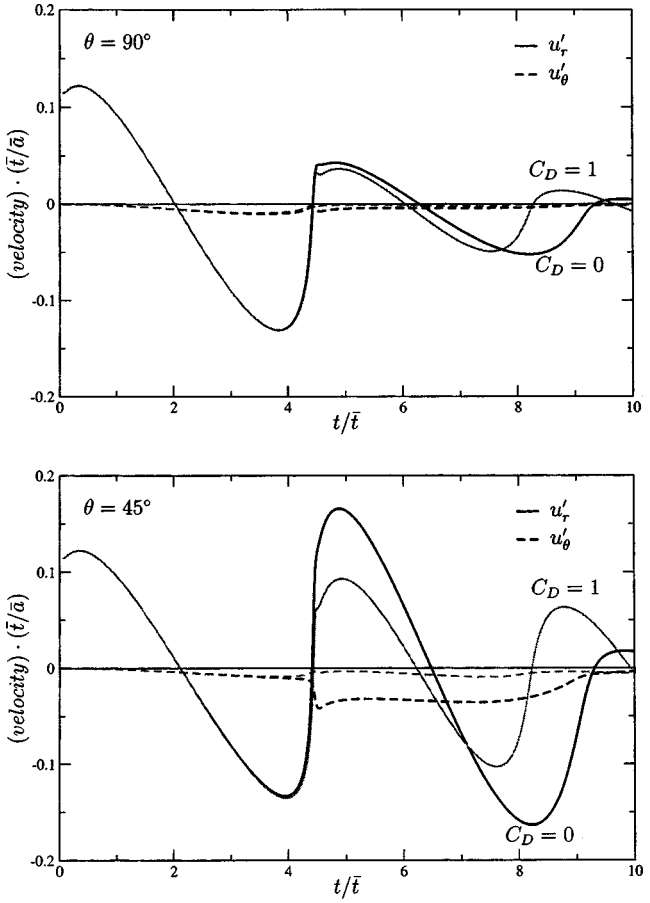


FIG. 11. Radial- and meridional-velocity histories, in the moving coordinate system, produced by the bubble-displacement histories of Fig. 8 with (59) and (60), $r=R=5\bar{a}$.

ated by an underwater explosion. These expressions are valid in the far field ($r^2>10\bar{a}^2$) when free-surface effects are negligible ($\bar{d}>10\bar{a},\ z<\bar{d}-10\bar{a}$), and incorporate the effects of bubble dilation and translation. As seen in Fig. 4 for the G&H bubble, computational results exhibited good agreement with experimental data for deep bubbles; experimental data are lacking for bubbles at intermediate depths.

Figure 4 demonstrates that the G&H bubble model is clearly superior to the Lamb and K&K models as inputs for the generation of pressure and velocity fields. Figure 6 shows that field generation based on an incompressible-flow model produces accurate results only in a thin annular region ($R-\Delta\leqslant r\leqslant R+\Delta$) with the charge as its center. Computational results for large bubbles at intermediate depth (Fig. 8) exhibit the limitations of spherical bubble models when translation is significant. They also show (Figs. 9 and 10) that dipole radiation can contribute substantially to the pressure and velocity fields. Finally, it is seen in Fig. 1 that the refined similitude relation given in Sec. II A constitutes a substantial improvement over previous relations. The models described herein have been implemented in the CFA2 code, which is distributed by Anteon Corporation (khunter@anteon.com).

ACKNOWLEDGMENTS

This work was supported under Contract No. N0014-01-1-0154 with the Office of Naval Research. The authors express their appreciation to Mr. Gregory Harris of the Naval Surface Warfare Center, Indian Head Division, and to Dr. Georges Chahine of Dynaflow, Inc., for their helpful comments. The authors also acknowledge the assistance of Mr. Kent Rye of the Naval Surface Weapons Center, Carderock Division, and of Professor Ross Chapman of the School of Earth and Ocean Sciences, University of Victoria, in seeking far-field experimental data.

Arons, A. B., Slifko, J. P., and Carter, A. (**1948**). "Secondary pressure pulses due to gas globe oscillation in underwater explosions I. Experimental data," J. Acoust. Soc. Am. **20**(3), 271–276.

Arons, A. B., and Yennie, D. R. (**1948**). "Energy partition in underwater explosion phenomena," Rev. Mod. Phys. **20**, 519–536.

Brinkley, S. R. J., and Kirkwood, J. G. (**1947**). "Theory of the propagation of shock waves," Phys. Rev. **71**(9), 606–611.

Cole, R. H. (**1948**). *Underwater Explosions* (Princeton University Press, Princeton, NJ).

Coles, J. S., Christian, E. A., Slifko, J. P., Niffennegger, C. R., and Rogers, M. A. (**1946**). "Shock-wave parameters from spherical TNT charges detonated underwater," in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 1, pp. 1085–1105.

Farley, T. E., and Snay, H. G. (**1978**). In Explosion Effects and Properties: Part II-Explosion Effects in Water, edited by M. M. Swisdak NSWC/WOL TR 76-116, page 17. Naval Surface Warfare Center. Unclassified data from classified source.

Fujikawa, S., and Akamatsu, T. (**1980**). "Effects of the nonequilibrium condensation of vapour on the pressure wave produced by the collapse of a bubble in a liquid," J. Fluid Mech. **97**, 481–512.

Geers, T. L., and Hunter, K. S. (**2002**). "A wave-effects model for an underwater explosion bubble," J. Acoust. Soc. Am. **111**(4), 1584–1601.

Herring, C. (**1941**). "Theory of the pulsations of the gas bubble produced by an underwater explosion," in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 2, pp. 35–131.

Hicks, A. N. (**1970**). "Effect of bubble migration on explosion-induced whipping in ships," Technical Report 3301, Naval Ship Research and Development Center, Bethesda, MD.

Keller, J. B., and Kolodner, I. I. (**1956**). "Damping of underwater explosion bubble oscillations," J. Appl. Phys. **27**, 1152–1161.

Kirkwood, J. G., and Bethe, H. (**1942**). "The pressure wave produced by an underwater explosion I. (OSRD No. 588)," in *Shock and Detonation Waves*, edited by W. W. Wood (Gordon and Breach, New York, 1967).

Lamb, H. (**1923**). "The early stages of a submarine explosion," Philos. Mag. **45**(Ser. 6), 257–265.

Landau, L., and Lifshitz, E. (**1951**). *The Classical Theory of Fields* (Addison-Wesley, Reading MA).

Leppington, F. G., and Levine, H. (**1987**). "The sound field of a pulsating sphere in unsteady rectilinear motion," Proc. R. Soc. London, Ser. A **412**, 199–221.

Lezzi, A., and Prosperetti, A. (**1987**). "Bubble dynamics in a compressible liquid. II. Second-order theory," J. Fluid Mech. **185**, 289–321.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).

Penney, W. G. (**1940**). "The pressure–time curve for underwater explosions," in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 1, pp. 273–287.

Penny, W. G., and Dasgupta, H. K. (**1942**). "Pressure–time curves for underwater explosions" (second paper), in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 1, pp. 289–299.

Price, R. S. (**1979**). "Similitude equations for explosives fired underwater," Technical Report NSWC TR 80-299, Naval Surface Warfare Center.

Prosperetti, A., and Lezzi, A. (**1986**). "Bubble dynamics in a compressible liquid. I. First-order theory," J. Fluid Mech. **168**, 457–478.

Rogers, P. H. (**1977**). "Weak-shock solution for underwater explosive shock waves," J. Acoust. Soc. Am. **62**(6), 1412–1419.

Shiffman, M., and Friedman, B. (**1943**). "Studies on the gas bubble resulting from underwater explosions—on the best location of a mine near the sea bed," in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 2, pp. 245–320.

Snay, H. G. (**1962**). "Underwater explosion phenomena: The parameters of migrating bubbles," Technical Report NAVORD 4185, U.S. Naval Ordnance Laboratory, White Oak, MD.

Snay, H. G., and Tipton, R. V. (**1962**). ''Charts for the parameters of migrating explosion bubbles,'' Technical Report NOLTR 62-184, U.S. Naval Ordnance Laboratory, White Oak, MD.

Swift, E., and Decius, J. C. (**1948**). ''Measurements of bubble pulse phenomena. III. Radius and period studies,'' in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 2, pp. 553–599.

Taylor, G. I. (**1942**). ''Vertical motion of a spherical bubble and the pressure surrounding it,'' in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 2, pp. 131–144.

Taylor, G. I., and Davies, R. M. (**1942**). ''The motion and shape of the hollow produced by an explosion in a liquid,'' in *Underwater Explosion Research* (Office of Naval Research, Washington, D.C., 1950), Vol. 2, pp. 227–243.

Tomita, Y., and Shima, A. (**1977**). ''On the behavior of a spherical bubble and the impulse pressure in a viscous compressible liquid,'' Bull. JSME **20**(149), 1453–1460.

Van Dyke, M. (**1964**). *Perturbation Methods in Fluid Mechanics* (Academic New York).

Acoustic identification of a single transmission at 3115 km from a bottom-mounted source at Kauai

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

(Received 4 October 2002; revised 19 December 2003; accepted 5 January 2004)

Sounds received in the Gulf of Alaska at 3115 km from the ATOC/NPAL source at Kauai (75 Hz, 0.027-s resolution, bottom-mounted) are compared with acoustic and oceanographic models. Unlike data collected at stationary SOSUS arrays, these data come from a towed horizontal array at 372-m depth of military origin. A plausible identification of the acoustic reception is made despite the fact that only one transmission is collected and sound interacts with the bottom near the source. The similarity between the modeled and measured impulse response here may be useful for understanding the signals between this same source and the NPAL array near southern California. The plausible identification of sound from the horizontal array here appears to point toward the feasibility of using other military platforms of opportunity besides SOSUS to study acoustic propagation and possibly map climatic changes in temperature by means of tomography. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1650014]

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

Pages: 1497–1504

I. INTRODUCTION

Synthetic aperture receivers may yield sufficient resolution to tomographically image climatic temperature variations in the ocean related to El Niño and the Southern Oscillation using signals from a few sources.[1] While sparse sampling may reveal interesting features of climatic change, there may not be enough sound surveillance systems (SOSUS) or other stationary receivers to map these features using tomography without interpolating between sections using a dynamical model of the ocean's circulation.[2] This paper indicates that a towed horizontal array can be used to process and identify acoustic paths over ocean basin scales from a single transmission from the ATOC[3] source at Kauai. Since the received data come from a military origin, it appears that similar arrays could augment stationary receivers to provide a synthetic aperture for mapping climatic variations as the ships move from day to day.

Historically, it was suggested that the mesoscale could be mapped with stationary sources and receivers.[4] A moving source was used to map the mesoscale in a 300 by 300-km^2 area.[5] A moving ship was used to map the mesoscale within a 1000-km-diam circle in the Atlantic.[6] These investigations used instruments with accurate time keeping and navigation.

Mapping the mesoscale seems possible in a mesoscale sized box using inaccurately navigated receivers.[7] Models indicate that synthetic aperture receivers can have location errors of O(1) km and still accurately map climatic variations of temperature such as those due to El Niño and the Southern Oscillation.[1,8] In particular, the simulations suggest that tomographic images of Rossby waves of order O(500) km and other large variations are well resolved using moving receivers with very poor navigation. The effects from a semi-realistic mesoscale are almost negligible in these simulations because of the dominant effects from large scales of temperature on acoustic travel time. These results suggest the possibility of using mobile receivers for studying climate change by means of tomography. A needed demonstration involves processing sounds and identifying acoustic paths that propagate over basin-scales to mobile receivers.

To do tomography between a fixed source on the bottom and mobile receivers, it appears to be useful to investigate two things. First, it is useful to see if acoustic paths can be identified with imperfect information concerning the bottom depth and subbottom properties near the source. It is possible that the difficulty in understanding the acoustic signal between the Kauai source and an array near southern California is due to misunderstood interactions between sound and the bottom near the source *and* the receiver.[3] On the other hand, sounds as received on the towed array discussed in this paper do not interact with the bottom near the receiver, so this geometry is simpler to analyze. A plausible identification is made by assuming that ray paths reflect specularly from the bottom near the source.

Second, it appears necessary to find out if path identification can be made from a single transmission from the source, instead of the traditional method of using numerous transmissions interspersed over a day or longer. With many transmissions, it is possible to average many records to establish stable arrivals that may otherwise fade for a minute or so due to scattering. The fading of acoustic paths is analogous to the twinkling of stars at night due to atmospheric turbulence. The results of this paper support the view that a single transmission can be enough to understand acoustic pulses at the receiver. This appears to be the first time that a model is used to identify signals from a single transmission at basin-scales.

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

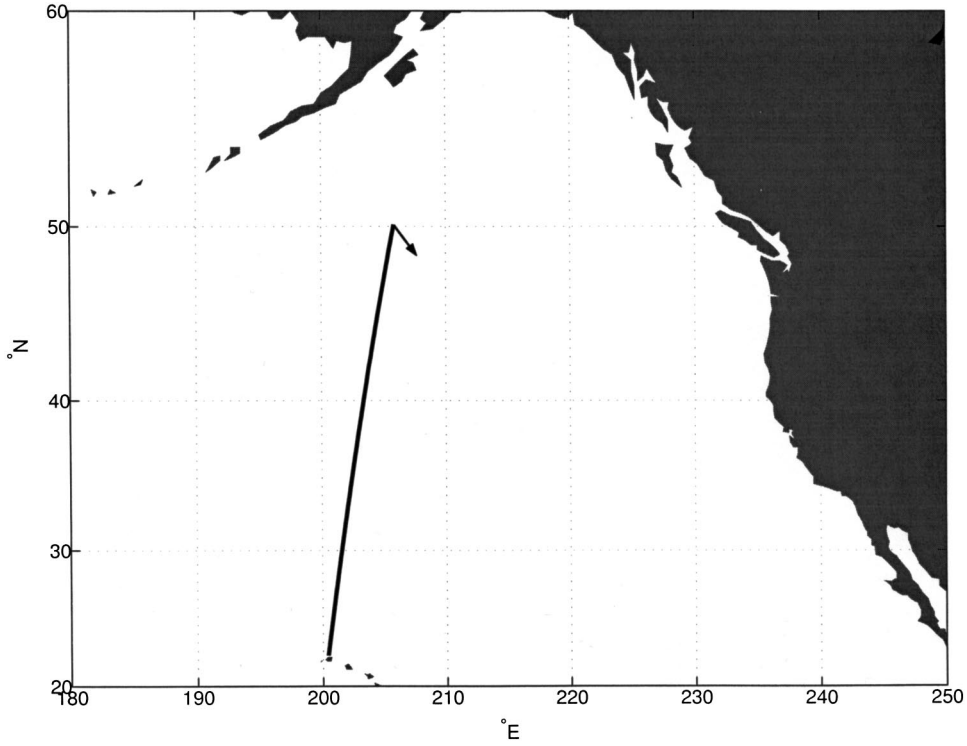


FIG. 1. Plan view of the experiment. Data come from a source at Kauai and are collected on a towed array of U.S. military origin in the Gulf of Alaska. The heading of the vessel is to the southeast and makes an angle of about 53° with the geodesic (arrow).

II. EXPERIMENT

The source[9] is mounted on the bottom at 816-m depth and 22.349°N, 200.43033°E near Kauai. Power and accurate timing are provided through a cable running to the shore. A broadband signal is transmitted at a center frequency of 75 Hz. The phase of the carrier is modulated every two cycles using a linear maximal shift register sequence having 1023 digits. The first zeros in the emitted spectrum are at 75/2 $=37.5$ Hz on either side of 75 Hz. The signal periodicity and level are $(1023)2/75=27.280$ s and 195 dB *re* 1 μPa @ 1 m, respectively. Of the 20 min or 44 transmitted periods beginning at 18:00 on 7 November 1997, 14 min, or 32 periods, are analyzed here.

The receiver is a horizontal array of U.S. military origin towed at 372-m depth. The forward modeling in this paper sets its location at 50.0967°N, 205.8467°E, which is written with much greater precision than its accuracy of a few kilometers. The precision is given so others might model this section using the same coordinates used here.

The geodesic length between the source and receiver is 3115.45 km (Fig. 1). During the arrival of the signal, the array speed is 2.5 m/s and its heading is 137°T. Since the bearing angle to the source is about 190°T, the incoming signal makes an angle of about 53° with respect to the heading of the ship. During the 14 min of data reception, the ship travels 2.2 km and is 1.3 km closer to the source at the end of the 14 min. The data are time stamped with an accuracy of 5 to 10 s. In this paper, it is not possible to compare predictions of acoustic travel time with the data because of the uncertainties of position and timing. However, the sample rate of the data is stable and accurate.

III. DATA

Using standard techniques the data are first beamformed toward the source. Next, each 27.280-s M-sequence period is adjusted for various possible Doppler corrections and correlated with a replica of the transmitted sequence. Replica correlation with a linear maximal shift register sequence compresses 27.280 s of energy from each acoustic path into a pulse of duration $2/75=0.026\,67$ s with a theoretical gain of $10\log_{10}(1023)=30$ dB. The Doppler correction yielding the largest signal-to-noise ratio is chosen for each period. This procedure yields sufficient signal-to-noise ratio to examine the impulse response.

A few words can be said about the coherent integration time of this signal, though it has no bearing on other results in this paper. If one chooses the average Doppler correction among the 32 processed periods, the signal can be coherently integrated for 80 s. It is not known if one could integrate for a longer period than this if one utilized a different Doppler correction for each period. This type of processing would explore the degree to which the ship's change in velocity affects coherent integration time versus effects due to fluctuations in the ocean.

A ''bit plot'' is shown for the output of each of 32 processed periods of the received signal (top, Fig. 2). Some features persist for 14 min and other do not.

Incoherent averages are formed using

$$a(m)=\left[\frac{1}{32}\sum_{r=1}^{32}\frac{\|d(m,r)\|^2}{\sigma^2(r)}\right]^{1/2},\quad m=1,2,\ldots,M, \tag{1}$$

where the mth complex demodulate of the rth record is $d(m,r)$. The variance of the noise for record r is $\sigma^2(r)$. It is included to give proper weight to records based on their signal-to-noise ratio. $\sigma^2(r)$ is estimated from each record where signal is not present. The energy arrives around 2097 s and ends around 2105 s, a duration of 8 s.

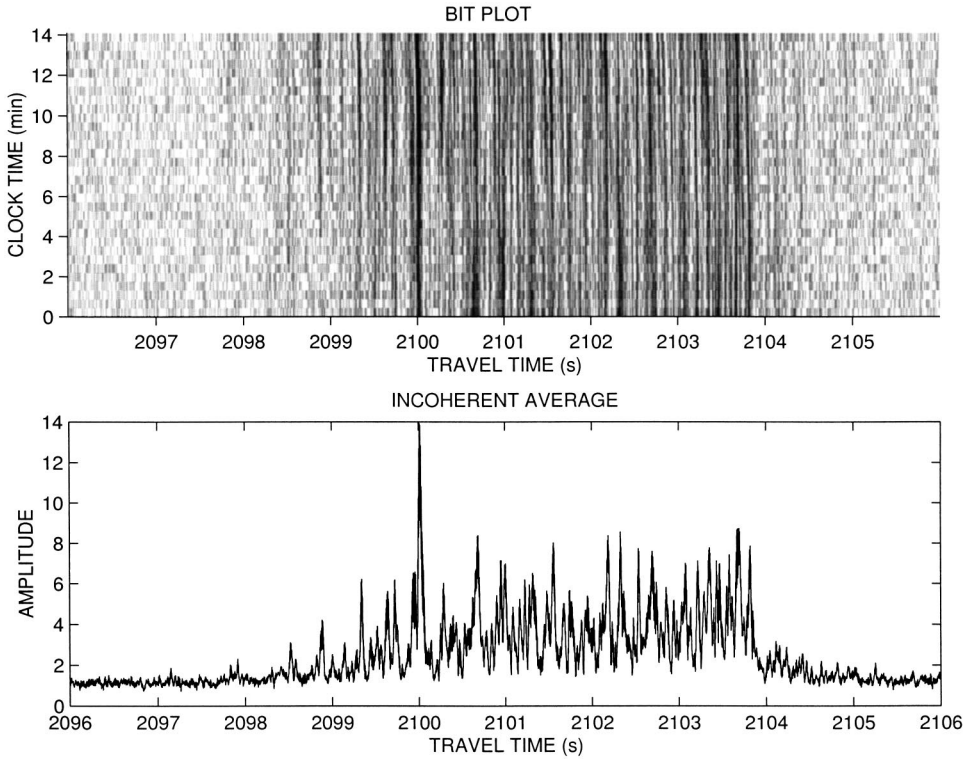


FIG. 2. The impulse response from the Kauai source for the section in Fig. 1 on 7 November 1997. The bit plot shows signal-to-noise ratio from −3 dB and less as white to 20 dB and more as black. Each of the 32 separate periods of received signal, covering a total of 14 min, is separately processed to correct for the motion of the towed array (Doppler) and to remove the shift register sequence code using replica correlation. The bottom panel is the incoherent average [Eq. (1)] of these 32 processed periods. The axis for travel time is inaccurate.

IV. MODELS FOR ACOUSTIC PROPAGATION

Different realizations of the internal wave field are modeled to ascertain their effects on acoustic variability. With previous data, this step has not been needed because many data sets could be averaged to determine the features of the data that were stable and could be expected to be identified from a model. With a single transmission, it is important to model the effects from a time-evolving internal wave field to estimate the portions of the model that ought to be stable and identifiable.

A. Environment

The speed of sound is computed using Del Grosso's algorithm[10] and Levitus' climatological averages[11] of temperature and salinity for Fall. The depth of minimum speed varies from 740 m at the source to 105 m at the receiver. Since the acoustic models use Cartesian coordinates, the sound speed profiles are translated to Cartesian coordinates using the Earth-flattening transformation.

Internal waves are modeled with the Garrett–Munk[12] spectrum. Currents are ignored, being two orders of magnitude less than sound speed perturbations arising from adiabatic vertical displacements of water in the upper ocean. The perturbations are added to the climatology of sound speed described above. Internal wave modes are precomputed and retrieved as needed at range intervals of 80 km to account for changes in water depth, buoyancy frequency, and sound speed. Vertical displacements of these modes are set to zero at the surface and bottom. For each 80-km interval, a three-dimensional field of internal waves is computed in a box of 80 km by 80 km by D m where D is the average depth of the ocean in that interval. A vertical slice through the box gives the vertical displacements along the geodesic. The energy of the internal wave field is taken to be that specified by Garrett and Munk[12] because that energy level has matched observations with this construction of internal waves before.[13] The literature is not unanimous in its adoption of the best energy level to use for acoustic modeling.[14–16] Further information on the construction of internal waves is found elsewhere.[13]

The depth of the bottom near 100 km of the source is taken from a SEABEAM[17] survey (Bruce Howe, personal communication). Beyond this range, depths are taken from a digital database[18] (Fig. 3).

For lack of a definitively better set of parameters for the subbottom near the source, parameters like those used for the Kaneohe source at Oahu[13] are used. The thickness of the sediment is 200 m. The sound speed at the top of the sediment divided by that at the bottom of the water column is 1.02. The density of the sediment is 1.7 gm cm^{-3}. The attenuation in the sediment is

$$\alpha(f) = \alpha_0 f^p (\text{dB m}^{-1}), \tag{2}$$

where f is the frequency in kHz, $p=1$, and $\alpha_0 = 0.02$ dB m^{-1} kHz^{-1}. The speed in the sediment is taken to increase with depth as 1 s^{-1}. The speed in the basement divided by that at the bottom of the sediment layer is 2. The density of the basement layer is 2.5 gm cm^{-3}. The attenuation in the basement is given by Eq. (2) except $\alpha_0 = 0.5$ dB m^{-1} kHz^{-1} and $p=0.1$. Geoacoustic parameters are needed within 500 km of the source because that is where the modeled sound field interacts with the bottom. Most of the interaction occurs in the first 50 km. Afterward, interactions occur with a few deep seamounts.

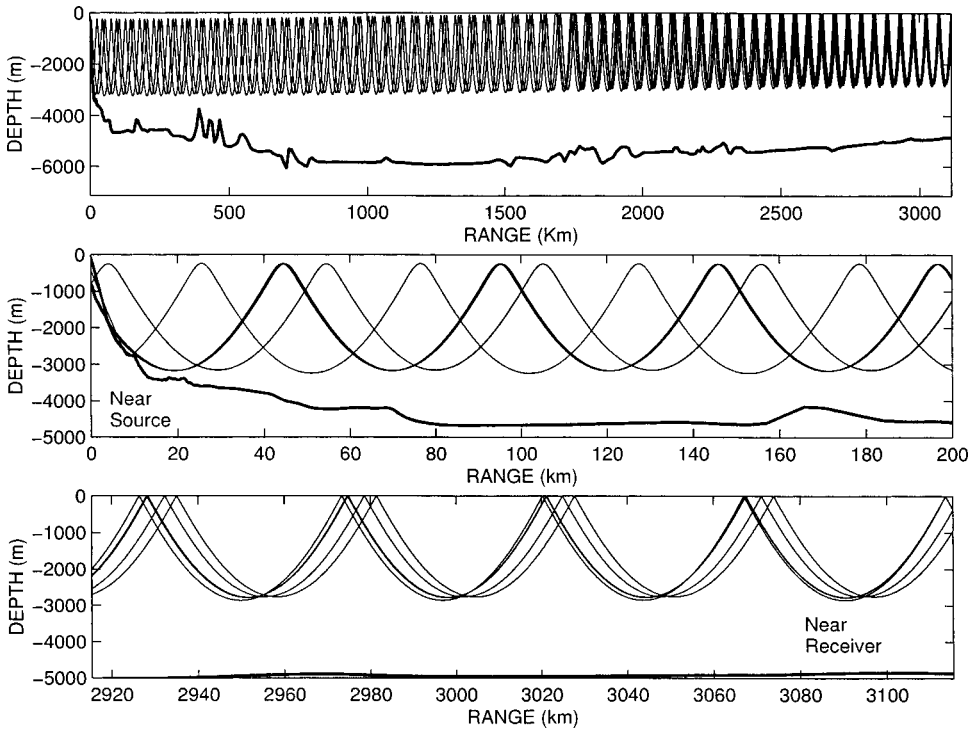


FIG. 3. The depth of the bottom is indicated along the section (Fig. 1) with details near the source and receiver. Rays in the top 16 dB making up arrival B1 in Fig. 7 are indicated.

B. Parabolic approximation

The sound speed insensitive parabolic approximation[19] is used to compute a two-dimensional field of sound along the geodesic from 0- to 8000-m depth. Tests[19] suggest that travel times of pulses are computed with an accuracy of a few milliseconds. The field is modeled at each of 2048 acoustic frequencies. The impulse response is synthesized with an inverse Fourier transform resulting in a time series with a period of 27.3 s. In the first 48 km where sound interacts with the steep slope of Kauai, the computational grid has an interval of 0.0333 km in range and 3.9 m in depth. At greater ranges, the grid interval is 0.1333 km in range and 7.8 m in depth. These values are sufficient to obtain convergence within a few decibels at the receiver.

C. Rays

Fans of rays are traced using a program, zray, that is a modification of ray.[20] Eigenrays are found using another program. These programs have been used to identify acoustic paths before.[21] Rays reflect specularly from the bottom. Both geometric and nongeometric arrivals are found. Geometric types are those that pass through the source and receiver. Nongeometric types are those that provide energy at the receiver on the shadow sides of caustics. For lack of a more reliable value, rays that reflect from the bottom suffer an attenuation of 3 dB per bounce. The sound speed field used for the ray trace is identical to that used for the sound speed insensitive parabolic model at its computational grid.

V. IDENTIFYING ACOUSTIC PATHS

A. Stable arrivals from models

Stable arrivals are those that can be tracked from day to day. This cannot be investigated from one data record. Instead, models are used to predict stability.

Eight impulse responses are computed from the sound speed insensitive parabolic approximation.[19] The sound speed field for each comes from Levitus' climatology and a field of internal waves. Internal waves are generated at intervals of a day using the linear dispersion relation. Two incoherent averages are computed from these eight impulse responses (Fig. 4). Seven stable arrivals are labeled (A–G). Arrivals after 2105 s do not appear to be stable. Arrival C does not look very stable but this is probably because there are not enough model realizations to show its stability. Stable arrivals before A are unlabeled because they have very low signal-to-noise ratios in the data. Arrivals A–D appear to consist of two resolved arrivals, which can be observed in modeled time fronts (Fig. 5). Evidently, at the receiver depth of 372 m, these can be resolved in the presence of the modeled field of internal waves. Temporal separations of later doublets cannot be resolved. Without internal waves, more stable arrivals are predicted than when internal waves are added (Fig. 5). Figure 4 appears to predict stability better than the time fronts, especially between 2104 and 2105 s where it is difficult to judge stability from time fronts.

B. Ray approximation for tomographic inversion

The impulse response from rays looks similar to that from the parabolic approximation (Fig. 6). To use rays for tomographic inversions, it is necessary to show that, for each stable arrival, the corresponding ray paths are themselves sampling a similar region of the ocean and that those ray paths have travel times that are sufficiently accurate.

The first issue can be dealt with by finding the eigenrays for different representations of ocean fields, and examining the paths for each realization for each stable arrival. Our computer resources are insufficient to compute eigenrays through the internal wave field. Investigators have found that at long distances, stable arrivals are composed of many rays.

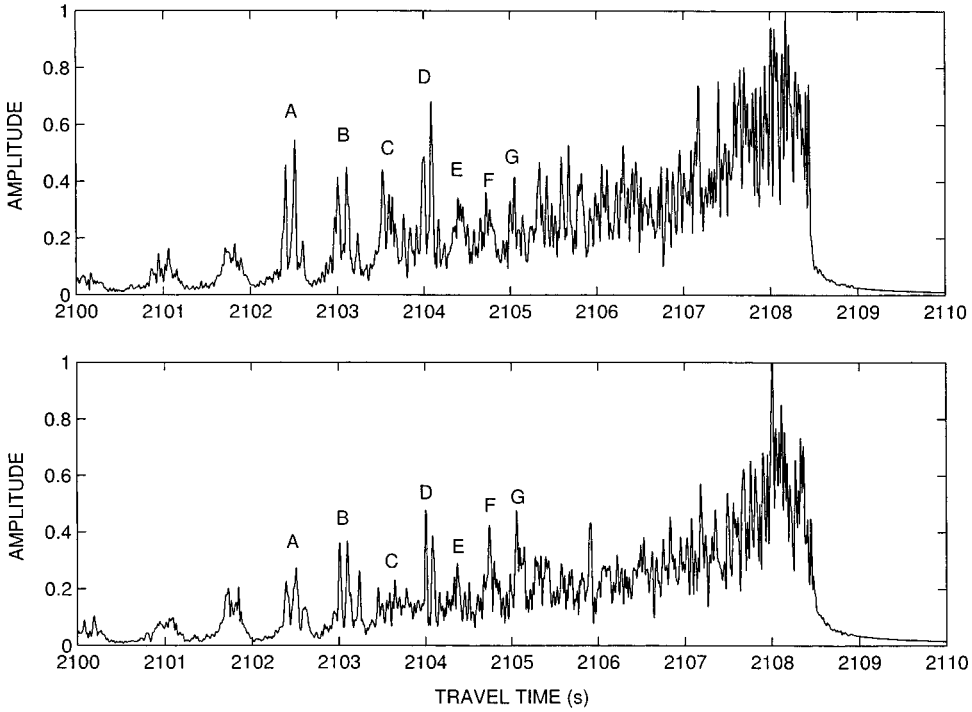


FIG. 4. Investigation of predicted path stability. Top: Incoherent average of four modeled impulse responses at one-day intervals where the modeled time variation is given by the linear dispersion relation and the Garrett–Munk[12] spectrum of internal waves. Bottom: Same except these are from four other impulse responses at one-day intervals. The acoustic receptions are modeled with the sound speed insensitive parabolic approximation.[19] The letters indicate arrivals which appear to be stable from one panel to the next.

Some of the arrivals are suitable for tomographic inversion because their constituent ray paths have similar upper turning depths even though their turning ranges are quite different.[16,21,22] In this experiment, some stable arrivals are composed of several rays (A1, A2, B, C2, D1, E2, F1, F2, G1, G2) even though the mesoscale and internal wave scales are not incorporated into the sound speed field (Fig. 3). The notation ''A1'' or ''A2'' denotes the first or second arrival of a doublet, respectively. Upper turning depths for arrivals A–G are 237 to 358 m, respectively, near the source and 61 to 1 m, respectively, near the receiver. Most stable arrivals appear to be suitable for tomographic inversions.

The second issue is whether the travel times from rays are close enough to a full wave solution of the wave equation to warrant their use without significant modification. In this paper, the full wave solution is given by the parabolic approximation. Travel times of rays and stable arrivals from the parabolic model can differ because of diffraction[23] and due to the fact that the parabolic model includes propagation into the subbottom. Rays are only allowed to reflect specularly from the bottom. Despite these differences, the differences in modeled travel times are less than 0.03 s, and are typically

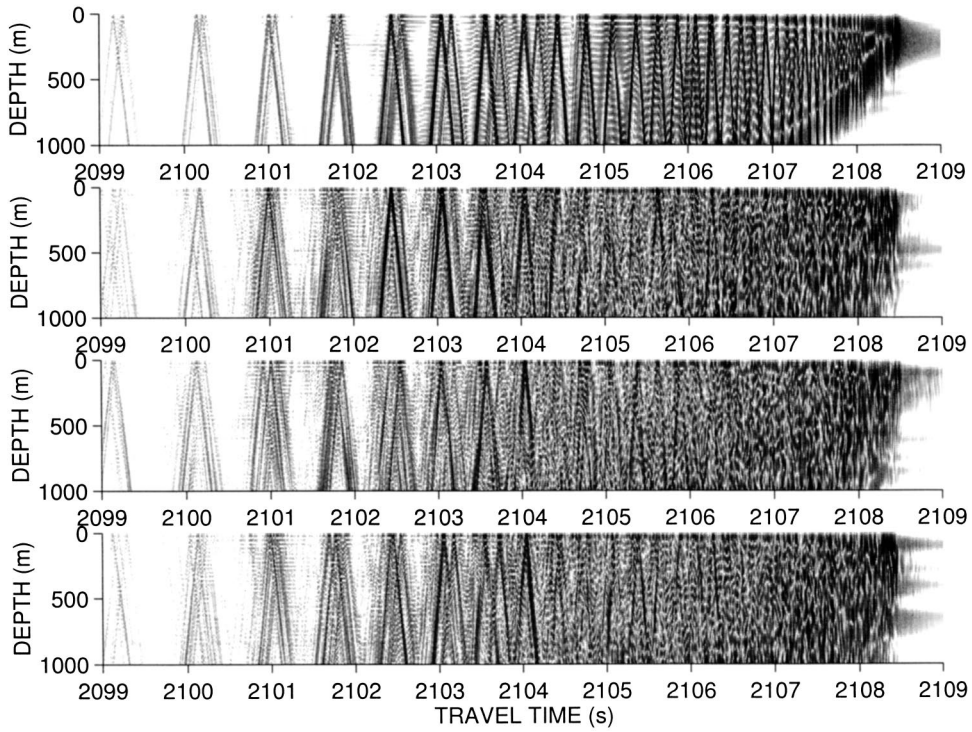


FIG. 5. Modeled time fronts in the top 1000 m at the receiver without internal waves (top) and with internal waves at daily intervals (bottom 3 rows). The receiver is at 372-m depth. The models are generated with the sound speed insensitive parabolic approximation[19] and a Garrett–Munk[12] spectrum of internal waves.

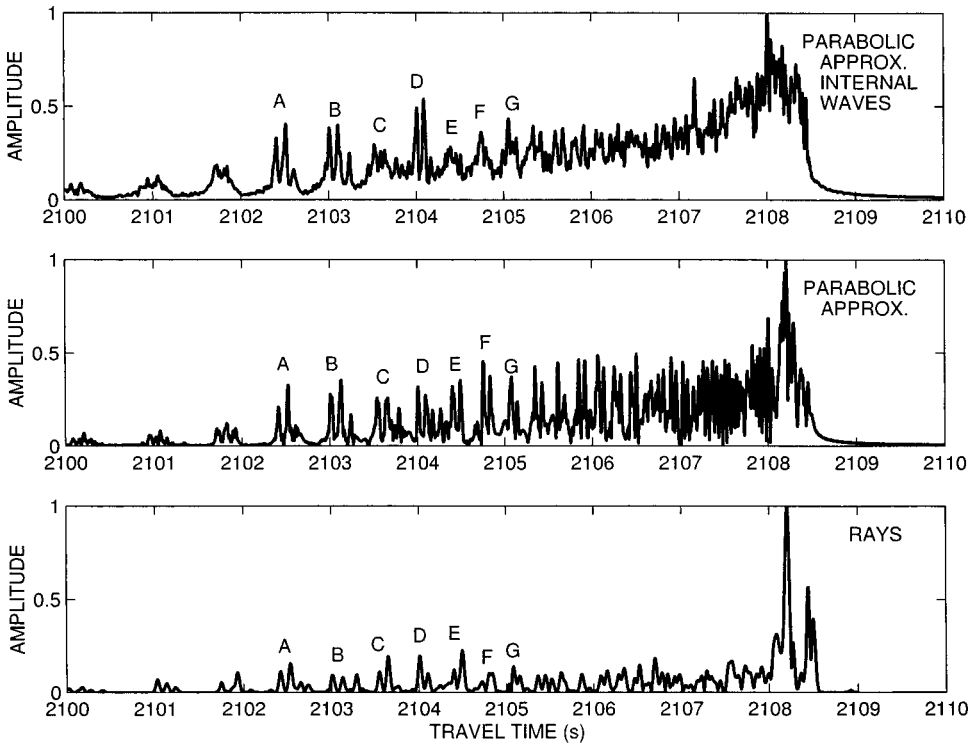


FIG. 6. Three models for the propagation of sound. Top: Same as top panel, Fig. 4 except this uses all eight internal wave fields to construct the incoherent average. Middle: Sound speed insensitive parabolic approximation[19] where the sound speed field is from Levitus' climatology[11] for Fall. Bottom: Eigenrays where the sound speed field is identical to that in the middle panel. The seven arrivals that appear to be stable in the models are labeled A–G.

0.005 s (Table I). These errors are small compared to the climatic signal of about 1 s expected from Rossby waves linked to El Niño and the Southern Oscillation.[2]

C. Identifying arrivals in the data

The most accurate model for identifying paths is the incoherent average of the output from the parabolic equation for eight internal wave fields (Fig. 7). Using a visual alignment, the association with the data is plausible. The seven stable arrivals, A–G, from the model appear in the data. The cutoff times from the data and model are similar, but the data have an extra second of energy at lower levels at the end. This second of energy may come from a positive bias in the travel time of sound that is trapped near the depth of minimum speed in the waveguide. Mesoscale eddies appear to have caused biases of this order in a different transmission at basin-scales.[13]

TABLE I. The difference in travel time between peaks from the parabolic approximation (pe) and ray models from the bottom two panels of Fig. 6. Models are computed for the same sound speed field. Peaks are labeled A-G and the "1" and "2" denote the first and second arrivals of the doublets.

Peak	$T_{pe}-T_{ray}$ (s)
A1	−0.006
A2	−0.009
B1	−0.005
B2	0.
C1	−0.010
C2	−0.014
D1	−0.005
D2	−0.011
E1	−0.003
E2	−0.006
F1	+0.029
F2	−0.008
G1	−0.004
G2	−0.009

The lack of an accurate time base probably does not significantly affect the confidence with which the data can be identified with these models because other oceanographic variability associated with ENSO can modify travel times by O(1) s (Ref. 2). With a 1-s error from any model, there is a possibility that the alignment in Fig. 7 is incorrect. However, it appears to be problematic to shift model times earlier or later here because both the data and model exhibit a dramatic drop in level at about 2108.5 s with the alignment in Fig. 7. It is plausible that the "extra energy" that arrives late in Fig. 7 is due to a bias caused by the mesoscale. Thus, the alignment in Fig. 7 is the best one, with other alignments appearing significantly worse. It seems that all acoustic models have difficulty identifying acoustic arrivals from any basin-scale section in the North Pacific without any doubt whatsoever in the face of the inherent O(1) s uncertainty in travel time due to unknown variability at climatic scales.

VI. CONCLUSIONS

Acoustic signals at 3115 km from a source near Kauai can be coherently processed to yield large signal-to-noise ratios on a towed array in the Gulf of Alaska (Fig. 1). The acoustic paths that are predicted to be stable by means of acoustic and oceanographic models can be identified in the data. There are discrepancies between the models and data, particularly during the last second when weak arrivals are observed that are not present in the model. The models use climatological and internal wave variations, but do not include a mesoscale. A previous study[13] indicates that the mesoscale is responsible for a 0.6-s addition of energy at the end of the reception due to a positive bias in the travel times

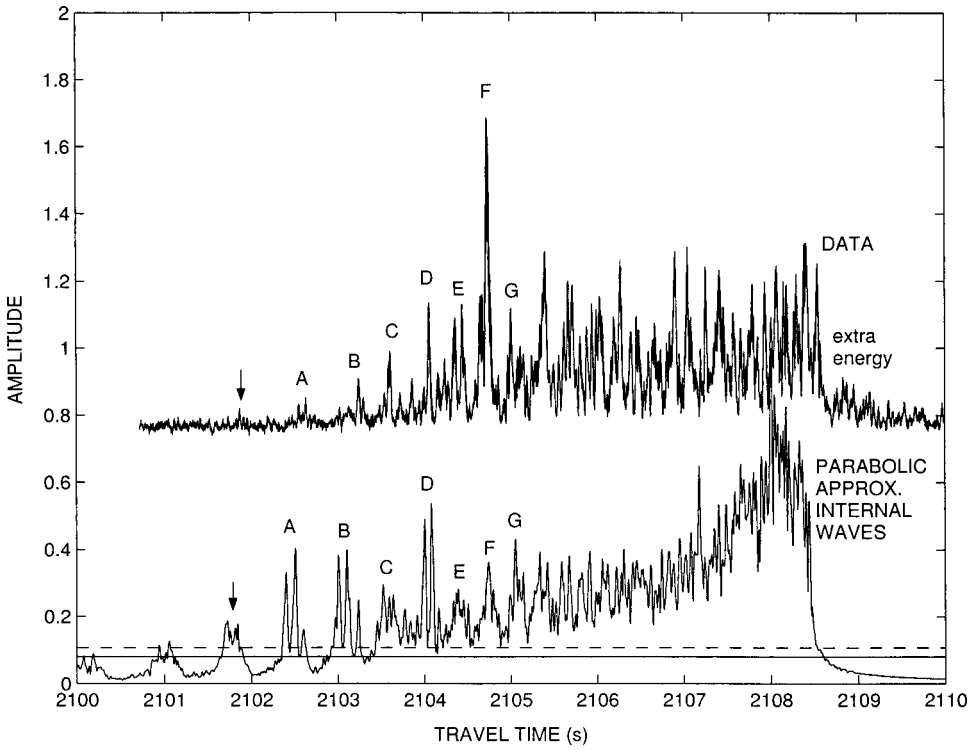


FIG. 7. Incoherent average of the data (Fig. 2) compared with the incoherent average from the parabolic approximation model (top, Fig. 6). Amplitudes of the data are offset so they do not overlap the model. Modeled time variability is generated using eight realizations of the Garrett–Munk[12] spectrum of internal waves at daily intervals. Stable arrivals from the model in Fig. 4 are shown with their associated counterparts from the data. The arrow identifies a weak arrival in the data that corresponds to a modeled arrival. The peak signal-to-noise ratio in the data is 21 dB. The solid line through the model shows this same ratio. The dashed line shows the upper limit for noise variations given the statistical fluctuations of noise in the data. The weak ''extra energy'' in the data is not observed in the model. About 4 s are added to the data to align with the model.

of sounds trapped near the axis of the acoustic waveguide. Unlike internal waves which have a universal spectrum,[12] the mesoscale is not so simply predicted. It is possible that some ocean circulation models have a realistic mesoscale that can be used to predict this bias.

Mobile receivers allow studies of acoustic propagation to be conducted at relatively little cost for many different geographic regions. Although SOSUS stations and the single section studied here yield data that appear to be interpretable with acoustic and oceanographic models, the outcome may be more complicated in different areas of the ocean. Being able to process and identify acoustic paths from the towed array here seems to keep open the possibility of using mobile systems for studying climatic temperature changes in the ocean using tomography and synthetic apertures. Ray models for some identified features from the data here indicate that tomographic inversions are feasible (Fig. 3).

At first thought, inaccurate timing at a receiver may seem to preclude tomography in the ocean. The literature contains a paper showing that models of climatic variability can be accurately mapped from a few sources and 20 mobile receivers even when the time base in the sources is off by hours.[24] The same principles used in that paper could be applied to ask if climatic variability could be mapped if the time bases on mobile receivers were off by hours, but their time bases were stable over months at a time. It is too early to conclude that inaccurate time bases at mobile receivers either could or could not produce accurate maps of climatic variability. On the other hand, accurate timing could be provided for towed arrays if desired.

A significant effort is still required to interpret signals transmitted over basin-scales despite the contemplation of this effort 83 years ago.[25] The process cannot be done in real-time at this point in history. Once a seemingly correct choice of acoustic and oceanographic models is used to identify acoustic paths at a receiver, the computer times required to do the modeling are counted in weeks rather than in the 30 min or so required for the signal to propagate from the source to the receiver. Finding a suitable set of acoustic and oceanographic models usually takes much longer than a few weeks. Perhaps during the next decades, enough experience with other data sets and models will make it possible to accurately predict acoustic signals in many regions of the ocean. The speeds of computation necessary to make predictions will probably not be a limiting factor when predictions can be made reliably and routinely.

The modeled signal penetrates the bottom near the Kauai source. Even though the geoacoustic parameters for this region are currently not well known from the literature, a plausible identification is made at basin-scales, even when the bottom properties are ignored and only specular reflection from the bottom is allowed (ray model, bottom, Fig. 6, Table I). This identical finding[13,21] is obtained from the Kaneohe source near Oahu (133 Hz, 16 Hz bandwidth, 3709 km section). Is it luck that the subbottom need not be well modeled for these sections, or are other acoustic sections more sensitive to models for propagation in the solid Earth? In any case, the data from the Gulf of Alaska can be used to test the way bottom interactions are modeled near the Kauai source.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research (ONR) Contract No. N00014-00-C-0317. John Spiesberger thanks the following individuals and institutions. The data were provided by Jim Phalen of the Applied Physics Laboratory at Johns Hopkins University and Ed McWethy at the ONR. Arthur Bisson at the ONR facilitated access to

these data. Bruce Howe and others at the University of Washington provided information regarding the location and time of the emitted signal. They also deployed the source at Kauai. Bruce Einfalt (Penn State University), Carter Ackerman (Penn State University), and Kurt Metzger (University of Michigan) provided some computer programs used to process the signal for Doppler corrections and replica correlation. This paper is dedicated to Fred Tappert for his scientific contributions and the interesting discussions and collaborations we had over many years, and to Anisim Silivra whom the author worked with on the problem of timing errors in tomography. The author thanks the reviewers for their suggestions.

[1] A. Fabrikant, J. L. Spiesberger, A. Silivra, and H. E. Hurlburt, ''Estimating Climatic Temperature Change in the Ocean with Synthetic Acoustic Apertures,'' IEEE J. Ocean. Eng. **23**(1), 20–25 (1998).

[2] J. L. Spiesberger, H. E. Hurlburt, M. Johnson, M. Keller, S. Meyers, and J. J. O'Brien, ''Acoustic thermometry data compared with two ocean models: The importance of Rossby waves and ENSO in modifying the ocean interior,'' Dyn. Atmos. Oceans **26**, 209–240 (1998).

[3] M. D. Vera, M. A. Dzieciuch, and the NPAL Group, ''Modeling the acoustic receptions at the NPAL array from the Kauai source,'' J. Acoust. Soc. Am. **113**, 2279 (2003).

[4] W. Munk and C. Wunsch, ''Ocean acoustic tomography: A scheme for large scale monitoring,'' Deep-Sea Res., Part A **26**, 123–161 (1979).

[5] B. Cornuelle, W. Munk, and P. Worcester, ''Ocean acoustic tomography from ships,'' J. Geophys. Res. **94**, 6232–6250 (1989).

[6] The AMODE-MST GROUP, ''Moving ship tomography in the North Atlantic,'' EOS Trans. Am. Geophys. Union **75**, 17,21,23 (1994).

[7] T. F. Duda, R. A. Pawlowicz, J. F. Lynch, and B. D. Cornuelle, ''Simulated tomographic reconstruction of ocean features using drifting acoustic receivers and a navigated source,'' J. Acoust. Soc. Am. **98**, 2270–2279 (1995).

[8] J. L. Spiesberger, A. Fabrikant, A. Silivra, and H. E. Hurlburt, ''Mapping Climatic Temperature Changes in the Ocean with Acoustic Tomography: Navigational Requirements,'' IEEE J. Ocean. Eng. **22**(1), 128–142 (1997).

[9] ATOC Instrumentation Group: B. M. Howe, S. G. Anderson, A. B. Baggeroer, J. A. Colosi, K. R. Hardy, D. Horwitt, F. W. Karig, S. Leach, J. A. Mercer, K. Metzger, Jr., 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. Worcester, ''Instrumentation for the Acoustic Thermometry of Ocean Climate (ATOC) prototype Pacific Ocean network,'' OCEANS 95 Conference Proceedings, San Diego, CA, 9–12 October 1995, pp. 1483–1500.

[10] V. A. Del Grosso, ''New equation for the speed of sound in natural waters with comparisons to other equations,'' J. Acoust. Soc. Am. **56**, 1084–1091 (1974).

[11] S. Levitus, ''Climatological atlas of the world ocean,'' in NOAA Prof. Pap. 13 (U.S. Government Printing Office, Washington, DC, 1982).

[12] C. Garrett and W. Munk, ''Space-time scales of internal waves,'' Geophys. Fluid Dyn. **2**, 225–264 (1972).

[13] M. A. Wolfson and J. L. Spiesberger, ''Full wave simulation of the forward scattering of sound in a structured ocean: A comparison with observations,'' J. Acoust. Soc. Am. **106**, 1293–1306 (1999).

[14] J. A. Colosi, E. K. Scheer, S. M. Flatte, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).

[15] S. M. Flatte, J. A. Colosi, M. A. Dzieciuch, and P. F. Worcester, ''Acoustic observations of internal-wave strength in the Mid-Pacific in 1989 and 1996,'' J. Acoust. Soc. Am. **100**, 2582 (1996).

[16] J. Simmen, S. M. Flatte, and G. Wang, ''Wavefront folding, chaos, and diffractions for sound propagation through ocean internal waves,'' J. Acoust. Soc. Am. **102**, 239–255 (1997).

[17] V. Renard and J. P. Allenon, ''Sea Beam, multi-beam echo-sounding in 'Jean Charcot': Description, evaluation and first results,'' Int. Hydrogr. Rev. **56**, 35–67 (1979).

[18] National Geophysical Data Center, Boulder, CO, ''5 minute gridded world elevations and bathymetry-a digital database,'' 1987.

[19] F. Tappert, J. L. Spiesberger, and L. Boden, ''New full-wave approximation for ocean acoustic travel time predictions,'' J. Acoust. Soc. Am. **97**, 2771–2782 (1995).

[20] J. B. Bowlin, J. L. Spiesberger, T. F. Duda, and L. E. Freitag, ''Ocean acoustical ray-tracing software RAY,'' Woods Hole Oceanographic Technical Report, WHOI-93-10 (1993).

[21] J. L. Spiesberger and F. D. Tappert, ''Kaneohe acoustic thermometer further validated with rays over 3700 km and the demise of the idea of axially trapped energy,'' J. Acoust. Soc. Am. **99**, 173–184 (1996).

[22] J. L. Spiesberger, ''Ocean acoustic tomography: travel-time biases,'' J. Acoust. Soc. Am. **77**, 83–100 (1985).

[23] A. Draganov and J. Spiesberger, ''Diffraction and pulse delay in a structured ocean,'' J. Acoust. Soc. Am. **98**, 1065–1074 (1995).

[24] A. Silivra, J. L. Spiesberger, A. Fabrikant, and H. E. Hurlburt, ''Acoustic tomography at basin-scales and clock errors,'' IEEE J. Ocean. Eng. **22**(1), 143–150 (1997).

[25] V. H. Lichte, ''Über den Einflußhorizontaler Temperaturschichtung des Seewassers auf die Reichweite von Unterwasserschallsignalen,'' Phys. Z. **17**, 385–389 (1919). English translation by A. F. Wittenborn, with a forward by R. J. Urick is available from Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

Sub-bottom profiling using ocean ambient noise

Chris H. Harrison[a)]
SACLANT Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy

(Received 11 July 2003; accepted for publication 11 December 2003)

Spectral factorization is shown to restore the phase of an incoherent layered sediment reflection coefficient so that its Fourier transform is the minimum phase impulse response at each angle. The method requires the reflection coefficient to be known over a range of frequencies and the grazing angles in question to be above critical. It is developed here in the context of another recently established technique for extracting the seabed's plane wave reflection coefficient from ambient noise data measured on a moored or drifting vertical array (VLA). Thus it offers the possibility of sub-bottom profiling from a single platform with no sound source. Limitations of the phase restoration method are discussed and, using modeled data, comparisons are made between the ''true'' impulse response derived from the known complex reflection coefficient and the result of applying spectral factorization to the absolute value of the reflection coefficient. The method is also demonstrated on experimental reflection loss inferred from ambient noise measurements at three moored VLA sites and one VLA drift track in the Mediterranean Sea. Sub-bottom profiles (impulse response versus position) are shown for the drift track demonstrating that one can indeed survey with only a single directional receiver. The technique appears to perform well when compared with other profiling techniques and published results.
[DOI: 10.1121/1.1645854]

PACS numbers: 43.30.Pc [AIT] Pages: 1505–1515

I. INTRODUCTION

In sonar applications of underwater acoustics bottom reflection loss is usually treated as a frequency domain quantity (Jensen *et al.*, 1994; Tolstoy and Clay, 1987; Brekhovskikh and Lysanov, 1982). That is to say, at each frequency and grazing angle it has an amplitude and a phase. In this sense the reflection loss is the frequency response or transfer function (Rabiner and Gold, 1975) of the seabed, and with realistic seabeds this function can be complicated because of multiple arrivals and internal reflections from sediment layers (Brekhovskikh, 1980), not to mention various types of wave (compression, shear, visco-elastic, and so on) (Jensen *et al.*, 1994). Looking at the same phenomena in the time domain we see explicit arrivals from the layers and again later arrivals from interlayer reflections (e.g., Holland and Osler, 2000). In this alternative sense the time domain picture is an impulse response (Rabiner and Gold, 1975) of the seabed. This arrival structure manifests itself as the ''layering'' seen in most geophysical surveying and sub-bottom profiling techniques (Trabant, 1984). Clearly the two view points are equivalent. In fact the complex plane wave reflection coefficient (the transfer function) at a particular grazing angle is the Fourier transform (FT) of the impulse response for that angle. Thus knowing the impulse response for a set of angles one can calculate the reflection coefficient, or more importantly, in the current context, one can calculate the layer picture from the reflection loss. Indeed knowing the complex reflection coefficient as a function of geographic position one can reconstruct the sub-bottom profiling result.[1]

Before discussing the practical complications stemming from lack of phase information we set these thoughts in the context of a recently established ''noise inversion'' technique where the modulus square of the reflection coefficient is deduced from beam-steered ambient noise measurements on a vertical array (VLA) (Harrison and Simons, 2002; Tang, 2003; Desharnais *et al.*, 2003; Donelly and Matt, 2003). Harrison shows that the up-to-down-ratio at each angle is, in fact, the modulus square of the plane wave reflection coefficient. This technique has already been shown theoretically (Harrison and Baldacci, 2002a, b) and experimentally (Harrison, 2002) to function as a survey technique with a drifting VLA measuring ''local'' properties with a footprint of order the water depth (depending on angle). Therefore the possibility of converting this result directly to a sub-bottom profile is an intriguing prospect. The rest of this paper is devoted to exploring the most contentious step which is how to reconstruct the apparently lost phase of the reflection coefficient. Techniques are already available in the literature, in particular ''spectral factorization'' which has been used in geophysical prospecting and seismic inverse methods (Claerbout, 1985; Robinson, 1983), detection of submerged objects (McDaniel, 1999; McDaniel and Clarke, 2001) and antenna design (Sarkar, 1998). These techniques are closely related to the Hilbert transform and the theory of analytic functions (Morse and Feshbach, 1953), and these are, in turn, related to the Kramers–Kronig relations in optics (Ditchburn, 1963; Kramers, 1927; Kronig, 1926).

A separate implication of this frequency to time domain transformation is for geoacoustic inversion. Search techniques have been applied to VLA noise data (Simons and Harrison, 2004), but, quite generally, when searching for a fit to any acoustic measurement by varying geoacoustic parameters in a model, there is the problem of setting the search bounds. In particular one usually has to assume a number of

a)Electronic mail: harrison@saclantc.nato.int

layers, and one is tempted to find effective parameters (Harrison and Siderius, 2003) for this assumed number, in other words, one finds simplified nonunique solutions. In the present case, given the reflection loss, we transform it rather than invert it, the result being an impulse response which explicitly shows layer boundaries.[2] The number of layers is therefore predicted by the method and an assumption is not necessary.

In Sec. II we demonstrate explicitly the Fourier transform relationship between modeled reflection coefficient and impulse response. First, this is shown starting with the calculated complex reflection coefficient where we expect a perfect result. In passing we look at the autocorrelation function of the impulse response. Then we develop the spectral factorization technique, summarize its known performance characteristics, and demonstrate a comparable (but not identical) result starting with only the modulus of reflection coefficient. Finally, in Sec. III the approach is applied to some reflection loss data, inferred from experimental noise measurements at three moored VLA sites and one 5-mile drifting VLA track in the Mediterranean Sea. The results are compared with earlier findings for the same sites and a recent seismic boomer record along the same track. There are some points of particular relevance to this noise inversion technique. Otherwise the transformation from frequency domain reflection loss to time domain layer structure is quite general.

II. SUB-BOTTOM PROFILE FROM MODELED REFLECTION LOSS

Before launching into the difficulties of retrieving the reflection coefficient's phase it is educational to model the plane wave reflection coefficient of a hypothetical layered medium. In this way we have knowledge of the true phase and we can investigate the relationship between the complex reflection coefficient and the impulse response, and also the relation between the power reflection coefficient (a power spectrum) and its Fourier transform (the autocorrelation function of the impulse response).

There are many ways of calculating reflection loss of a layered half-space. For instance the model OASES (Schmidt, 1999) offers a solution for a point source and receiver with many layers, each with sound speed varying linearly with depth. One option (OASR) calculates the plane wave reflection coefficient explicitly as a function of frequency and angle. Various analytical approaches have also been used to include shear and density gradients (Ainslie, 1995, 1996; Robins, 1990, 1994). Here we take the straightforward numerical approach for plane waves given in Jensen *et al.* (1994). In their Sec. 1.6.4 an iterative technique is described in which one first calculates the response of a single fluid sediment layer over a fluid or solid substrate. Then, knowing the reflection coefficient at the upper boundary, one can add another fluid layer on top, thus building up an arbitrary number of layers. In a few lines of code one can handle a small number of isovelocity layers.

A. Complex reflection coefficient

As an example we take two sediment layers overlying a half-space with properties given in Table I. The interference pattern as a function of frequency and angle is shown in Fig. 1 and is easily understood qualitatively as a low loss total-internal-reflection region on the left with interference fringes on the right caused by the angle-dependent path differences between the layer boundaries (see Fiorito and Uberall, 1979; Jensen *et al.*, 1994; Harrison and Simons, 2002). The figure shows the modulus square reflection coefficient $|R|^2$ in dB, but because this was modeled we know the complex reflection coefficient, too. Therefore at each angle we can Fourier transform from frequency to time. Provided we take care to ensure that complex R is conjugate symmetric in (angular) frequency ω, i.e., $R^*(-\omega)=R(\omega)$, we can make use of negative frequencies (or frequencies between the Nyquist and the sampling frequency in a DFT). In the case of modeled R there is unlimited frequency range available, and we are at liberty to use any windowing function. We can also zero-pad to improve time resolution if desired.

TABLE I. Bottom properties used in the modeled example.

	Thickness (m)	Sound speed (m/s)	Specific gravity	Attenuation (dB/λ)
Water column	∞	1500	1	0
Sediment 1	1.5	1600	1.4	0.14
Sediment 2	2.0	1700	1.6	0.15
Sub-bottom	∞	1800	1.8	0.15

The result is shown in Fig. 2. A perfect reflection with $R=+1$ would result in the Fourier transform of the windowing function (a "pulse") being centered on time zero. For this reason a small time offset has been included in the illustrations. Throughout this paper we have adopted the convention that travel time increases down the page, so that later results resemble sub-bottom profiles.

Although one knows what to expect, there are several interesting and striking features:

(i) Below the critical angle the angle-dependent phase change (that may also be seen in the Rayleigh reflec-

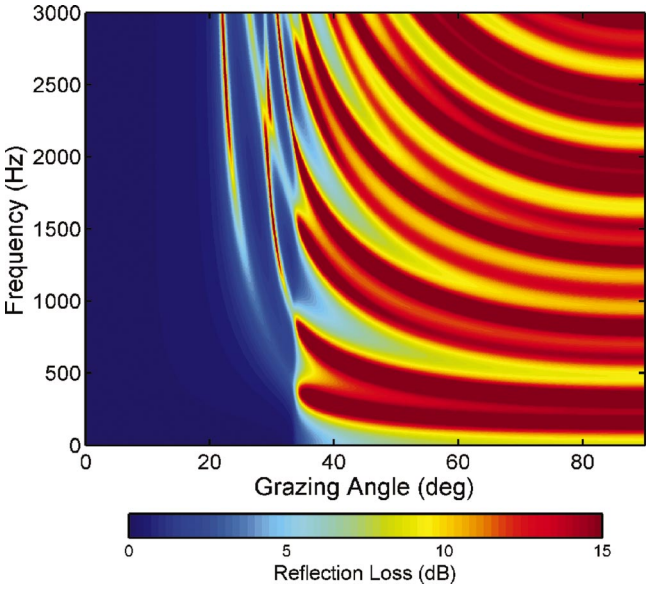


FIG. 1. Modeled reflection loss for two sediment layers (thicknesses 1.5 and 2 m) in between two half spaces; parameters given in Table I.

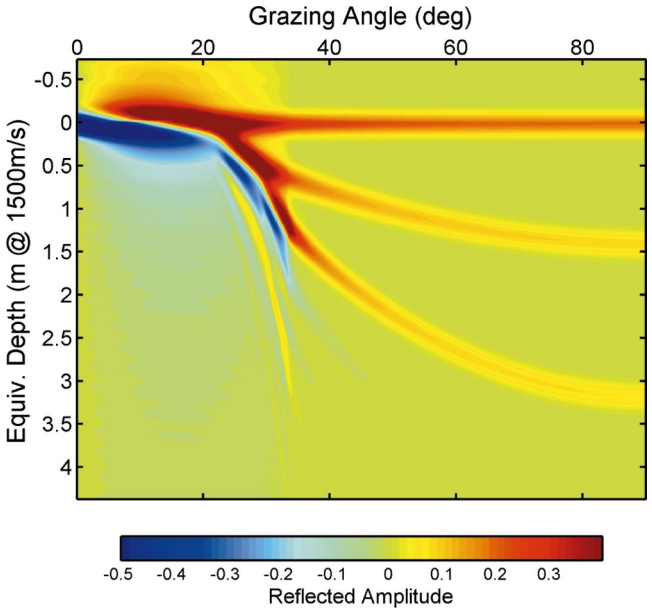


FIG. 2. The impulse response corresponding to the modeled reflection loss in Fig. 1 derived by taking the Fourier transform of the complex reflection coefficient separately for each grazing angle. Two-way travel time is converted to an equivalent depth at 1500 m/s.

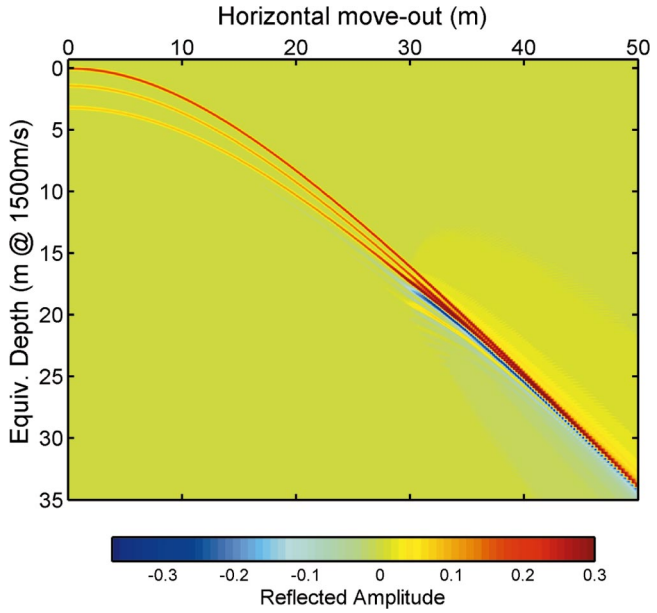


FIG. 3. The modeled impulse response from Fig. 2 remapped into equivalent depth versus horizontal move-out distance showing hyperbolic behavior (hypothetical monostatic source at 20 m above the seabed, i.e., equivalent depth $=-20$ m), clear layer arrivals near normal incidence, and merging of these arrivals at the critical angle, around move-out $=30$ m.

tion coefficient of a single boundary) can be seen as a negative pulse at very low grazing angles changing through a negative-pulse-with-overshoot to a positive pulse near critical.

(ii) The monotonically increasing sound speeds in this example produce three critical angles, one for each boundary. The two lowest critical angles can be seen as branches in the pulse arrival. The reason for this is easily seen from ray theory. The path difference in traversing a single layer of thickness h at angle θ in that layer is $s=\omega t=2h\sin\theta$. At normal incidence the path difference is a maximum, $s=2h$, and the angle in the layer reduces to zero at the critical angle so that the two arrivals merge. Below the critical angle there is only a single arrival. At the third critical angle we see no branch because there is no reflection from the bottom of the infinite half-space.

(iii) At each critical angle the arrival also has a slight slope discontinuity in angle-frequency space. This is perhaps clearest for the highest critical angle where there is no branch. The slight alteration of the travel time on the total-internal-reflection side of the branch point is caused by the ''lateral shift'' (Brekhovskikh and Lysanov, 1982), a manifestation of the reflection coefficient for this boundary becoming complex.

(iv) At angles of 25° and higher one can see clear signs of multiple reflections at delays corresponding to multiples and combinations of the two layer thicknesses. They remain visible until the lower boundary reflection coefficient deviates significantly from unity.[3]

(v) Notwithstanding the above multiples, there are three clear layer arrivals for high angles (i.e., three boundaries; four layers). Thus the *number* of layers can be seen in the time domain view even though it is not so clear in the frequency domain view (Fig. 1).

(vi) The spacing of the arrivals in time agrees unambiguously with the known layer spacings used in calculating the reflection coefficient (see Table I).

(vii) A vertical cross section of Fig. 2 is an impulse response rather than a true geological cross section. The strength of deep layer reflections is affected by transmission through the upper layers. Nevertheless, provided volume absorption produces only weak attenuation over the vertical extent of the layers and provided reflections are weak,[3] each pulse arrival amplitude is representative of that boundary's reflection coefficient in isolation and therefore the properties on either side of the boundary. Much effort has been expended by geophysicists in deducing geo-acoustic properties from the impulse response under more general conditions (Claerbout, 1985; Robinson, 1983)

It is interesting to compare the presentation in Fig. 2 with that of the standard ''move-out'' technique where we might have a fixed receiver and a horizontally moving source (Holland and Osler, 2000). In that case the travel time varies hyperbolically with move-out distance x. Making the transformation $x=z\cot\theta$ where z is the notional source/receiver height and shifting by an angle-dependent delay $t=(z/c)(\mathrm{cosec}\,\theta-1)$, Fig. 2 can be converted to the move-out format of travel time versus horizontal position, as shown in Fig. 3. Because this operation is simply a mapping, the branching phenomenon is still visible, and in principle gives a direct measure of the critical angle for each layer.

B. Modulus square reflection coefficient

Suppose we start by taking the FT of $|R|^2$ instead of complex R. By the Wiener–Khinchin relation the result of inverse Fourier transforming the modulus square of the re-

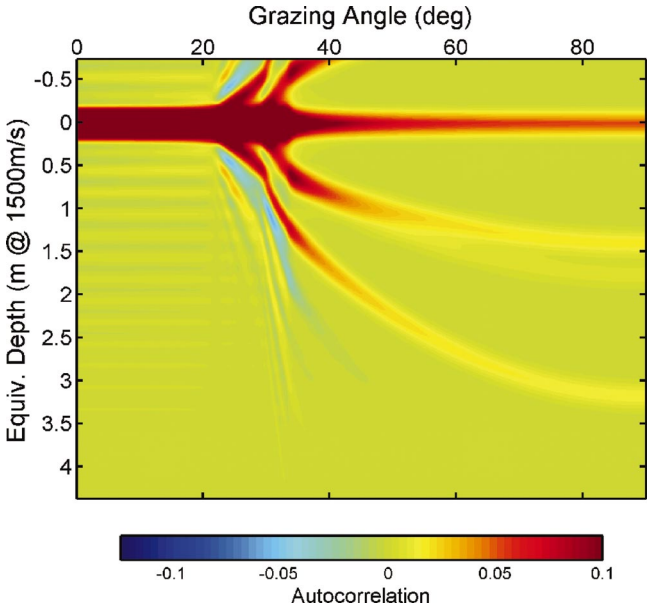


FIG. 4. The autocorrelation function of the modeled impulse response (shown in Fig. 2) derived by taking the Fourier transform of the modeled power reflection coefficient. The function is symmetrical about zero. In addition to the two arrivals corresponding to boundaries at 1.5- and 3.5-m depth there is a faint branch corresponding to the 2-m difference seen at 90° and equivalent depth 1.8 m.

flection coefficient (a power spectrum) is the autocorrelation function of the impulse response. Clearly this function must be symmetrical in time, with its largest value at time zero. One will see arrivals at combinations of layer thicknesses but layer ordering will not be distinguishable and the number of layers will not be evident, although one might be able to deduce upper or lower bounds on the number. The result of performing this operation on Fig. 1 is shown in Fig. 4. The lost information is evidently contained in the phase. Therefore the question arises, is there a way of estimating the reflection coefficient's phase that is good enough to provide a solution more like Fig. 2 than Fig. 4? Such a method for reconstructing the phase is spectral factorization (Claerbout, 1985).

1. Spectral factorization: A review

The reflection coefficient can be regarded as an analytic function since it is a continuous function of the complex variable ω as defined by Morse and Feshbach (1953, p. 356). This forces a relationship between the real and imaginary parts through the Cauchy–Riemann equations, and this relation can be expressed as the Hilbert transform. In the frequency domain form shown in Morse and Feshbach (1953, pp. 370–374) the Hilbert transform is a convolution with the function $1/\omega$. This is equivalent to multiplication by a step function in the time domain which is an alternative statement of the fact that the reflection process is a causal filter with a real impulse response.

By swapping domains the Hilbert transform can also be regarded as a quadrature filter, i.e., a 90° phase shift filter (Bellanger, 1994, p. 249). In the frequency domain this filter is a step function of amplitude $-i$ for positive frequencies and $+i$ for negative frequencies. In the time domain it corresponds to convolution by the function $1/t$. A simple numerical implementation is to take the FT of the given real function of time, set the values at negative frequencies to zero, and inverse FT. Applying this filter to the real time series x_R we obtain a corresponding imaginary series x_I. Alternatively one can construct an ''analytical signal'' (Bellanger, 1994, p. 247), $x=x_R+ix_I$, directly from x_R by multiplying the values for positive frequencies by $+2$ and zeroing the negative frequencies.

In the present case we are given a (real) power spectrum and we require a corresponding causal time series. Spectral factorization is an operation that finds a unique ''minimum phase'' time series that has the given spectrum (see, e.g., Proakis and Manolakis, 1996, p. 359). The Kolmogoroff spectral factorization method makes use of the Hilbert transform by moving it to the exponent (Claerbout, 1985).

Given a real spectrum h_R we shift it to the exponent by writing it as $h_R=\exp(\ln(h_R))\equiv\exp(g_R)$. Applying the Hilbert transform to g_R we obtain an imaginary part g_I and we construct the complex function $g=g_R+ig_I$. Now the complex h becomes $h=\exp(g_R)\exp(ig_I)$. Thus the original real spectral quantity h_R now has (an unchanged) amplitude $\exp(g_R)(=h_R)$ and a new phase g_I. Having retrieved the phase we can now Fourier transform complex h to obtain the desired impulse response.

In short, the operation to retrieve a time series $r(t)$ from the modulus of the reflection coefficient $|R|$ is

$$r(t)=\mathcal{F}^{-1}\{\exp(\mathcal{H}(\ln(|R(\omega)|)))\}, \tag{1}$$

where $\mathcal{F}$ is the Fourier transform and $\mathcal{H}$ is the operation of adding the Hilbert transform generated imaginary part to the original real part. This can be defined in terms of the Fourier transform using the unit step function w as

$$\mathcal{H}(x)=\mathcal{F}(2w\times\mathcal{F}^{-1}(x)). \tag{2}$$

There are several variants of this approach, such as Whittle's, Root, and Toeplitz (Claerbout, 1985). The result is always minimum phase. This follows from the fact that h_R is periodic in frequency which leads to g_R being periodic so that the net phase change after one period is zero. This means in practical terms that there is a tendency in these solutions for the strongest impulses to arrive first and for the time spread to be minimized. A clear discussion of the meaning of ''minimum phase'' is given by Proakis and Manolakis (1996, p. 359). The main problem with any phase retrieval method is one of uniqueness or ambiguity. From a filter theory point of view given a power spectrum there are many possible minimum, maximum, or mixed phase filters with that spectrum and we cannot tell whether the actual filter (the seabed) is really minimum phase or not. However, the seabed is not a completely arbitrary filter and there are other physical constraints which may be useful. Sarkar (1998) pointed out that using search techniques it is still possible to find mixed phase solutions but these are still not unique. These points are taken up later in the next section.

2. Spectral factorization: An example

Applying the Kolmogoroff method to the modulus of reflection coefficient shown as reflection loss in Fig. 1 we

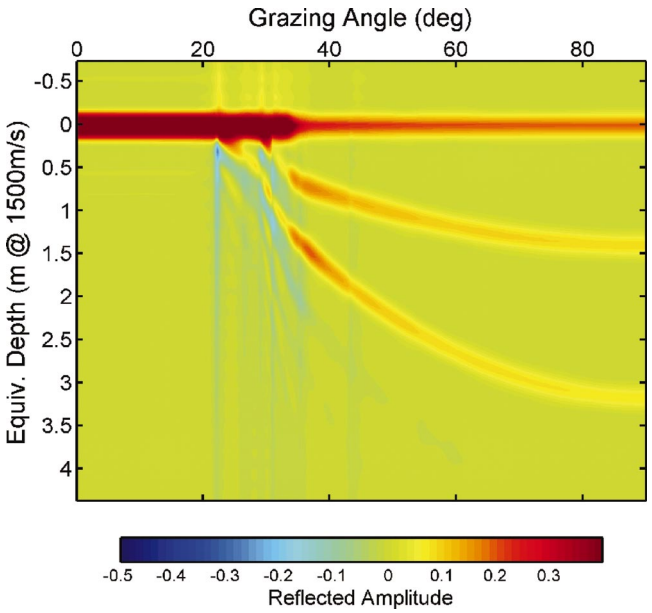


FIG. 5. The result of spectral factorization applied to the modulus of the reflection coefficient depicted in Fig. 1. Note that layers are in the correct order without any ghosting for high angles. Below critical there is no hint of the arrival's sign change seen in Fig. 2.

obtain the impulse responses shown in Fig. 5. Comparing this with the true time series (Fig. 2) we see that at large angles the layer boundaries are well distinguished. Their order and their number are (in this example) clear despite earlier reservations. This is already a benefit from the inversion point of view. Their intensities are also close. One can also see a multiple reflection at around 30°–40°. At angles lower than the biggest critical angle there are differences between Figs. 5 and 2 caused by the minimum phase condition; the variation of the pulse shape with angle seen in Fig. 2 is absent in Fig. 5. This is hardly surprising since at these angles there is hardly any frequency variation for the Hilbert transform to work on. Thus the result has a uniform zero phase for all angles. Also the branching is unclear, and the spectral factorization, method seems to settle down only when the reflection coefficient is real (i.e., when clear of all critical angles).

It is interesting to compare these plots with the less contentious autocorrelation function of the impulse response (Fig. 4) derived by simply Fourier transforming the power reflection coefficient. The same time shift of the origin has been imposed for clarity, but obviously the whole function is symmetrical about zero. Notice that although many of the features are similar the autocorrelation function is really showing a response to differences in layer boundary depth rather than just depth. Thus (with layer thickness 1.5 and 2.0 m, i.e., layer boundaries at depths 0, 1.5, and 3.5 m) we see all possibilities for depth differences $1.5=|0-1.5|$, $2.0=|1.5-3.5|$, and $3.5=|3.5-0|$ m So the "new" layer at 2.0 m is just a manifestation of the autocorrelation function's ambiguity.

This example is reasonably realistic, but it is impossible to be representative without a large number of examples. For instance, one anticipates the existence of pathological cases such as layers with impedance changes designed to give boundary reflections whose strength increases progressively with depth from a small value at the shallowest. Although this physical filter is patently causal the minimum phase response will tend to reverse the boundary order (Claerbout, 1985). It is easy to demonstrate this effect numerically by inventing a time sequence of separated spikes, taking the Fourier transform, then applying Eq. (1), and comparing the result with the original. In fact the test shown in Fig. 5 follows precisely this process except that the constraints of reflection physics have been incorporated in the otherwise arbitrary input impulse response. Even with these additional constraints, if one is prepared to turn the entire environment upside down (including water and bedrock), one obtains the exact inverse impulse response (provided there is no absorption)[4] and therefore the same power reflection coefficient (as a function of horizontal wavenumber), as can easily be verified numerically. Fortunately in the acoustics of bottom reflection the density difference between water and sediment typically makes the first return stronger than the later ones even though the sound speeds may differ from layer to layer, ultimately becoming substantially greater than that in water. So the minimum phase condition is often likely to be met. This contrasts with findings in other domains (McDaniel, 1999; McDaniel and Clarke, 2001; Sarkar, 1998). In general spectral factorization tends to work well if the real response has a strong first arrival followed by not too large a number of weaker ones—larger numbers are tolerated if they are weak enough.

The conclusion is that the method appears to extract valuable information, in particular the number of effective layers, from the modulus of the reflection coefficient which would otherwise require considerable effort in a parameter search technique.

III. SUB-BOTTOM PROFILE USING REFLECTION LOSS INFERRED FROM AMBIENT NOISE MEASUREMENTS

We now apply the spectral factorization process to some experimental data obtained with a moored or drifting vertical array (VLA). In all cases 32 equally spaced elements of the VLA (separation 0.5 m) were sampled at 6 KHz, beamformed, and analyzed as described in Harrison and Simons (2002) to obtain a reflection loss as a function of angle and frequency. Experiments at the three moored sites during MAPEXbis2000 included a sand site with a clear critical angle south of Sicily, and two silt/mud sites with low sound speeds close to Elba. In experiments during BOUNDARY2002 (Harrison, 2002) the VLA was allowed to drift for about 12 h covering about 5 miles near the Ragusa Ridge south of Sicily. Maximum likelihood adaptive beam forming was used in this case. In all cases stable reflection loss images are obtained every few minutes by averaging 128-point FFTs.

So far in this paper it has been assumed that the starting point is a perfect and complete modulus of reflection coefficient over all angles and a range of frequencies big enough to perform FTs without windowing problems. The design frequency of the array is 1500 Hz, and the implications for performance and resolution in angle and frequency are dis-

cussed by Harrison and Baldacci (2003). In short there is a window in frequency whose extent is angle dependent. The upper frequency limit is set by the onset of grating lobes

$$f_U = 2\, f_o/(1+\sin\theta), \tag{3}$$

where $f_o = c/2a$ is the design frequency, related to sound speed c and hydrophone spacing a. The lower frequency limit is set by beam angle resolution, roughly

$$f_L = (2\, f_o/N)/\sin\theta, \tag{4}$$

where N is the number of hydrophones. In an ideal arrangement we would have large N (regardless of array length) to avoid end effects. In these examples we treat each angle independently and we have a rather short usable range of frequencies although we can still make use of the conjugate symmetry of complex R and the symmetry of $|R|$. We therefore need to devise some preprocessing before attempting spectral factorization.

A. Preprocessing the noise data

Simple tapered windowing is inclined to overwhelm the limited data we have while we have strong *a priori* reasons to believe that the pattern repeats in frequency beyond the truncation. Spectral estimation techniques (Marple, 1987) handle this problem, but in this context we also wish to impose the constraint of symmetry on the periodic components of $|R|^2$. In other words we should be able to write

$$|R|^2 = \sum_{n=0} a_n \cos(\omega \tau_n), \tag{5}$$

where n could be finite or infinite. Also the main problem is found to be bridging the gap across zero frequency from positive to negative frequencies. In other words the lower frequency limit is more problematic than the upper. The approach used here is to extrapolate across the gap and to use a windowing function symmetric about zero frequency. Also $|R|^2$ is extrapolated above the upper frequency to 3 KHz (arbitrarily chosen as twice the upper frequency limit) in the knowledge that a window function such as hamming will, in any case, deemphasize this region.

Thus the preprocessing algorithms extrapolate $|R|^2$ retaining the unchanged original, as explained below. The main processing is then to apply spectral factorization [the curly bracketed part of Eq. (1)], then apply a hamming window (symmetrical about zero), then zero-pad (to obtain good time resolution), and finally apply a FFT. Therefore the same operation was performed on the extrapolated experimental data as on the modeled data in Sec. II.

The extrapolation technique was as follows. Remembering that we have good reason to believe that the experimental data must be a truncated version of Eq. (5) we suspect that there are retrievable Fourier components but that the truncation degrades their frequency resolution. If we can somehow find these component frequencies and their amplitudes (by inspection the number of components is small), we can reconstruct the function with cosines at those frequencies. The result is then truly periodic and extends to infinity. The actual procedure is iterative and it is easy to check the fidelity of the fit against the original. Obviously we do not trust the extrapolation too far in frequency from the experimental data. In each iteration we determine a single frequency component by first subtracting the mean, then taking the FFT and finding the real part of the value at the position of the absolute maximum. This allows for the possibility of negative cosine coefficients. At the end of each iteration the sum of all components found so far is subtracted from the original, and then the next iteration works on the residual. In this way one can pick out a predetermined small number of cosine components. In practice the number of terms appears not to be important and is usually between about one and five. Although one could find a perfect fit, it is not necessary in this case because the final step is to patch the extrapolations to the unchanged original using a weight function of the form w for the experimental data and $(1-w)$ for the extrapolation:

$$w = \{1+\tanh((f-f_L)/f_b)\} \times \{1+\tanh((f_U-f)/f_b)\}/4.$$

Here f_L, f_U, f_b are respectively lower and upper frequency limits of the experimental data and a small changeover width. So w is, to all intents and purposes, flat in between f_L and f_U.

With anything more than a single cosine component there is a risk of the extrapolation going negative at some point with disastrous consequences for spectral factorization. This can be averted by the simple ploy of taking the log of the quantity before finding the Fourier components, then taking the exponential of the result (which is guaranteed positive). One might object that this could interfere with the spectral factorization, however it is stressed that we are retaining the original data intact, and this extrapolation is merely a means of allowing standard windowing in subsequent FTs without throwing away the original data.

B. Moored experiment examples

1. S. Sicily, sand site

The impulse response as a function of angle is shown in Fig. 6. The y axis is a two-way-travel time converted to an ''equivalent'' depth using a speed of 1500 m/s. Considering that the processing was carried out independently for each angle (1° separation), the plot is remarkably clean from 90° down to 30°. At angles below this the zero time arrival contains no extra information as was found with the hypothetical case. The other features at angles lower than 10° extending to a few meters depth are believed to be artifacts generated by residuals in the inferred reflection loss. Taking this plot at face value and comparing it with Fig. 2 we see that a critical angle is predicted near 25°. However, we note that the first and second arrival do not quite merge. The reason for this is that in the frequency domain the fringes should theoretically go rapidly to infinity at the critical angle (see Fig. 2) while the impulse separation goes to zero (Fig. 1), but the fringes cannot go to infinity in practice because of the angle blurring effect of the beam forming particularly near the critical angle.

This example corresponds to the experimental inferred reflection loss in Fig. 5 of Harrison and Simons (2002) where a 1-m-thick layer with speed 1554 m/s was found by a hand search. When converted to equivalent thickness, this is 0.96

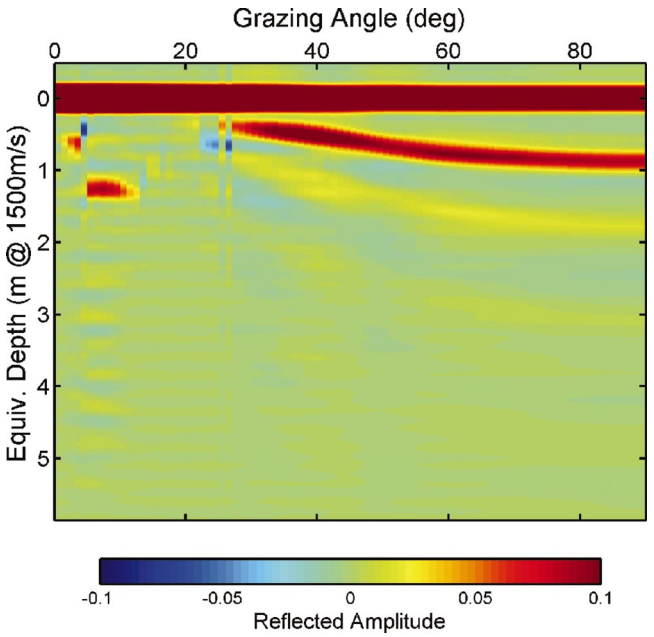


FIG. 6. Experimental impulse response derived by spectral factorization from ambient noise measured on a moored VLA at a sand site south of Sicily. There is one clear arrival corresponding to a 1-m layer with a critical angle. The later, fainter arrival is believed to be another layer rather than a multiple reflection.

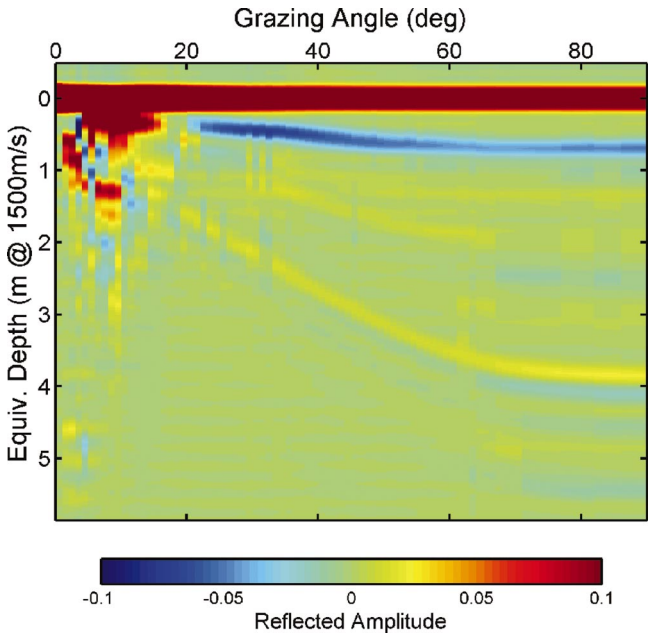


FIG. 7. Experimental impulse response derived by spectral factorization from ambient noise measured on a moored VLA at a mud site east of Elba. There is a clear negative arrival (blue) and a deeper positive arrival (yellow) corresponding respectively to the bottom of a thin high speed layer and the bottom of a deeper low speed layer.

m, which is close to the thickness seen in Fig. 6 at 90° of 0.88 m. Interestingly, there appears to be another weaker layer at 1.78 m, suspiciously close to being a multiple reflection, although at 30° the timing is not quite right. It is also possible to comment on this third arrival from the point of view of absolute intensity. Each arrival amplitude is absolute and corresponds to one of the terms in the usual derivation of the joint reflection coefficient in terms of reflection and transmission coefficients for each boundary (see Jensen *et al.*, 1994, p. 50). The first arrival (at zero time) has amplitude $a_1 = 0.344 = R_{12}$, the second at 0.877 m has amplitude $a_2 = 0.086 = R_{23}\, T_{12}\, T_{21}$ (with unambiguous interpretation), and the third at 1.78 m has amplitude $a_3 = 0.0157$. Regardless of the geoacoustic parameters leading to these numbers, one can say that the first multiple between the first two boundaries cannot be stronger than $R_{23}R_{21}R_{23}T_{12}T_{21} = R_{21} \times (R_{23}\, T_{12}\, T_{21})^2/(T_{12}\, T_{21}) = a_1\, a_2^2/(1-a_1^2) = 0.0059$. Furthermore, the sign must be the opposite of a_1, in other words negative. Therefore the third reflection cannot be a multiple. It could be an artifact of the present implementation of spectral factorization although its presence does not depend strongly on the exact implementation. It is more likely that it was not observed by Harrison and Simons (2002) because, being near to exact double spacing, it was difficult to detect. In principle one could use these amplitudes for geoacoustic inversion although we will not attempt it here. Also it is not obvious that there would be any benefit over the usual search techniques in the frequency domain.

2. E. Elba, mud site

In the vicinity of this site other authors (e.g., Murphy *et al.*, 1976) have found a thin layer (<0.5 m) of high speed sediment over a thicker layer of lower speed. The inferred reflection loss is shown in Fig. 7 of Harrison and Simons (2002), and from a hand search two sediment layers were found with thicknesses 0.8 and 3.5 m and speeds 1530 and 1471 m/s which agrees well with Murphy *et al.* (1976) allowing for geographic variation. Figure 7 in the present paper shows the impulse response versus angle. Again the angle variation is very smooth and clean from 90° down to 20°. The two obvious features are the negative arrival (blue) which reaches 0.693 m at 90° and the weaker positive arrival at 3.85 m. These agree well with the equivalent depths of 0.78 and 3.57 m from Harrison and Simons (2002). Furthermore, the signs of the three arrivals are exactly what one would expect from the sequence of sound-speed-density products deduced by Harrison and Simons (2002) (i.e., $1509 \times 1 = 1509$, $1530 \times 1.4 = 2142$, $1471 \times 1.2 = 1765$, $1530 \times 1.8 = 2754$), which lead to vertical impedance differences of 633, −377, and 889 m/s.

Noting the angle dependence of the second and third arrivals, there appears to be no tendency to merge with each other or with the first arrival. In fact, they both line up quite well with the origin, implying that there is no critical angle. We cannot see this merging at the origin explicitly because below about 15° the impulse response is simply obscured by what are most likely artifacts from the original inferred reflection loss. The amplitudes of the three arrivals at 90° are respectively +0.264, −0.051, and +0.023. One could use these numbers for inversion, but we do not attempt this here.

3. N. Elba, silt site

This area was known to have a more complicated layer structure. Murphy *et al.* (1976) suggested for a nearby site a thin high speed layer in the middle of a low speed sediment layer, but Harrison and Simons (2002) found a best fit for the

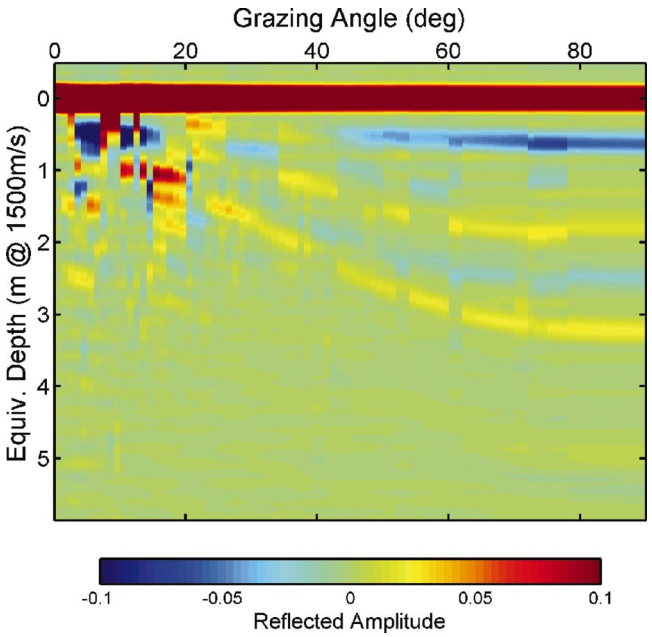


FIG. 8. Experimental impulse response derived by spectral factorization from ambient noise measured on a moored VLA at a silt site north of Elba. There is a clear negative arrival (blue) corresponding to the boundary between a thin high speed layer and deeper low speed layers (yellow).

thin layer at the top. Figure 8 shows that the structure is slightly more complicated. The first arrival (amplitude 0.334) is followed by a negative arrival of amplitude −0.75 at 0.66 m and two positive arrivals of 0.032 at 1.85 m and 0.023 at 3.23 m. One can also see fainter negative echoes in between. The equivalent depths from Harrison and Simons (2002) are close at 0.66 and 2.06 m, and again the signs of these arrivals fit the inference of a thin fast layer over a thicker slow layer from Harrison and Simons (2002). The spectral factorization algorithm seems slightly less stable with angle change, but nevertheless the arrivals appear to converge on the origin again implying a lack of a critical angle at least in the lower layers.

C. Drift experiment example

S. Sicily

In the earlier examples the VLA was moored at one site. In a drift experiment we have the same type of information at our disposal for all positions along the drift track. So there is the potential to construct a bottom profile with a beam looking at any angle (notwithstanding the low angle limitations and the insensitivity to azimuth). Note that in the case of a vertical beam the reservations about performance of the spectral factorization technique are irrelevant. During the drift experiment, a reflection loss plot was extracted every 10 s before forming a running average with a time constant of a few minutes. So, in effect, the information is spatially sampled every few minutes. Drift rates were of order half a knot or 0.25 m/s and there was a footprint of order the water depth (Harrison, 2002a) which varied by a few meters around 130 m. So in 5 min the VLA had moved by 75 m, well inside the footprint.

The 12-h, 5-mile profile constructed from arrivals between 80° and 90° is shown in Fig. 9(a). The horizontal scale is geo-time (measured in hours after midnight on 22 April 2002) which translates into position as the line labeled with circle symbols on the map shown in Fig. 10. The beginning of the track is close to the sand site discussed above, and indeed for most of the track one can see a dominant single positive layer reflection which appears to change depth in a realistic fashion. In addition, at later delays up to 5-m depth there are apparently various layer structures whose depths change independently of the first arrival. Their amplitudes also vary and there are occasional strong negative arrivals that seem to be more than just an overshoot in delay time. The unusual feature at geo-time about—4.5 h was caused by interference from a well-documented large ship very close by and a fishing boat almost on top of the VLA. Interestingly, layer depths are almost unperturbed although amplitudes are significantly altered.

Even though we cannot guarantee that the true impulse response is minimum phase [so that Fig. 9(a) would correspond with what would be seen by a comparable specification active profiling system] we can guarantee that the autocorrelation function is immune from these reservations, and in this case the autocorrelation (not shown) has similar behavior (namely a strong first arrival followed by many independently shifting arrivals) though with the expected ambiguities.

The first draft of this paper compared the spectral factorized result with three fairly nearby boomer tracks taken from Osler and Algan (1999). However, since then, in July 2003, the drift track was revisited during BOUNDARY2003 with the same boomer equipment, and the boomer record is shown in Fig. 9(b) aligned with the noise profile. Note that the boomer's two-way travel time naturally includes transit from sea surface to the seabed (see Fig. 11), whereas the spectral factorized ambient noise examples are referred to the first arrival. So, for the sake of comparison, in Fig. 9(b) the boomer has been zero-referenced by taking out the travel time to the seabed. The tracks (labeled A-B) are not quite coincident, as can be seen in the map of Fig. 10, but they are much closer than the nearest tracks from Osler and Algan (1999). The boomer does not have a perfect impulse response; it has a very sharp initial pulse which can be seen as the thin line at the top of Fig. 9(b) with ringing in the form of the thicker ghost line at about 0.4 m and the second ghost at about 1.5 m.

Bearing this in mind, there are a number of features in common between the boomer and the spectral factorized ambient noise (referred to below simply as the ''noise''). From A to C there is an echo at around 1 m, in agreement with the red line at comparable depth in the noise. Similarly, between E and G the boomer has an echo going from about 1.0 m at E to about 1.6 m at F, which agrees with the red line in the noise although it is slightly obscured by the constant delay ghost at 1.5 m (seen between F and G). From G to B the boomer echo at about 1.5 m starts to undulate rapidly in a similar way to the noise curve. Between D and E the boundary at less than 0.5-m depth in the noise corresponds to the breaking up region in the boomer (wavy layers). These wavy strata are probably oblique to the track and therefore grossly spatially averaged by the noise method. Between C and G

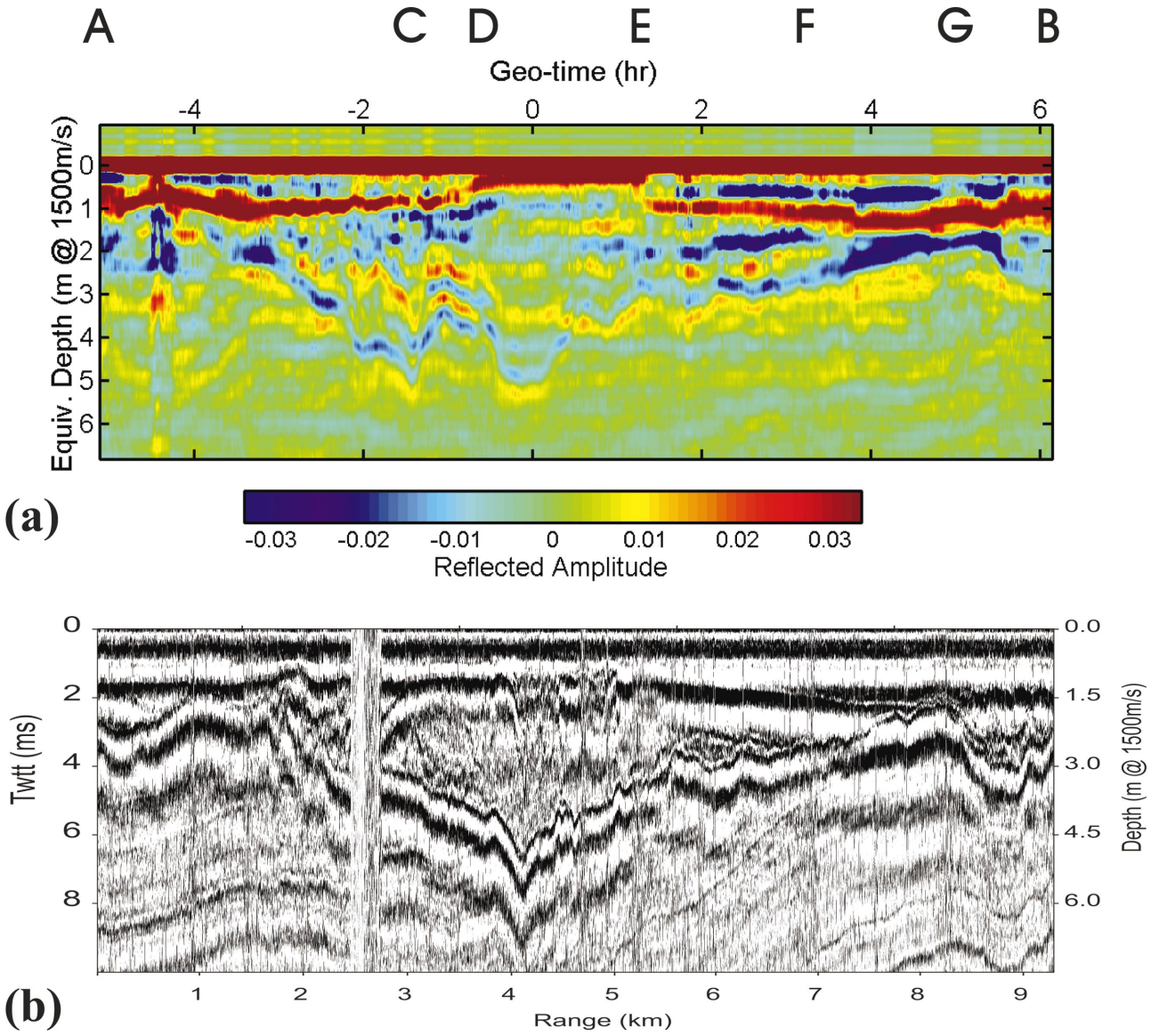


FIG. 9. (a) Sub-bottom profile derived, via reflection loss, from ambient noise received by a drifting VLA. (b) Zero-referenced and aligned boomer sub-bottom profile for comparison.

the deeper echo that goes down to about 5 m between C and E and then rises to a peak at about 1.5 m at G can be seen in the noise data with approximately the same depths. There are also hints of all the other boomer echoes in the noise data. However, one suspects that these echoes are often triplicated in the boomer as in the first return. For instance, evidence of this can be seen by inspecting the three black lines starting at depths of about 3.1, 3.5, and 4.6 m at C and following them to the right; they appear to be parallel. Considering that this comparison constitutes a blind test of the spectral factorization algorithm, agreement is quite remarkable.

IV. CONCLUSIONS

Spectral factorization has been used in a number of fields to restore lost phase information. In the present paper it was applied to the modulus of the seabed's reflection coefficient so that subsequent Fourier transformation of the new complex reflection coefficient produces a minimum phase impulse response for each angle. The performance was

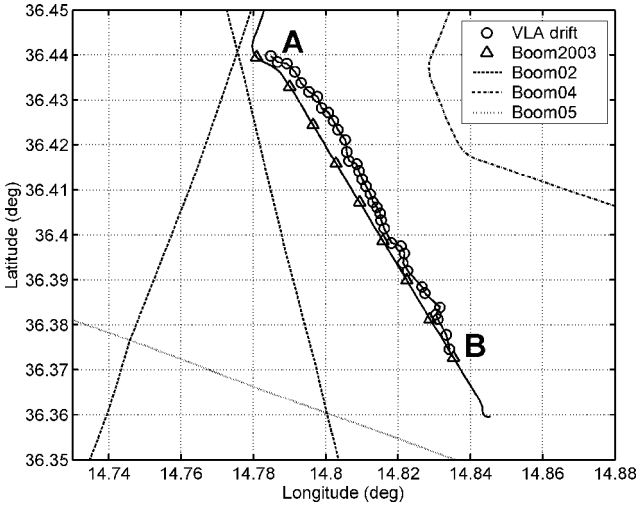


FIG. 10. Map showing the 5-mile VLA drift track (-○-), the boomer track from BOUNDARY2003 (-△-), and nearby boomer tracks from Osler and Algan (1999) labeled as Boom02, Boom04, Boom 05.

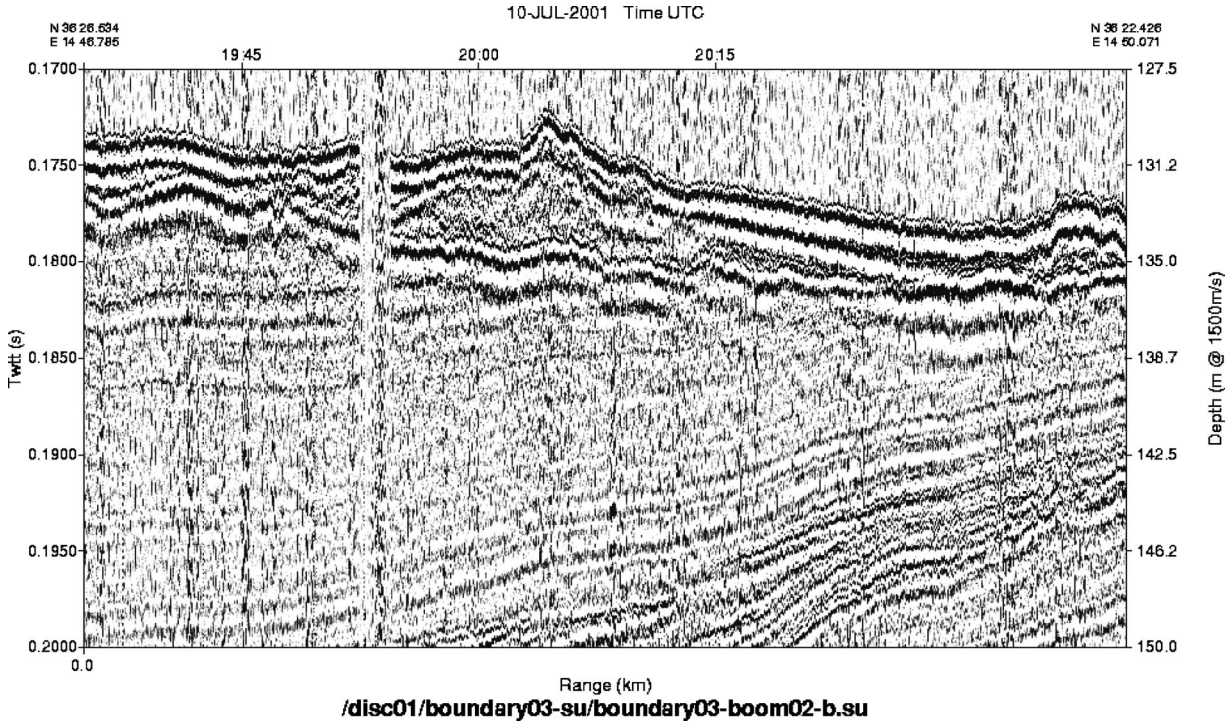


FIG. 11. Unmodified boomer sub-bottom profile taken during BOUNDARY2003 along more or less the same track (A-B) as the VLA drift (see Fig. 10).

checked first by modeling the joint complex reflection coefficient of a set of layers and taking its Fourier transform to obtain a ''true'' impulse response versus angle. The time domain behavior is interesting in its own right and clearly shows an arrival for each boundary at high angles. The chosen example had two sediment layers with monotonically increasing sound speed and therefore three critical angles. In time-angle space one could see that the arrivals for adjacent boundaries merged at their critical angles. The well-known phenomenon of the change of the reflection coefficient's phase angle between zero grazing and the critical angle is seen in the time-domain as a negative arrival transforming through negative-pulse-with-overshoot to a positive pulse at critical. It was also shown how the impulse response versus angle is related to the hyperbolic layering images typical of the geophysical ''move-out'' technique.

By applying spectral factorization to the modulus of the reflection coefficient and then Fourier transforming the result a similar angle-dependent impulse response was calculated. In this example the algorithm was able to determine layer structure including layer ordering, and results were indistinguishable at high angles from the true impulse response derived earlier. However, at angles below the highest critical angle the minimum phase response understandably fails to reproduce the detailed changes in pulse shape of the true response and yields instead a simple positive pulse at all angles.

The Fourier transform of the power reflection coefficient is the autocorrelation function of the impulse response, and clearly has the same layer ordering ambiguity as the power reflection coefficient. From this view point it is easy to see that pathological cases must exist where impedance differences are small for the first arrivals but large for later ones. Fortunately the density difference between water and sediment usually makes the first arrival strongest so that the true impulse response is indeed minimum phase.

Although the technique can be applied generally to incoherent layer reflections, the main interest in this paper is application to reflection loss inferred from directional ambient noise measurements made using a VLA. Examples taken from three moored sites and one drift track in the Mediterranean Sea were investigated. The moored sites produced convincing results that were in close agreement with earlier publications on this data and other measurement techniques. In particular spectral factorization appears to give a good idea of the ordering and number of layers. For instance, in the E. Elba mud example the impulse response shows a thin high speed layer over a thicker low speed layer. In the N. Elba silt case there appear to be three or four boundaries.

The time-varying reflection loss from the VLA drift experiment can be translated into depth profiles for beams at almost any angle since the data is the equivalent of a move-out experiment carried out about once per water depth in range. The general behavior of the layers, especially when compared with nearby boomer records, gives confidence that the result is close to a true sub-bottom profile. In particular, the boomer record along almost the same track is strikingly similar.

One could argue that the approach does not bring one any closer to geoacoustic parameters than the more conventional frequency domain inversion techniques. However, it does offer the possibility of at least determining the number of layers, if not determining their order. In addition there are many other time domain inversion techniques that are available such as forward and backwards linear prediction, the Levinson-Durbin algorithm, and the Schur algorithm for predicting reflection coefficients from a given autocorrelation sequence (Proakis and Manolakis, 1996).

Finally, the method offers the possible benefit of sub-bottom profiling with a single platform and without the necessity to make any additional sound, the frequency range and depth of penetration being determined by the array length and hydrophone density.

ACKNOWLEDGMENTS

The author thanks the Captain and crew of the *RV Alliance*, Enzo Michelozzi for engineering coordination, Pierro Boni for data acquisition, and particularly Peter Nielsen and Martin Siderius who allowed time for noise inversion experiments and spent many hours as Scientist-in-Charge planning respectively the MAPEXbis2000, BOUNDARY2002, and BOUNDARY2003 experiments. He would also like to thank Evan Westwood for asking the question, ''Can you deduce the phase from the noise inversion technique?'' at the end of an ASA meeting talk. Without this question and subsequent discussions with Henrik Schmidt the work would never have drifted in this direction.

[1]Note that reconstructing the true physical layer structure, as opposed to the position-dependent impulse response, is quite a different problem which is addressed by geophysicists (Claerbout, 1985).
[2]Multiple reflections are discussed later.
[3]It is easy to see that subsequent multiples are weakened in proportion to the product of the reflection coefficients of the upper and lower boundary of the layer (Jensen *et al.*, 1994, Eq. 1.53).
[4]This follows from the combination of reciprocity applied to the transmission coefficient and conservation of energy.

Ainslie, M. A. (**1995**). ''Plane-wave reflection and transmission coefficients for a three-layered elastic medium,'' J. Acoust. Soc. Am. **97**, 954–961.
Ainslie, M. A. (**1996**). ''Reflection and transmission coefficients for a layered fluid sediment overlying a uniform solid substrate,'' J. Acoust. Soc. Am. **99**, 893–902.
Bellanger, M. (**1994**). *Digital Processing of Signals* (Wiley, New York).
Brekhovskikh, L. M. (**1980**). *Waves in Layered Media* (Academic, New York).
Brekhovskikh, L., and Lysanov, Yu. (**1982**). *Fundamentals of Ocean Acoustics* (Springer-Verlag, New York).
Claerbout, J. (**1985**). *Fundamentals of Geophysical Data Processing* (Blackwell Scientific, Palo Alto); now available as free download on *Stanford Exploration Project* web site.
Desharnais, F., Thomson, D. J., and Gillard, C. A. (**2003**). ''Source depth and array tilt effects on seabed inversion of ambient noise,'' J. Acoust. Soc. Am. **113**, 2204.
Ditchburn, R. W. (**1963**). *Light* (Blackie, London).
Donelly, M. K., and Matt, G. A. (**2003**). ''Geoacoustic inversion of ambient noise: a modelling study,'' Proceedings of Underwater Defence Technology (UDT), Stockholm, June 2003.
Fiorito, R., and Uberall, H. (**1979**). ''Resonance theory of acoustic reflection and transmission through a fluid layer,'' J. Acoust. Soc. Am. **65**, 9–14.
Harrison, C. H. (**2002**). ''Geoacoustic inversion of ambient noise,'' J. Acoust. Soc. Am. **112**, 2282.
Harrison, C. H., and Baldacci, A. (**2002a**). ''Bottom reflection properties by inversion of ambient noise,'' Proc. 6th European Conference on Underwater Acoustics (ECUA), Gdansk, Poland, 24–27 June 2002.
Harrison, C. H., and Baldacci, A. (**2002b**). ''Bottom reflection properties deduced from ambient noise: Simulation of a Processing Technique,'' SACLANT Undersea Research Centre, Report SM-392.
Harrison, C. H., and Baldacci, A. (**2003**). ''Bottom reflection properties deduced from ambient noise: simulation and experiment,'' Proc. Sixth International Conference on Theoretical and Computational Acoustics (ICTCA), Hawaii, 11–15 August 2003.
Harrison, C. H., and Siderius, M. (2003). ''Effective parameters for matched field geoacoustic inversion in range-dependent environments,'' IEEE J. Ocean. Eng. Special Issue edited by N. R. Chapman, **28**(3), 432–445.
Harrison, C. H., and Simons, D. G. (**2002**). ''Geoacoustic inversion of ambient noise: a simple method,'' J. Acoust. Soc. Am. **112**, 1377–1389.
Holland, C. W., and Osler, J. (**2000**). ''High-resolution geoacoustic inversion in shallow water: a joint time- and frequency-domain technique,'' J. Acoust. Soc. Am. **107**, 1263–1279.
Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (AIP, New York).
Kramers, H. A. (**1927**). ''La difusion de la lumiere par les atomes,'' Estratto Atti Congr. Int. Fis. Como. **2**, 545–557.
Kronig, R. de L. (**1926**). ''On the theory of the dispersion of X-rays,'' J. Opt. Soc. Am. **12**, 547–557.
Marple, S. L. (**1987**). *Digital Spectral Analysis* (Prentice-Hall, Englewood Cliffs, NJ), pp. 373–378.
McDaniel, J. G. (**1999**). ''Applications of the causality condition to one-dimensional acoustic reflection problems,'' J. Acoust. Soc. Am. **105**, 2710–2716.
McDaniel, J. G., and Clarke, C. L. (**2001**). ''Interpretation and identification of minimum phase reflection coefficients,'' J. Acoust. Soc. Am. **110**, 3003–3010.
Morse, P. M., and Feshbach, H. (**1953**). *Methods of Theoretical Physics* (McGraw–Hill, New York).
Murphy, E. L., Wasiljeff, A., and Jensen, F. B. (**1976**). ''Frequency-dependent influence of the sea bottom on the near-surface sound field in shallow water,'' J. Acoust. Soc. Am. **59**, 839–845.
Osler, J., and Algan, O. (**1999**). ''A high resolution seismic sequence analysis of the Malta Plateau,'' Saclantcen Undersea Research Centre Report SR-311, May 1999.
Proakis, J. G., and Manolakis, D. G. (**1996**). *Digital Signal Processing Principles, Algorithms, and Applications* (Prentice Hall, Englewood Cliffs, NJ).
Rabiner, L. R., and Gold, B. (**1975**). *Theory and Application of Digital Signal Processing* (Prentice-Hall, London).
Robins, A. J. (**1990**). ''Reflection of plane acoustic waves from a layer of varying density,'' J. Acoust. Soc. Am. **87**, 1546–1552.
Robins, A. J. (**1994**). ''Generation of shear and compression waves in an inhomogeneous elastic medium,'' J. Acoust. Soc. Am. **96**, 1669–1676.
Robinson, E. A. (**1983**). *Digital Siesmic Inverse Methods* (Reidel, Dordrecht).
Sarkar, T. K. (**1998**). ''Generation of nonminimum phase from amplitude-only data,'' IEEE Trans. Microwave Theory Tech. **46**, 1079–1084.
Schmidt, H. (**1999**). ''OASES user's guide and reference manual,'' Dept. of Ocean Eng., M.I.T.
Simons, D. G., and Harrison, C. H. (**2004**). ''A performance assessment of genetic algorithms and differential evolution for acoustic inversion problems,'' Proc. 7th European Conference on Underwater Acoustics (ECUA), Delft, The Netherlands, 5–8 July 2004.
Tang, D. J. (**2003**). ''Inversion of sediment property using ambient noise,'' J. Acoust. Soc. Am. **113**, 2204.
Tolstoy, I., and Clay, C. S. (**1987**). *Theory and Experiment in Underwater Sound* (AIP, New York).
Trabant, P. K. (**1984**). *Applied High-resolution Geophysical Methods* (Reidel, Dordrecht).

Acoustic remote sensing of swimbladder orientation and species mix in the oreo population on the Chatham Rise

Roger F. Coombs and Richard Barr[a)]
National Institute of Water and Atmospheric Research, Private Bag 14-901, Kilbirnie, Wellington, New Zealand

(Received 24 May 2002; revised 13 June 2003; accepted 11 December 2003)

A method for combining *in situ* measurements and theoretical swimbladder-derived estimates of target strength of the deep-water fish, black and smooth oreos, is described. The technique uses Monte Carlo simulation and yields fish length–target strength relationships suitable for use in estimating biomass from echo integration acoustic surveys. The relationships are derived from estimates of the mean and standard deviation of the tilt angle distributions of the wild fish generated by the method. The relationships may also be used to estimate proportions of the two oreo species in the wild. The mean tilt angle of black oreos in the wild was about 10° with a standard deviation of 8°. For smooth oreos it was close to zero with a standard deviation of about 4°. The target strength relationships derived for biomass estimation purposes were $TS_B = -78.05 + 25.3\log_{10}(L) + 1.62\sin(0.0815L + 0.238)$ and $TS_S = -82.16 + 24.6\log_{10}(L) + 1.03\sin(0.1165L - 1.765)$, where L is the fish length and TS_B and TS_S are the target strengths of black and smooth oreos respectively. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1649998]

PACS numbers: 43.30.Sf, 43.20.Fn [WMC] Pages: 1516–1524

I. INTRODUCTION

Black oreos (*Allocyttus niger*) and smooth oreos (*Pseudocyttus maculatus*) are widely distributed in the southern half of New Zealand's EEZ. Black oreos live in depths of 600–1300 m and smooths in 650–1500 m. Both species have gas-filled swimbladders. They support a substantial fishery mainly on the southern slopes of the Chatham Rise. The fishery is a recent one, starting about 1980, and biomass has mostly been estimated from research trawl surveys and catch per unit fishing effort (Hart and McMillan, 1998). However, both species are schooling fish with the attendant difficulties that this creates in estimating abundance using trawling (McMillan *et al.* 1996). As a consequence, since 1997, acoustic surveys using echo integration (Dragesund and Olsen, 1965) to estimate biomass have also been carried out.

Echo integration requires the acoustic backscattering cross section or target strength of the fish under investigation, usually in the form of a target strength–length relationship (Foote *et al.*, 1986; Foote and Traynor, 1988), and the primary purpose of the work described here was to determine such relationships for the two oreo species.

The three principal approaches to target strength estimation are *in situ*, *ex situ*, and theoretical (Stepnowski and Moszynski, 2000). In the *in situ* method, the target strength of the fish is measured directly in its home environment in the open sea. Difficulties with this approach include identifying which echoes come from single fish targets and determining the species mix and size of those targets (Cordue *et al.*, 2001; Foote and Traynor, 1988). In the *ex situ* technique, measurements are made on captive fish, eliminating the problem of species mix. However, the orientation of the fish in an artificial environment, and hence target strength, may not be representative of their free-ranging colleagues (Nakken and Olsen, 1977; Foote, 1980b). In the case of deepwater, swimbladdered species such as oreos, which normally live below 600 m, *ex situ* methods are not practical because of the damage caused by the pressure change when they are brought to the surface.

The theoretical approach uses a model of the acoustic backscattering properties of the target species and an assumed target tilt distribution to estimate the tilt averaged target strength as a function of fish length. For fish with air-filled swimbladders, the dominant source of acoustic backscatter is the swimbladder itself (Foote, 1980a). A variety of methods have been used to determine swimbladder dimensions including measurements made directly on fish and measurements of swimbladder casts (McClatchie *et al.*, 1996). For the work described here we have used swimbladder casts. The acoustic back scattering cross section and hence target strength of the swimbladders, or the fish, can then be computed from the measurements using one of a number of possible approximate numerical methods (Clay and Horne, 1994; Foote, 1985; McClatchie *et al.*, 1996).

In this study we combine *in situ* and theoretical swimbladder modeling methods to estimate black and smooth oreo target strength and length relationships. Our overall approach is comparable to that of Foote and Traynor (1988) on walleye Pollock but we make a much more detailed least squares comparison of the *in situ* and computed target strengths. Our procedure for the latter is similar to that used by Chu *et al.* (1993) for Antarctic krill. However, we have been able to extend the method to estimate mean tilt angle and tilt angle standard deviation in addition to length–target strength relationships. Using these results we further show that it is possible to estimate the proportions of black and

[a)]Author to whom correspondence should be addressed. Electronic mail: r.barr@niwa.cri.nz

 0001-4966/2004/115(4)/1516/9/$20.00

smooth oreos in the wild population from *in situ* target strength measurements.

II. MATERIALS AND METHODS

A. Target strength estimation from swimbladder modeling

1. Swimbladder casts

Fish samples for the swimbladder modeling were collected on the south Chatham Rise during two biomass surveys using the New Zealand National Institute for Water and Atmospheric Research (NIWA) 70 m research vessel TANGAROA (voyages TAN9713, 10 November to 19 December 1997 and TAN9812, 28 September to 30 October 1998). Since samples were obtained by trawling from depths greater than 600 m, the bladders were subject to great stress from gaseous expansion when the fish were brought to the surface. Most of the smooth oreo swimbladders, which appear to have a tough swimbladder membrane and the ability to release gas relatively quickly, survived this trauma. Black oreo swimbladders were found to be more delicate and most had burst by the time they arrived at the surface.

Swimbladder casts were made from epoxy resin, with filler added to improve their durability. Swimbladders were injected while still in the fish's body cavity and dissected out when the epoxy had cured. Casts were only made from intact swimbladders that were not obviously badly distorted or stretched. To assess the amount of resin to use, a rough estimate of the fish density was made by weighing some of the fish in both air and salt water, taking care in the latter case that any trapped air bubbles were removed (A. Hart, private communication). The volume of resin used was the same as the volume of air needed to make the fish neutrally buoyant. However, despite these measures, there was a substantial degree of subjectivity in deciding how full the bladders should be, particularly for black oreos which were of poorer quality. It is also not necessarily the case that swimbladder volume is adjusted for neutral buoyancy in fish in the wild and we introduce an additional parameter to accommodate this in our fitting procedure (see Sec. II C).

The smooth oreo swimbladder casts, which were typically shaped like prolate spheroids (see lower panel of Fig. 1), were cut into slices with a diamond saw along and at right angles to their long axis. The outlines were then digitized. The slices were 2.5 mm thick where the shape changed rapidly and 5 mm thick elsewhere. Black oreo casts, which had more the shape of a short sausage [see Fig. 1(b)], were not sectioned. Instead representative measurements of diameter as a function of length were made (A. Hart, private communication). As the smooth oreo swimbladders were found to be largely circular in cross section it was decided to model both black and smooth oreo swimbladders as having cylindrical symmetry. This had the advantage of greatly reducing the complexity of the integrals needed to evaluate the swimbladders target strengths (see Sec. II A 3) while providing a realistic physical model.

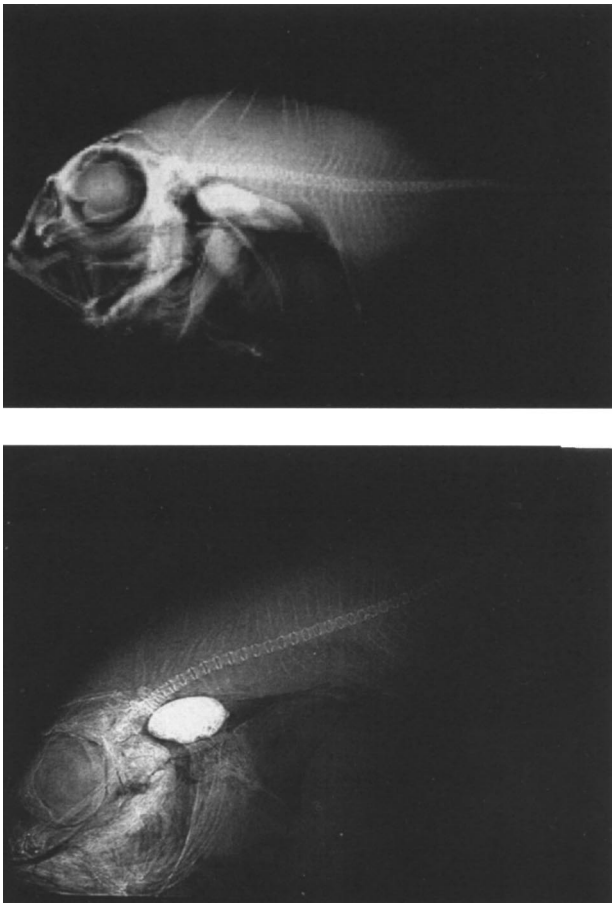


FIG. 1. Lower panel: x ray of a smooth oreo with a ''barium loaded'' swimbladder. Upper panel: x ray of a black oreo with a ''barium loaded'' swimbladder.

2. Swimbladder attitude

The swimbladder offset angle to the horizontal axis of the fish (which we define as the line joining the anterior tip of the jaw to the end of the vertebrae in the tail) was measured by filling the swimbladder with epoxy resin loaded with barium sulfate and taking x ray photographs of the whole fish (see Fig. 1). Five x rays each of black and smooth oreos were taken and these gave angles of $27.3 \pm 2.6°$ for black and $20.6 \pm 0.9°$ for smooth oreos. As sample sizes were small, these values should only be taken as a rough guide to swimbladder orientation. Nevertheless, they do show that the swimbladder offset in oreos is much greater than in more slender fishes such as Atlantic cod (Clay and Horne, 1994; Nakken and Olsen, 1977).

We consider only the scattering produced by the fish swimbladders in this paper. Consequently, the estimates of tilt angles we derive for free-swimming fish in the wild are for the swimbladders, not the fish. It is necessary to adjust the swimbladder offset angle to obtain the tilt of the fish about its horizontal axis. We assume that the tilt angle distribution is normal and this is supported by field observations on a range of fish (see, for example, Olsen, 1971; Coombs and Cordue, 1995).

3. Swimbladder modeling

Casts were made of 32 black oreo swimbladders covering a range of fish total length of 25.7–36.8 cm and 27 smooth oreos covering a length range of 22.0–54.0 cm.

The swimbladder measurements for both species were used to construct equivalent cylindrically symmetric spheroids. The scattering from these was estimated using the Kirchhoff approximation (Clay and Medwin, 1977; McClatchie *et al.*, 1996). By using swimbladder models with cylindrical symmetry, the time for numerically evaluating the surface integral was reduced considerably compared to the fully general case.

B. *In situ* target strength estimation

1. Acoustic system

A split beam version of NIWA's Computerized Research Echosounder Technology (*CREST*) (Coombs, 1994) was used to collect the *in situ* target strength data. *CREST* is computer based, using the concept of a ''software echo sounder.'' The transmitter was a switching type with a nominal power output of 80 W rms, an operating frequency of 38 kHz and a transmit pulse length of 0.32 ms (12 cycles). Time between transmits was 1.4 s. All four transducer quadrants (beams) were energized simultaneously from the transmitter but on receive the system operated as four semi-independent echosounders. Each receiver channel had dual broadband, wide dynamic range preamplifiers and dual serial analog-to-digital converters (ADCs) feeding a digital signal processor (DSP56002). Analog Devices AD7723 ADCs were used, operating in ''band-pass'' mode at a conversion rate of 62.5 kHz (equivalent to a rate of 125 kHz in a conventional ADC). The ADC data were complex demodulated, filtered, and decimated to 15.625 kHz. The filter was a 100-tap, linear-phase finite impulse response digital filter with a 3-dB bandwidth of 3.0 kHz. After filtering and decimation a $40 \log R$ time varied gain was applied and the results shifted to give 16-bit resolution in both the real and imaginary terms. These complex data were stored for later processing. The transducer used was a Simrad model ES38DD split-beam transducer with a depth rating of 1500 m. The combined source level and transducer receive response (SL +SRT) was 48.8 dB *re* 1 V and the receiver gain at 1 m was 60.4 dB.

The split-beam transducer and receiver electronics were mounted in an open frame towed body deployed on a 2-km tow cable at depths of between 900 and 1000 m. Data were collected while drifting or steaming at very slow speeds (<2 knots) in ''drops'' typically lasting about an hour. Acoustic data from each drop were stored in computer files identified by a ''d'' followed by a number. The system was ordinarily used with a 38.1 mm±2.5-μm-diam tungsten carbide calibration sphere suspended about 20 m below the transducer. The sphere tended to swing in the beam and was often not visible so sphere data from all acoustic drops were aggregated to estimate the SL+SRT quoted above.

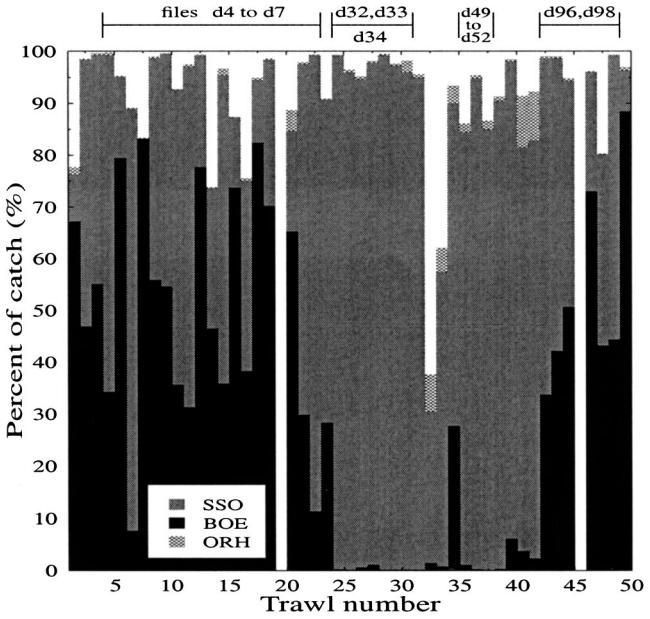


FIG. 2. Percentage of black oreo (BOE), smooth oreo (SSO), and orange roughy (ORH) in trawl catches.

2. In situ data processing

The *in situ* acoustic data were collected during TANGAROA voyage TAN0011 (23 October to 5 November 2000). Echosounder ''marks'' that were expected to be either black or smooth oreos were located and the towed transducer deployed 50–100 m above the mark for 1–2 h with the aim of maximizing the number of echoes captured from individual fish.

After collecting the acoustic data, the marks were trawled to identify the fish and estimate their size composition. For the latter as many fish as possible were measured, where practical the whole catch. Fish were measured to the millimeter below using NIWA's electronic fish measuring system. We carried out 49 trawls and the main species caught were black and smooth oreos and orange roughy. The proportions of these species are shown in Fig. 2. The numbers along the upper *x*-axis of the figure give the acoustic data file numbers associated with the trawls numbered on the lower *x*-axis.

The acoustic data were first filtered to remove echoes not originating from individual fish. Echoes were initially identified by locating peaks in the combined beam signal. The measurements from individual quadrants of the transducer were combined in pairs to calculate the position of the selected echoes in the beam (Ehrenberg, 1979) and the amplitude corrected for the position in the beam. The maximum amplitude of each echo was then estimated by fitting a quadratic to the three values bracketing the peak of the echo. The maximum of this quadratic was taken as the target strength value for subsequent analysis. Selection of echoes from individual fish was based on the criteria given in Barr *et al.* (2000).

3. In situ data selection

In addition to being selective about the individual targets used, we also selected the data sets gathered from different

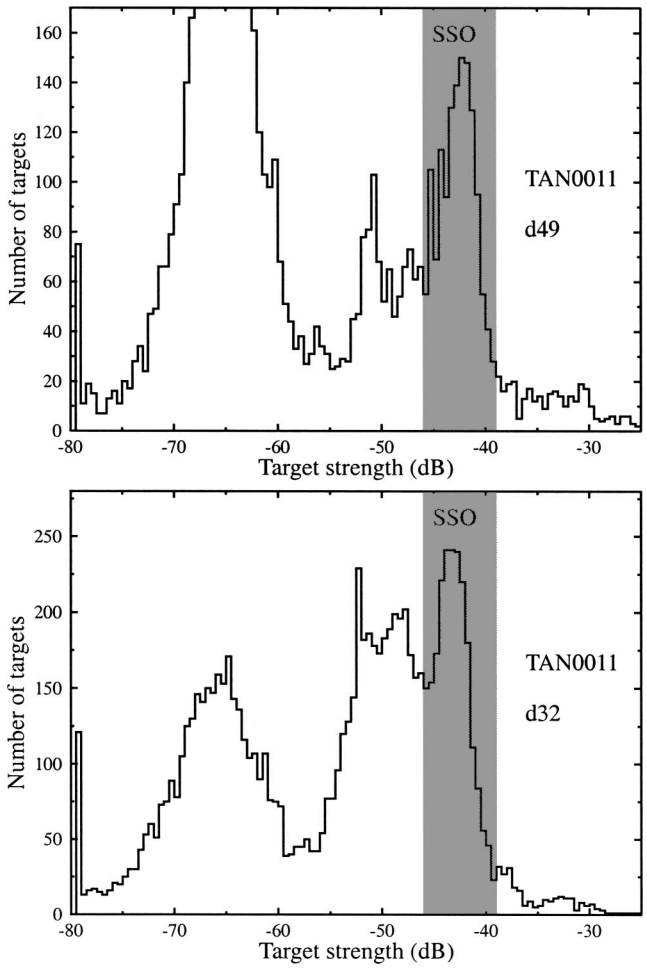


FIG. 3. *In situ* target strength distributions for smooth oreo (SSO) data files. Lower panel: file d32. Upper panel: file d49.

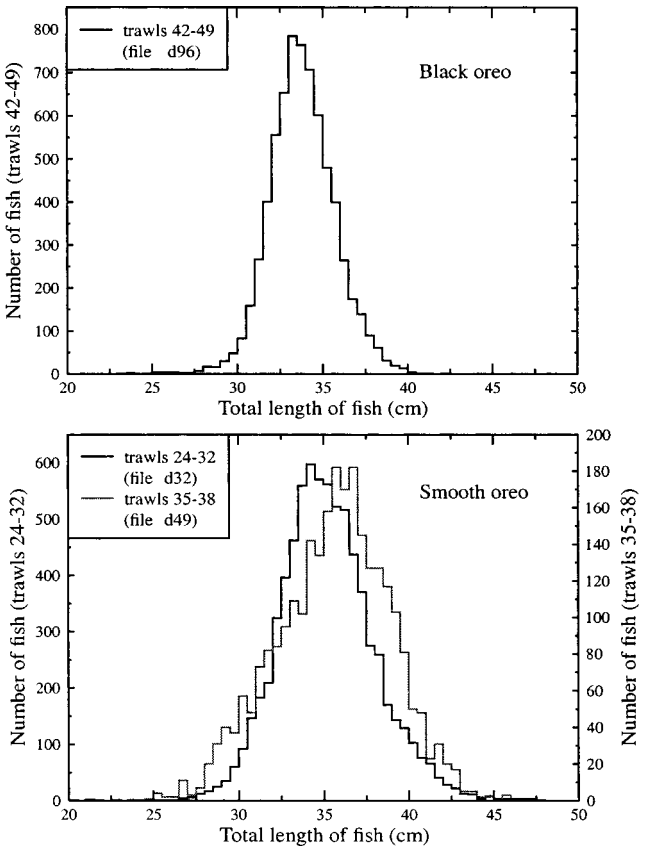


FIG. 4. Fish length data. Lower panel: Trawl data associated with smooth oreo acoustic files d32 and d49. Upper panel: Trawl data associated with black oreo acoustic file d96.

fish marks and only used those where the associated trawl catches were strongly dominated by one or other oreo species. Smooth oreos formed more than 90% of the trawl catches associated with drop data files d49, d50, and d52 and for files d32, d33, and d34 more than 95%. The *in situ* target strength distributions for files d32 and d49 are shown in the lower and upper panels of Fig. 3, respectively.

By using both target amplitude and phase (see Barr *et al.*, 2000) and the information about the species mix gained from trawling, we can tentatively identify the targets generating the four dominant modes in the target strength distribution shown in the lower panel of Fig. 3. The sharp peak near −53.5 dB is thought to be produced by myctophids and the broader peak near −49 dB by orange roughy, the latter peak being better characterized in the phase data (see Barr and Coombs, 2001). The very broad mode near −67 dB we ascribe to bathypelagic zooplankton. This leaves the peak near −44 dB which we attribute to smooth oreos. In the data for file d49, shown in the upper panel of Fig. 3, only three peaks are evident: an enhanced peak near −67 dB, ascribed to bathypelagic zooplankton, a narrow peak near −52 dB due to myctophids and a broader peak near −43.5 dB produced by smooth oreos.

The light gray shaded regions in Figs. 3 and 5 cover the range of target strengths that we attribute to smooth oreos and that are largely uncontaminated by larger fish with air-filled swimbladders or by orange roughy. We have used this information in the Monte Carlo simulations and the *in situ* data were only evaluated over this region. Thus our technique is an advance on directly ascribing a modal peak in target strength to modal peak in fish length, yet it is more constrained than the general search technique suggested by Cordue *et al.* (2001).

From the trawls made in the same area as the acoustic drops, we obtained the smooth oreo length distributions presented in the lower panel of Fig. 4. It can be seen that the trawls associated with file d49 produced slightly larger fish than the trawls associated with file d32. This length difference may explain why the peak smooth oreo target strength shown in the upper panel of Fig. 3 is slightly higher than in the lower panel of the same figure.

As can be seen from the trawl catch data in Fig. 2, there were no trawl areas where the catch was >90% black oreos. However, trawls 46 and 49 were 73% and 88% black oreos, respectively, and the acoustic drops in the same area (d96 and d98) were therefore chosen to give *in situ* target strength data strongly dominated by black oreos. The *in situ* target strength distribution for file d96 is shown in Fig. 5 and can be compared with the smooth oreo file, d32, in the lower panel of Fig. 3. Whereas there are very few smooth oreo targets stronger than −40 dB, it can be seen that black oreo targets extend in range up to −33 dB. The dark gray region in Fig. 5 covers the range of target strength data we assume

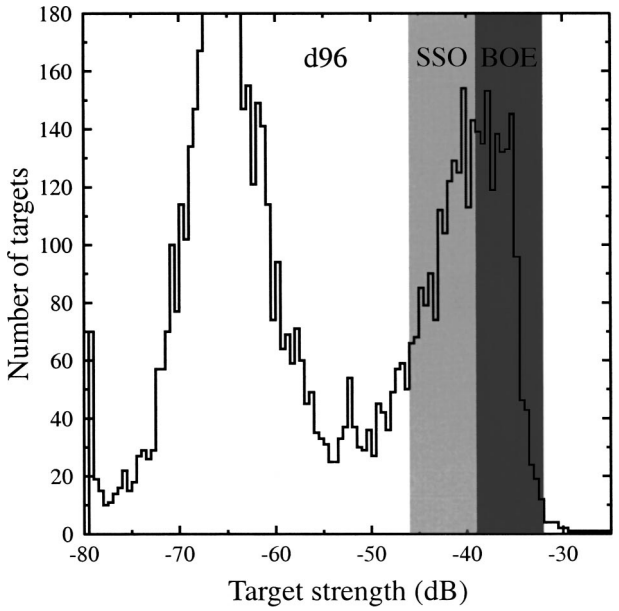


FIG. 5. *In situ* target strength data for file d96.

to be dominated by black oreos. As for smooth oreos, only these data were used in the Monte Carlo simulation. The black oreo length distribution from trawls 42–49 is shown in the upper panel of Fig. 4.

C. Monte Carlo analysis of combined data

Three experimental data sets were used in the Monte Carlo modeling, swimbladder cast data, fish length distribution data, and *in situ* target strength data. The swimbladder measurements were combined with the length data and an assumed fish tilt angle distribution to generate a target strength distribution. To do this, a simulated population of fish with randomly chosen lengths matching the observed length distribution was created. Each fish was given a randomly chosen tilt angle such that the overall tilt angle distribution matched a chosen distribution. Each fish was also given a randomly chosen swimbladder, selected from the dataset of measured swimbladders, appropriately scaled for the fish length. The target strength of each fish was then estimated using the Kirchhoff procedure outlined in Sec. II A 3 above. As described in Sec. II A 1, an additional scale shift parameter was used in the target strength calculations to allow for the possibility that the swimbladder cast volume was unrepresentative of fish in the wild (and other measurement errors).

With this procedure, the tilt angle distribution and the ''true'' swimbladder volume were the only elements for which we did not have direct experimental data. The tilt angle mean and standard deviation were varied and the sum of the squares of the deviations (least squares error) between computed and *in situ* target strength measurements estimated for each distribution. Both the mean tilt angle and the tilt angle standard deviation were varied from 0° to 30° in 1° steps giving 900 different fish distributions. Errors in estimating the bladder volume and small errors in equipment calibration were accommodated, after the target strength distribution had been calculated, by shifting the whole distribution up and down in the range ±2 dB with a 0.1-dB resolution.

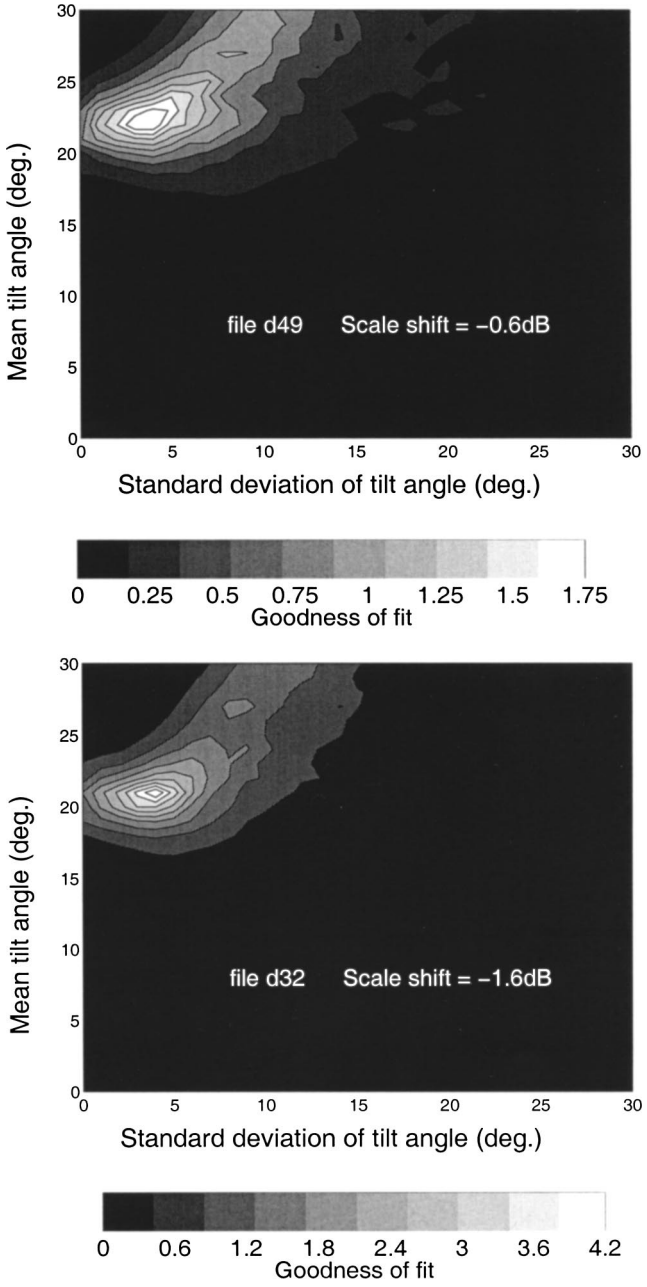


FIG. 6. Goodness of fit contours plotted in tilt angle versus tilt angle standard deviation space for smooth oreos. Lower panel: file d32. Upper panel: file d49.

We have evaluated the agreement between *in situ* target strength distributions and those computed from swimbladder models as a ''goodness of fit.'' This is calculated as the reciprocal of the square root of the sum of the squares of the differences between measured and computed target strength.

III. RESULTS

The results for smooth oreos are summarized in Fig. 6 in which the goodness of fit surface for data file d32 is shown in the upper panel, and for d49 in the lower. It can be seen that both datasets yield single, clear goodness of fit peaks at

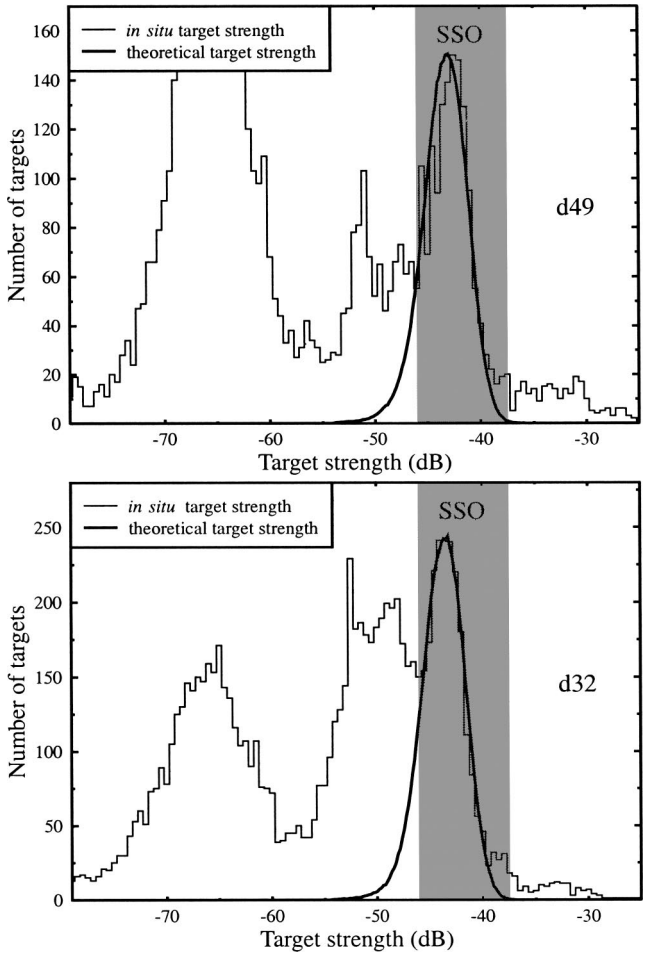


FIG. 7. Comparison of theoretical target strength estimates derived from modeling with *in situ* estimates for smooth oreo (SSO) data files. Lower panel: file d32. Upper panel: file d49.

very similar mean tilt angles. These were 20.8° for d32 and 22.4° for d49 with tilt angle standard deviations of 3.9° and 3.7°, respectively. The associated scale shifts for best fit were −1.6 dB in the computed target strength for file d32 and −0.6 dB for d49. Figure 7 shows the good agreement between *in situ* and modeled target strength distributions using these parameters.

Black oreo goodness of fit results, file d96, are shown in Fig. 8. Again there is a clear maximum region, in this case centered on a mean tilt angle of 17° with a standard deviation of 8°. The associated scale shift was −1.3 dB. The minor peaks in goodness of fit on either side of the main peak are thought to result from the statistical noise at the top of the mode of the black oreo *in situ* target strength distribution (see Fig. 5). Figure 9 shows the agreement between *in situ* and modeled target strength distributions using these parameters, which is again excellent, especially since the curves are constrained to fit in the gray shaded region. The good fit over the target strength range from −30 to −50 dB suggests that there were few smooth oreos present in the region surveyed.

The mean tilt angle, tilt angle standard deviation, and shift values required to best fit the *in situ* target strength data were used, in a Monte Carlo simulation, to derive target strength–length distributions describing smooth and black

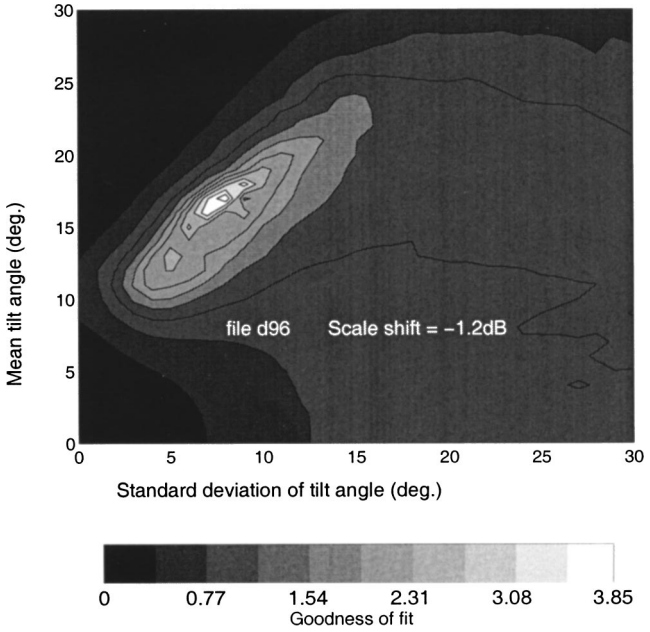


FIG. 8. Goodness of fit contours plotted in tilt angle versus tilt angle standard deviation space for black oreo data file d96.

oreos. These distributions are presented in Fig. 10 where the small circles represent the means of the distributions at each length in the simulation. The two lower plots, for smooth oreo data, can be seen to be very similar to each other and are displaced about 5 dB below the plot for black oreos. This 5 dB displacement appears to be almost independent of fish length for the oreos typically caught on the south Chatham Rise.

The solid lines in Fig. 10 are simple functional fits to the simulated points which were used for biomass estimation (Doonan *et al.*, 2003). For black oreos the target strength, TS, of a fish of length, L, is given by

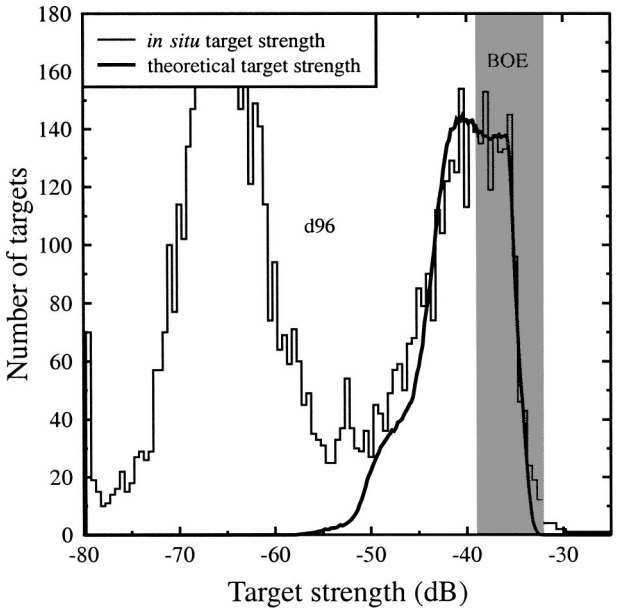


FIG. 9. Comparison of theoretical target strength estimates derived from modeling with *in situ* estimates for black oreo (BOE) data file d96.

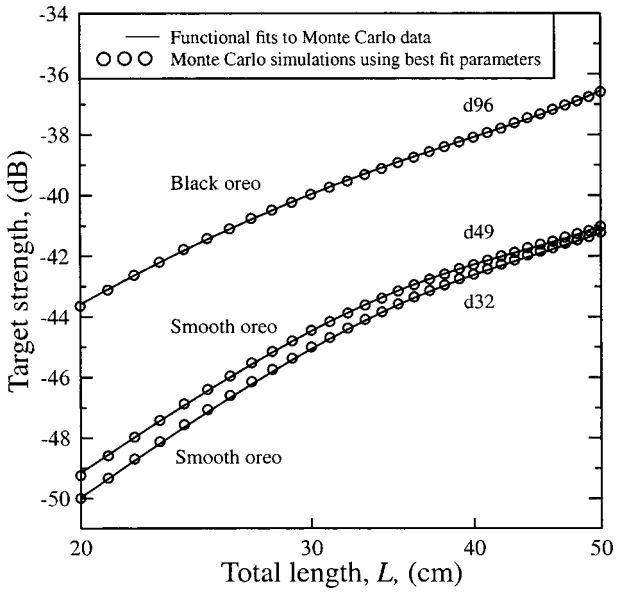


FIG. 10. Tilt averaged target strength data plotted as a function of fish total length for black and smooth oreo.

$$TS = -78.05 + 25.3 \log_{10}(L) + 1.62 \sin(0.0815L + 0.238). \quad (1)$$

For smooth oreo data files d32 and d49 the functions are of the form

$$TS = -83.78 + 25.5 \log_{10}(L) + 1.068 \sin(0.111L - 1.59), \quad (2)$$

and

$$TS = -80.54 + 23.76 \log_{10}(L) + 0.987 \sin(0.122L - 1.94), \quad (3)$$

respectively.

Combining the points in Eqs. (2) and (3) gives the relationship used for smooth oreo biomass estimation:

$$TS = -82.16 + 24.6 \log_{10}(L) + 1.03 \sin(0.1165L - 1.765). \quad (4)$$

These formulations were chosen by inspection of the residuals which resulted when fitting the data to simple logarithmic plots. They are not based on any insight into the physics of the scattering process. They do, however, provide extremely good fits to the computations.

IV. DISCUSSION

In applying our method we have made a number of assumptions. First, we have neglected changes in air density in the swimbladder with depth. Second, we have neglected the scattering effects of the body of the fish. Last, to allow for the likelihood that the swimbladder casts differ in volume from those of wild fish, we have applied a variable scaling shift to the theoretical target strength values to obtain an overall best fit between theory and experiment.

With respect to the last point, all computed target strengths have been reduced to best fit the *in situ* data, on average by about 1.1 dB. Although our rationale for this shift is a mismatch in swimbladder volumes, calibration errors could provide an alternative explanation. However, we think this unlikely. As described in Sec. II B 1, a standard sphere was deployed with the system and used to derive the system calibration. The latter is closely similar (well within ± 0.2 dB) to other calibrations carried out on other surveys at several different locations.

A weakness of our simple computer models (the first point above) was the lack of any allowance for the higher air density in the swimbladder of a wild fish which would reduce the density contrast with the fish body and hence reflectivity. A simple calculation based on theory in Medwin and Clay (1998) indicates this approximation resulted in an overestimate of target strength by ~ 0.45 dB for fish at depths of 1000 m. This reduces the difference between the *in situ* data and modeling to only 0.7 dB, to be accounted for by fish body effects and swimbladder volume mismatch, yielding excellent agreement between theory and experiment.

Although in our fitting procedure we make no assumptions about what the swimbladder tilt angle might be, our expectation was that, given their body form, oreos are oriented in the wild with their nose-tip-to-tail axis horizontal. With this in mind it can be seen that the estimated *in situ* mean swimbladder tilt angles for smooth oreos of 20.8° and 22.4° are closely similar to the swimbladder offset angle of $20.6 \pm 0.9°$ from the radiographs implying a zero tilt angle for the fish itself. However, the mean *in situ* tilt angle of 17° for black oreos is not nearly such a good fit to the respective radiograph value of $27.3 \pm 2.6°$, suggesting a nose-up attitude of about 10°. More work is needed in this area and there are several possible causes of the difference. First and foremost, perhaps, is that black oreos may naturally swim nose-up. Distortions in the fish body or swimbladder are another possibility and most black oreo swimbladders burst on being brought to the surface. Trawl catches always contained a significant proportion of smooth oreos, so another possibility is contamination of the black oreo *in situ* target strength data. A last possibility is that the scale shift yielding the maximum goodness of fit was actually too small, another reflection on the poor quality of the swimbladder casts for black oreos.

Changing the scale shift factor is not a completely independent operation to changing the swimbladder tilt when it comes to evaluating a change in the target strength. To show this effect we have fixed the smooth oreo swimbladder tilt angle standard deviation at 4° and then allowed mean tilt angle and scale shift to vary together to best fit the *in situ* data for file d49. Figure 11 (lower panel) shows how the goodness of fit maximizes when the scale shift is -0.6 dB, but the goodness of fit changes only slowly with the shift. Also plotted in the same figure is the mean tilt angle as a function of the scale shift required to keep the goodness of fit maximized. It can be seen that the two are almost equivalent in their effect with a 1-dB scale shift being equivalent to a 2.9° change in mean tilt angle. Since Macaulay *et al.* (2001) estimate target strength uncertainties due to swimbladder volume mismatches to be a maximum of about 1 dB, this translates to an uncertainty of 2.9° in our mean swimbladder orientation estimates. The upper panel of Fig. 11 shows a

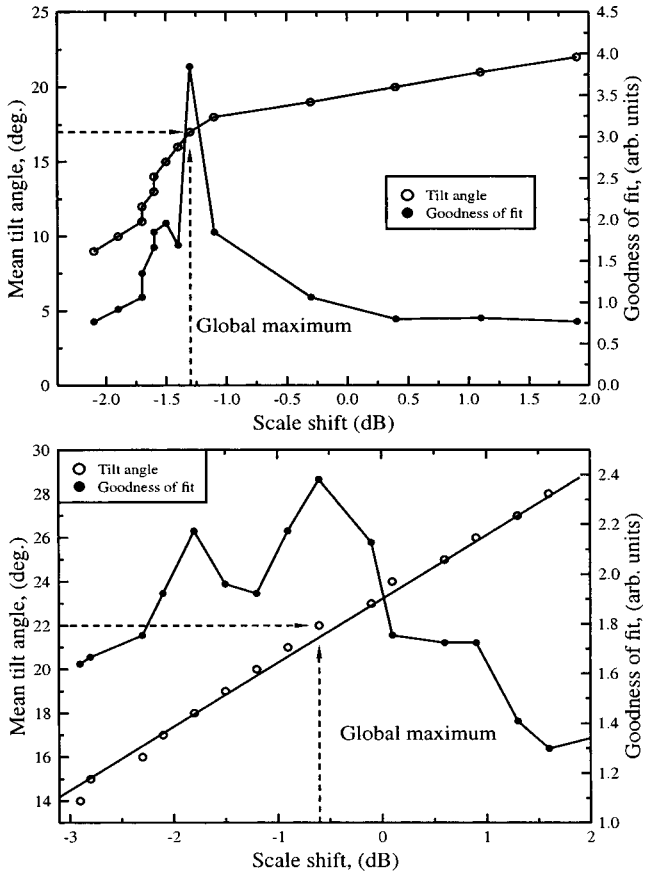


FIG. 11. Mean tilt angle and goodness of fit plotted as a function of the scale shift. Lower panel: file d49. Upper panel: file d96.

similar plot for the black oreo data file d96. In this case the goodness of fit shows a clear peak at a scaling offset of -1.3 dB and a tilt angle of 17°. However, were the scale shift to be >0 dB, then a tilt angle of $>20°$ would be more appropriate and more in line with the radiograph data.

Although the interdependence of scaling offset and tilt angle is annoying when one is trying to make deductions about fish biology and fish behavior, it is of only minor importance to biomass estimation. What is important here is that we have produced a mathematical fish model which accurately describes the *in situ* target strength data. As we have noted earlier, our relationships are entirely empirical and we do not ascribe any physical or biological significance to the formulation. In general we believe that relationships used for biomass estimation should reflect the data and for this reason do not use the "$20\log_{10}L$" formulation for this purpose (see McClatchie *et al.* 2003). However, to allow comparison with other results, the black oreo relationship in this form is

$$TS = -69.88 + 20\log_{10}(L) \tag{5}$$

and the smooth is

$$TS = -74.72 + 20\log_{10}(L). \tag{6}$$

Having derived tilt angle distributions and scale shift factors appropriate for free swimming black and smooth oreos from areas with a strong preponderance of one or the other, we are now in the position to be able to apply our Monte Carlo modeling technique to finding the proportions

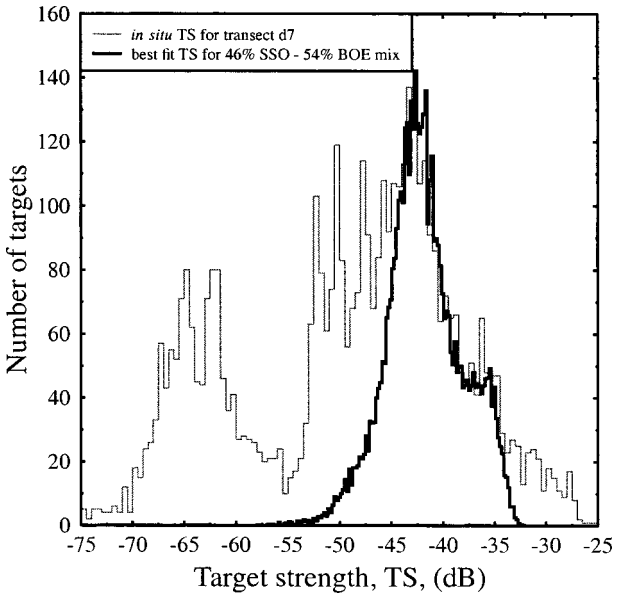


FIG. 12. Comparison of target strength data derived from modeling (best fit) with target strength data measured *in situ*, data file d7. The best fit oreo model assumes a species mix of 46% smooth oreo (SSO) and 54% black oreo (BOE).

of black and smooth oreos in regions where the trawl data indicated a significant mixture of the two (see Fig. 2), solely from the *in situ* acoustic data collected. The thin gray trace in Fig. 12 shows the *in situ* target strength distribution for file d7, recorded in the same region as trawls 11, 13, 15, 20, and 21 shown in Fig. 2. The trawl data suggest a ratio of about 50:50% blacks to smooths but with significant spatial variability. The *in situ* data were fitted by allowing the species mix (black:smooth) to change from 0:100% to 100:0% in 1% steps. The thick black trace in Fig. 12 shows the theoretical target strength distribution which best fitted the *in situ* data. It was produced by a mix of 46% smooth and 54% black oreos. The goodness of fit plotted against the species mix is presented in Fig. 13. Also shown is the mean of the individual trawl catch percentages and the mean percentage mix derived from the total catch of all the trawls in the area. The value derived from the *in situ* acoustic data can be seen to be in good agreement with the trawl catch data.

V. CONCLUSIONS

Simple acoustic backscatter models of fish, based on measurements of swimbladder casts, have been used to compute target strength distributions of black and smooth oreos for direct comparison with target strength distributions measured *in situ*. Only three parameters were used in the computer model and these were mean tilt angle, tilt angle standard deviation, and an amplitude scale shift factor. The latter was necessary to allow for mismatches between the volumes of swimbladder casts and the swimbladders of wild, free-swimming oreos. The *in situ* and computed target strength distributions were compared with a goodness of fit statistic which showed clear, single, isolated maxima.

The agreement between measured and computed target strength distributions is excellent and the values of mean swimbladder tilt angle derived from this fitting process are

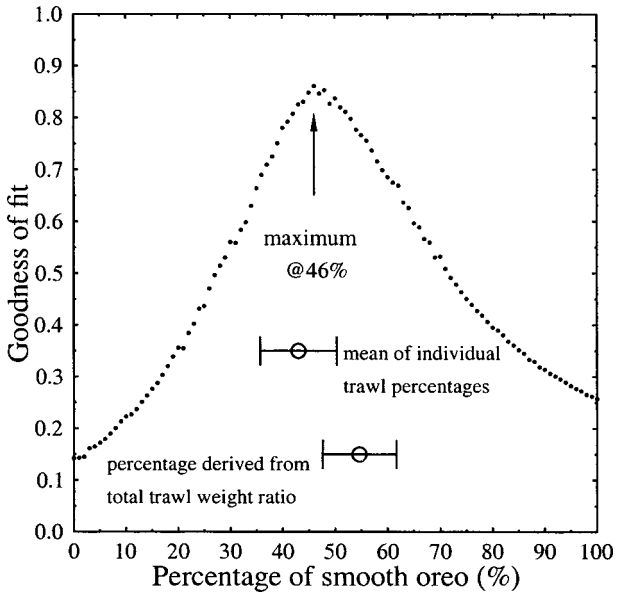


FIG. 13. Goodness of fit (of modeled to *in situ* target strength) plotted as a function of the percentage of smooth oreo in the mix.

closely similar to the tilt of the swimbladder in the fish measured from x rays, particularly for smooth oreos. In regions with a mix of black and smooth oreos our acoustic modeling technique was able to estimate the proportions of the two species, producing values similar to those obtained by trawling.

There may be some uncertainties in the tilt angles as a consequence of mismatches between the volume of swimbladder casts and wild fish swimbladders using this method, especially for black oreos. However, we believe the target strength–length relationships we have derived by Monte Carlo modeling are a significant step forward over the more basic technique of matching modes in the target strength and length distributions.

ACKNOWLEDGMENTS

We would like to thank Alan Hart and Paul Grimes for their effort in producing swimbladder casts of most of New Zealand's deepwater fishes. We thank Adam Dunford for his internal review of this manuscript. Funding for this research was provided by the New Zealand Ministry of Fisheries under Project No. OEO2000/01A.

Barr, R., and Coombs, R. F. (**2001**). ''*In situ* target strength measurements and chirp responses of orange roughy (*Hoplostethus atlanticus*),'' NIWA, Wellington, Final Research Report for Ministry of Fisheries Research Project ORH1999/01A, Objective 1, p. 45.

Barr, R., Coombs, R. F., and Macaulay, G. J. (**2000**). ''Can we discriminate between different deepwater fishes using a standard acoustics target strength ping?'' in *Proceedings of the 10th International Symposium on Acoustic Remote Sensing of the Atmosphere and Oceans and Associated Techniques*, Auckland, New Zealand.

Chu, D., Foote, K. G., and Stanton, T. K. (**1993**). ''Further analysis of target strength measurements of Antarctic krill at 38 and 120 kHz: Comparison with deformed cylinder model and inference of orientation distribution,'' J. Acoust. Soc. Am. **93**, 2985–2988.

Clay, C. S., and Horne, J. K. (**1994**). ''Acoustic models of fish: The Atlantic cod (*Gadus morhua*),'' J. Acoust. Soc. Am. **96**, 1661–1668.

Clay, C. S., and Medwin, H. (**1977**). *Acoustical Oceanography* (Wiley, New York).

Coombs, R. F. (**1994**). ''An adaptable acoustic data acquisition system for fish stock assessment,'' in *International Conference on Underwater Acoustics*, 5–7 December, 1994, Australian Acoustical Society, Univ. of New South Wales, Sydney, pp. 18–22.

Coombs, R. F., and Cordue, P. L. (**1995**). ''Evolution of a stock assessment tool: acoustic surveys fo spawning hoki (*Macruronus novaezelandiae*) off the west coast of South Island New Zealand, 1985–91,'' N. Z. J. Mar. Freshwat. Res. **29**, 175–194.

Cordue, P. L., Coombs, R. F., and Macaulay, G. J. (**2001**). ''A least squares method of estimating length to target strength relationships from *in situ* target strength distributions and length frequencies,'' J. Acoust. Soc. Am. **109**, 155–163.

Doonan, I. J., McMillan, P. J., Hart, A. C., and Coombs, R. F. (**2003**). ''Smooth oreo abundance estimates from the October–November 2001 survey of the south Chatham Rise (OEO 4),'' *New Zealand Fisheries Assessment Report 2003/26*.

Dragesund, O., and Olsen, S. (**1965**). ''On the possibility of estimating year-class strength by measuring echo-abundance of 0-group fish,'' Fiskeridir. Skr., Ser. Havunders. **13**, 47–75.

Ehrenberg, J. E. (**1979**). ''A comparative analysis of *in situ* methods for directly measuring the acoustic target strength of individual fish,'' IEEE J. Ocean. Eng. **OE-4**, 141–152.

Foote, K. G. (**1980a**). ''Importance of the swimbladder in acoustic scattering by fish: A comparison of gadoid and mackerel target strengths,'' J. Acoust. Soc. Am. **67**, 2084–2089.

Foote, K. G. (**1980b**). ''Effect of fish behaviour on echo energy: the need for measurements of orientation distributions,'' J. Cons., Cons. Int. Explor. Mer **39**, 193–201.

Foote, K. G. (**1985**). ''Rather high frequency sound scattering by swimbladdered fish,'' J. Acoust. Soc. Am. **78**, 688–700.

Foote, K. G., and Traynor, J. J. (**1988**). ''Comparison of walleye pollock target strength estimates determined from *in situ* measurements and calculations based on swimbladder form,'' J. Acoust. Soc. Am. **83**, 9–17.

Foote, K. G., Aglen, A., and Nakken, O. (**1986**). ''Measurement of fish target strength with a split-beam echo sounder,'' J. Acoust. Soc. Am. **80**, 612–621.

Hart, A. C., and McMillan, P. J. (**1998**). ''Trawl survey of oreos and orange roughy on the south Chatham Rise, October–November 1995 (TAN9511),'' NIWA Technical Report 27.

Macaulay, G. J., Hart, A. C., and Grimes, P. (**2001**). ''Estimation of the target strength of orange roughy by-catch species,'' NIWA, Wellington, Final Research Report for Ministry of Fisheries Research Project ORH1999/01A, Objective 2, p. 11.

McClatchie, S., Macaulay, G. J., and Coombs, R. F. (**2003**). ''A requiem for the use of 20 log *Length* for acoustic target strength with special reference to deep sea fishes,'' ICES J. Mar. Sci. **60**, 419–428.

McClatchie, S., Alsop, J., Ye, Z., and Coombs, R. F. (**1996**). ''Consequence of swimbladder model choice and fish orientation to target strength of three New Zealand fish species,'' ICES J. Mar. Sci. **53**, 847–862.

McMillan, P. J., Doonan, I. J., Coburn, R. P., and Hart, A. C. (**1996**). ''Is the south Chatham Rise trawl survey series providing an index of smooth oreo abundance in OEO 4,'' New Zealand Fisheries Assessment Research Document 96/16, p. 18.

Medwin, H., and Clay, C. S. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, London).

Nakken, O., and Olsen, K. (**1977**). ''Target strength measurements of fish,'' Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **170**, 52–69.

Olsen, K. (**1971**). ''Orientation measurements of cod in Lofoten obtained by underwater photography and their relation to target strength,'' ICES CM 1971/B:17, (mimeo).

Stepnowski, A., and Moszynski, M. (**2000**). ''Inverse problem solution techniques as applied to indirect *in situ* estimation of fish target strength,'' J. Acoust. Soc. Am. **107**, 2554–2562.

Echo-to-reverberation enhancement using a time reversal mirror

S. Kim,[a] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. Edelmann, and T. Akal[b]
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 9 May 2003; revised 19 December 2003; accepted 30 December 2003)

Reverberation from rough ocean boundaries often degrades the performance of active sonar systems in the ocean. The focusing capability of the time-reversal method provides a new approach to this problem. A time-reversal mirror (TRM) focuses acoustic energy on a target enhancing the target echo while shadowing the boundaries below and above the focus in a waveguide, thereby reducing reverberation. The resulting echo-to-reverberation enhancement has been demonstrated experimentally using a time-reversal mirror in the 3–4 kHz band in shallow water. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1649737]

PACS numbers: 43.30.Vh, 43.30.Gv, 43.30.Hw [WLS] Pages: 1525–1531

I. INTRODUCTION

A time reversal mirror (TRM)[1,2] retransmits a received signal in time-reversed order (backpropagation) resulting in a focus at the origin of the signal. Focusing can be achieved effectively with a vertical source–receiver array (SRA) in a waveguide,[3–6] where most acoustic energy is confined within the waveguide by reflection from the boundaries and refraction from the inhomogeneous sound-speed environment. In a range-independent environment, a vertical SRA generates an azimuthally symmetric focal structure.

The focusing capability of a TRM suggests potential applications to the active detection problem in the ocean where backscattering from rough ocean boundaries often masks a weak target echo and degrades detection performance. In a waveguide, a TRM can generate a focused acoustic field at the origin of a probe source and, as we demonstrate, less energy at the rough ocean boundaries below and above the focus, referred to as shadowing in this paper. This property can be used to enhance the echo-to-reverberation ratio in the returning backscattered field, resulting in improved detection performance.

For the purpose of studying TRM echo-to-reverberation enhancement, we will assume that a probe source near the target is available that facilitates retransmission of the initial signal from the TRM to generate a focus on the target. In practice, it is unlikely that the target actually will be in the immediate vicinity of a probe source. However, two approaches to focusing on the target can be implemented. First, we previously demonstrated the ability to shift the focus in range through an easily implemented frequency shifting procedure.[7] Thus, it is not required that the probe source actually be located at the focal position. Second, a surrogate probe source could be derived from the monostatic echo plus reverberation return from an initial conventional active transmission (e.g., a simple broadside ping). A selected region of that return can be time reversed and retransmitted as part of an iterative focusing procedure.[8]

This paper is organized as follows. In Sec. II, we use numerical simulations to demonstrate time-reversal focusing and shadowing and the resulting reduced reverberation. In Sec. III, we present experimental results obtained in shallow water north of Elba Island, Italy, demonstrating the echo-to-reverberation enhancement.

II. NUMERICAL SIMULATIONS

We have reported in our ocean TRM experiments[4,9,10] that the acoustic field above and below the focus is typically 15–20 dB less than the field at the focus at 450 and 3500 Hz. We consider an environment similar to that of the experimental results that will be presented in Sec. III. To calculate the reverberation from a rough interface in a waveguide, a

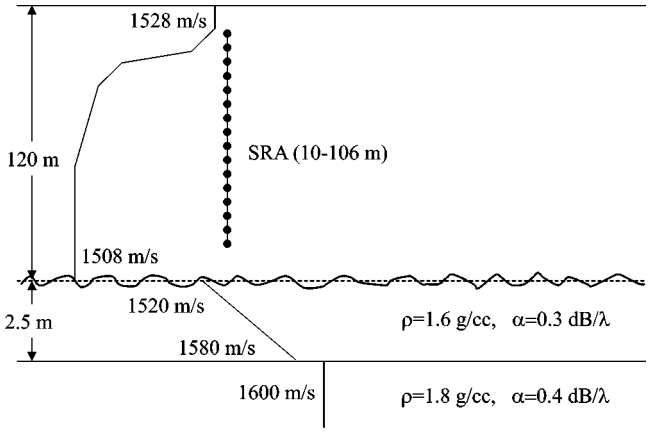


FIG. 1. Ocean environmental model used for the backscattered field simulations. The SRA spans the water column from 10 to 106 m with 3 m element spacing in a 120 m water depth. The bottom roughness has 0.1 m rms height and 15 m correlation length with a Goff–Jordan power law spectrum. The sound-speed profile shows a typical downward refracting environment with the thermocline spanning 20–50 m. The sound speeds are 1528 m/s at the surface and 1508 m/s at the ocean bottom.

[a]S. Kim is presently with the Agency for Defense Development, Chinhae, 645-600, South Korea. Electronic mail: sikim@add.re.kr. The work was performed while the author was at the Marine Physical Laboratory.
[b]T. Akal was with NATO SACLANT Undersea Research Center, 19138 La Spezia, Italy. Current address: TUBITAK-MAN, Marmara Research Center, Earth and Marine Sciences Research Institute, P.K.21 Gebze, Kocaeli 41470, Turkey. Electronic mail: tuakal@yahoo.com

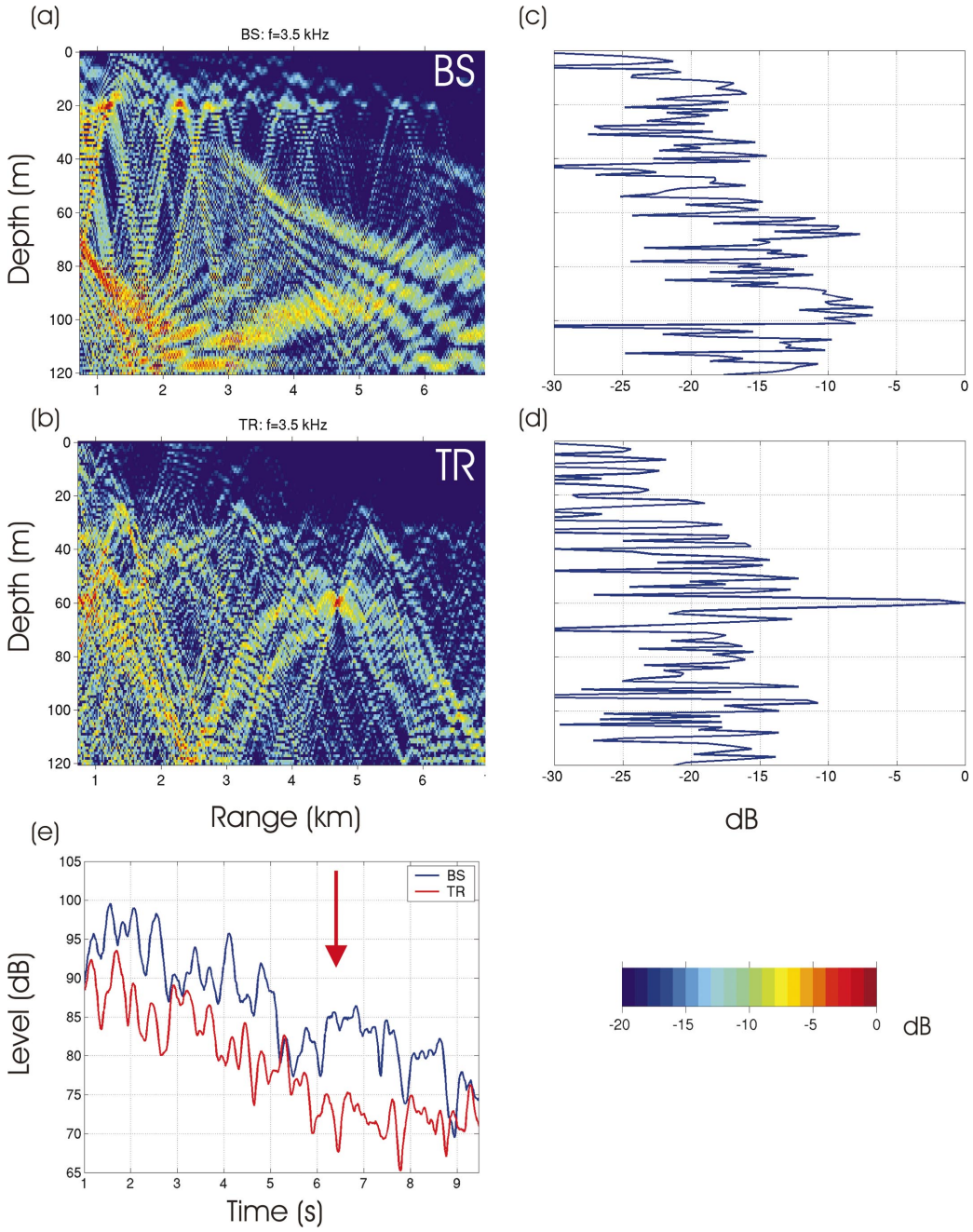


FIG. 2. Normalized transmission loss at a center frequency of 3.5 kHz: (a) broadside (BS) and (b) time-reversal (TR) focusing with the PS at 4.7 km range and 60 m depth. The geometrical spreading term $(1/r)$ is removed and the dynamic range is 20 dB. The plots are normalized by the largest values in their depth/range extent. (c) and (d) are the vertical slices of (a) and (d) at the PS range of 4.7 km. The curves in (e) show the reverberation levels using a 100 ms Gaussian-shaped pulse: BS (blue) and TR (red). The TR reverberation (red) shows about a 5 dB notch around 6.4 s, corresponding to the PS range. Note that the overall level of BS is 5–10 dB higher than the TR due to the waveform normalization at the SRA prior to retransmission.

normal mode scattering model based on the perturbation method[11,12] is implemented for realistic shallow water environments.

A. Environmental model

Figure 1 shows the waveguide environment model used for backscattered field simulations. The SRA consists of 33 elements spanning the water column from 10 to 106 m with 3 m interelement spacing in the 120 m deep water. The sound-speed profile indicates a typical downward refracting environment with the thermocline spanning 20–50 m depth resulting in substantial sound interactions with the ocean bottom. The ocean–sediment interface has a roughness of 0.1 m rms and 15 m correlation length generated using a power law spectrum.[13] The ocean bottom has a 2.5 m thick sediment layer with a sound speed of 1520 m/s at the top interface that is similar to the geoacoustic properties of the experimental area described in the next section. Since the water column sound speed at this depth is 1508 m/s, acoustic waves propagating above the critical angle of 7.2° are highly attenuated. We use a 100 ms Gaussian shaped pulse with a center frequency of 3.5 kHz. The vector time series transmitted by the SRA is normalized such that the maximum value across all

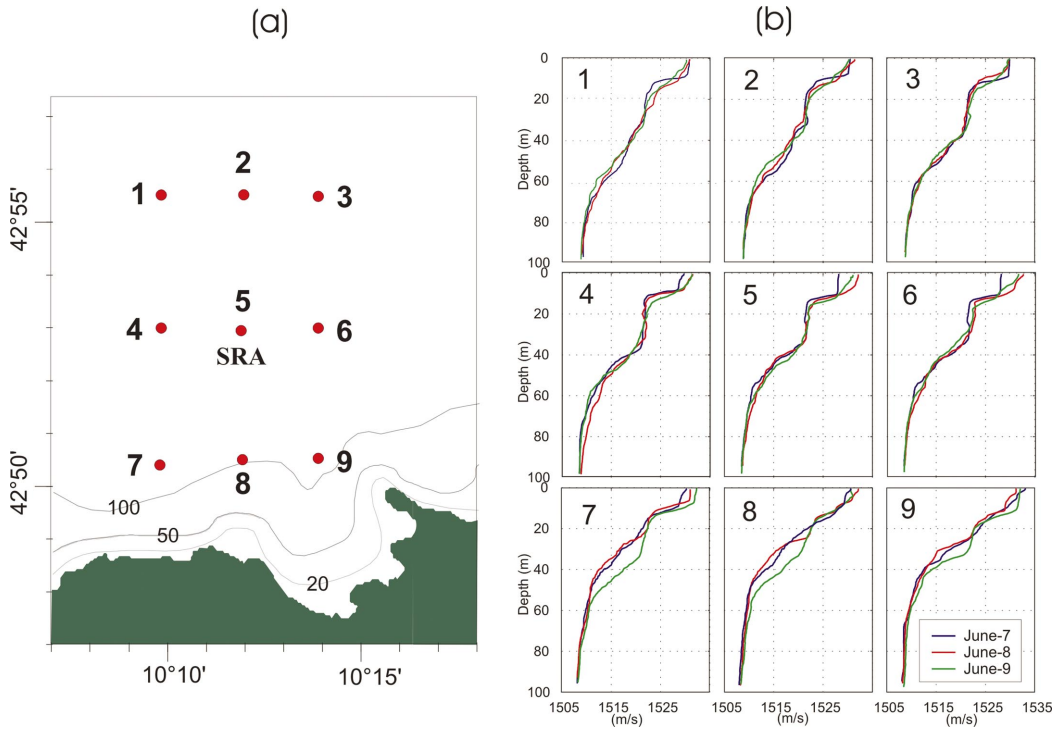


FIG. 3. The TRM Experiment carried out north of Elba Island off the west coast of Italy: (a) map around the SRA denoting the positions of the CTD measurements taken on three consecutive days and (b) sound speed profiles showing the spatial and temporal variability.

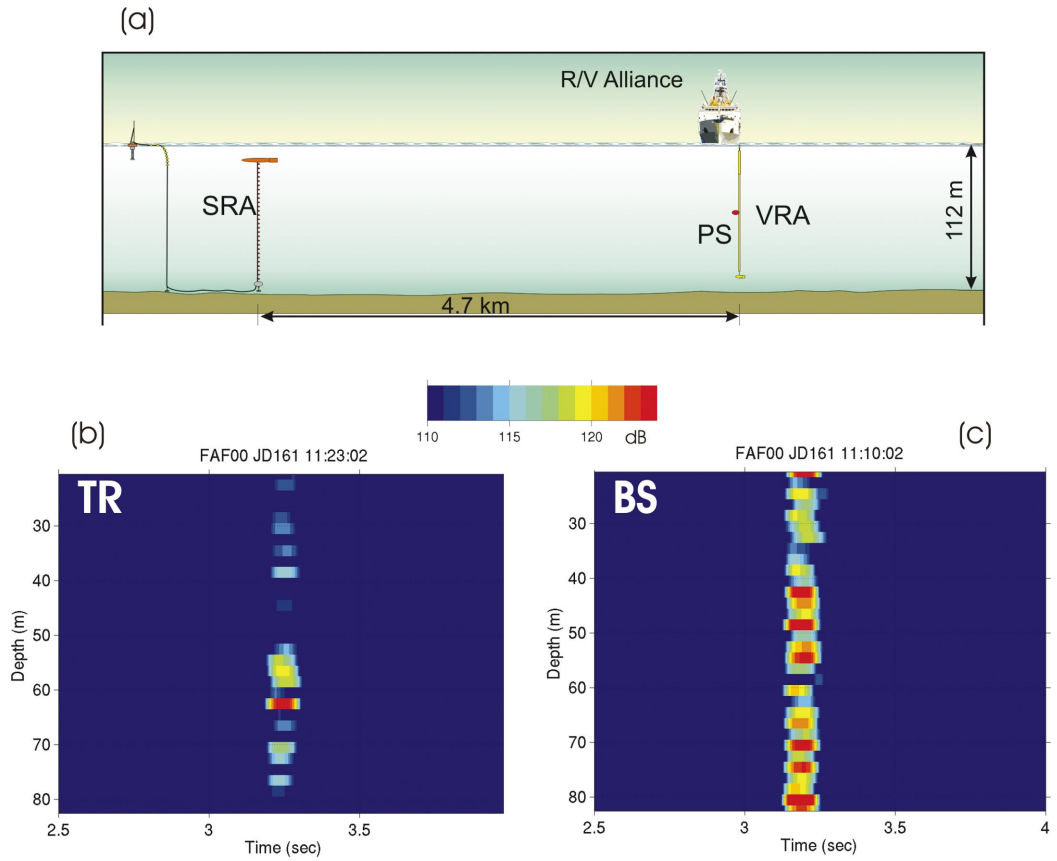


FIG. 4. (a) Experimental configuration for reverberation measurements. The PS was deployed from the R/V ALLIANCE at 60 m depth and 4.7 km range away from the SRA. The PS pulse was a 100 ms long pure tone at 3.75 kHz. (b) TR focusing recorded by the VRA near the PS. For comparison purposes, (c) shows a BS transmission received at the VRA that fills the water column.

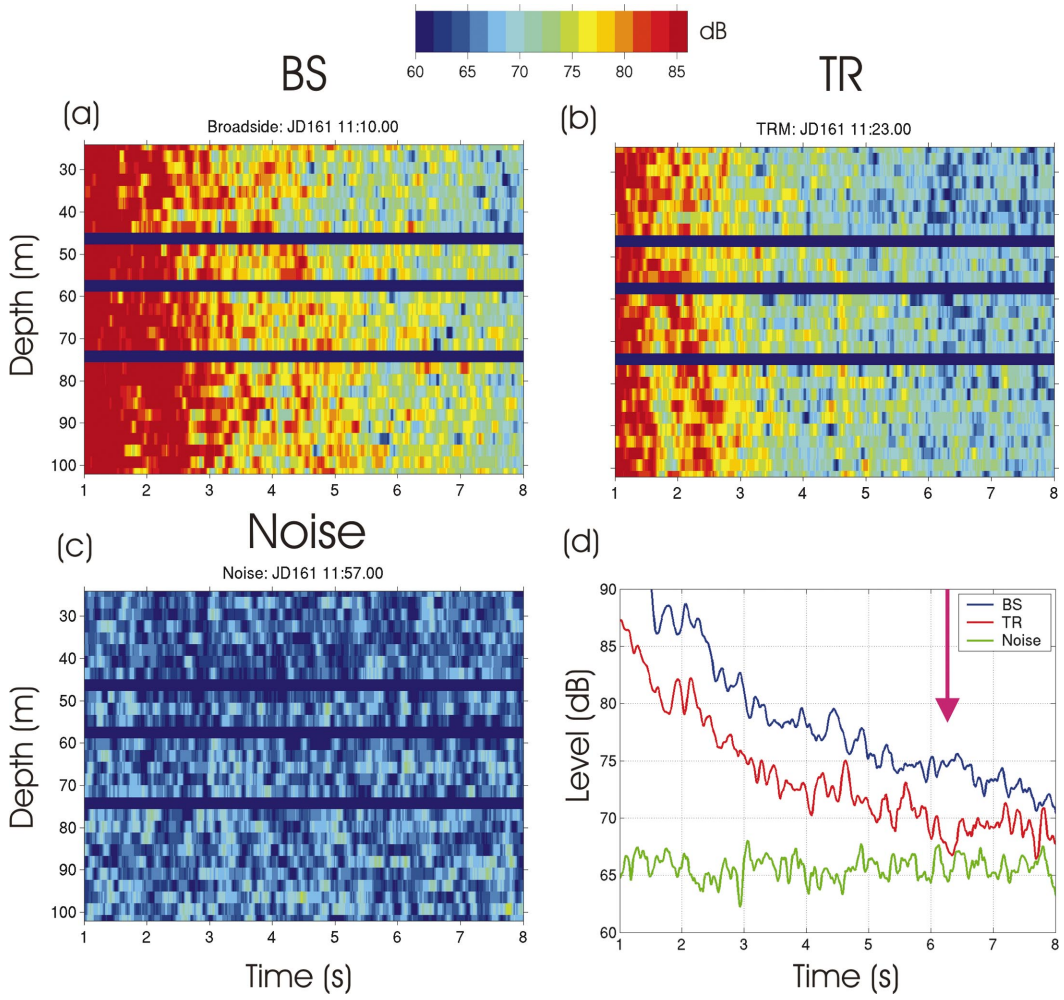


FIG. 5. Measured backscattered field at the SRA: (a) BS transmission and (b) TR transmission. The ambient noise field is also displayed in (c) as a reference. (d) Shows the corresponding reverberation level incoherently averaged across the upper SRA elements (8–29) along with the ambient noise level: BS (blue), TR (red), and noise (green). The TR reverberation indicates about a 3 dB notch around 6.3 s corresponding to the PS range of 4.7 km.

elements and time is equal to the maximum element source level of 174 dB re 1 μPa used during the experiment. This results in approximately 8 dB less energy being transmitted by the SRA for the time-reversal (TR) transmission compared to the conventional broadside (BS) transmission, as shown in the next section.

B. Reverberation

In this section we simulate the reverberation reduction using a time-reversal mirror. First consider non-TRM reverberation from a broadside (BS) SRA transmission. BS is an excitation of the SRA with equal amplitudes. Figure 2(a) shows one-way BS transmission loss in range and depth at 3.5 kHz. The geometrical spreading term $1/r$ has been removed and the dynamic range is 20 dB. On the other hand, Fig. 2(b) shows two-way time-reversal (TR) focusing when the probe source (PS) is at a 4.7 km range and 60 m depth. Note the triangle-shaped shadow zones formed above and below the focus. Figures 2(c) and 2(d) are vertical slices of (a) and (b), respectively, at the PS range of 4.7 km, displaying the energy distribution across the depth. The TR field in the shadow of (d) is lower than that of the focus by 15–20 dB, consistent with the experimental results reported in Refs. 9, 14, 10.

Figure 2(e) shows calculated reverberation levels for both BS and TR transmissions using a 100 ms Gaussian-shaped pulse at 3.5 kHz. The reverberation field is averaged incoherently across the SRA elements. Two important observations can be made. First, the overall level of BS (blue) is higher (approximately 8 dB) than the TR level (red) due to the normalization applied at the SRA prior to retransmission, as predicted. Second, and more important, the TR retransmission produces about a 5 dB reverberation notch (red arrow) around 6.4 s corresponding to the PS range of 4.7 km.

The inherent assumption in our numerical modeling of bottom reverberation is that both the environment and interface roughness are axisymmetric so that there is no out-of-plane scattering. In addition, we neglect volume scattering and ocean surface scattering. These assumptions certainly are unrealistic in general and we do not attempt to do direct model/data comparisons in this paper. Therefore, the absolute levels are not intended to be comparable. However, the experimental results presented in the next section confirm the existence of the notch.

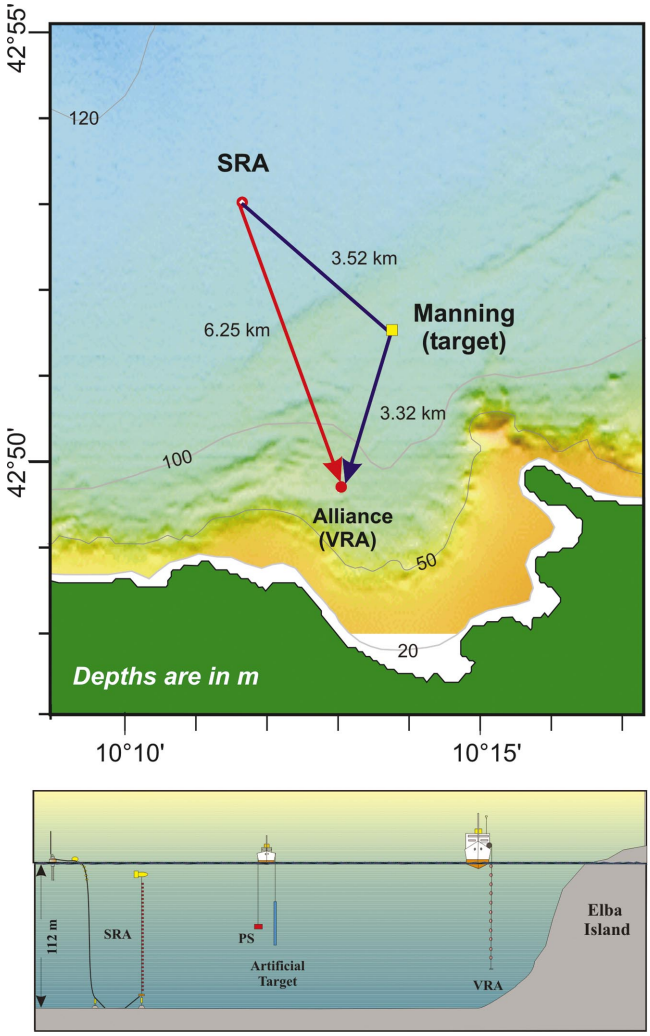


FIG. 6. Schematic for a bistatic scattering experiment. An air-filled tube simulating a target and a PS were deployed from the R/V Manning. The SRA transmitted BS pings as well as the time-reversed probe source reception. The transmissions were monitored bistatically by the VRA deployed from the R/V Alliance.

III. EXPERIMENTAL RESULTS

A time-reversal experiment was performed May/June 2000 north of Elba off the west coast of Italy, as shown in Fig. 3(a). The reverberation portion of the experiment was conducted 5 miles north of Elba in an area of water depth 112 m having a gentle downward slope to the north. A detailed oceanographic survey was carried out during the experiment. Sound speed profiles (SSP) were collected frequently by CTD casts that clearly indicate the downward-refracting structure with a slight range dependency in the north–south direction, as shown in Fig. 3(b). The SRA had 29 transducers spanning a 78 m aperture (2.786 m interelement spacing) with a center frequency of 3.5 and 1 kHz bandwidth. During this part of the experiment the SRA was operated at 3.75 kHz to avoid some unexpected electronic self-noise around 3.5 kHz. The maximum source level was 174 dB re 1 μPa per transducer. A detailed description of the TRM hardware can be found in Ref. 15. We also deployed a 32-element vertical receiver array (VRA) to measure bistatic reverberation as well as to monitor the time-reversal focus.

A. Reverberation reduction

The experimental configuration for reverberation measurement is shown in Fig. 4(a). The SRA covered the water column from 24 to 102 m. A probe source was deployed at 60 m depth from the R/V Alliance that was positioned 4.7 km south of the SRA. A 100 ms pure tone pulse at 3.75 kHz was transmitted from the PS and 3 min later (JD161 11:23:00) the SRA retransmitted the received probe signal in time-reversed order. For comparison purposes, a broadside transmission also was carried out (JD161 11:10:00). The transmission from the SRA were monitored using the VRA deployed next to the PS from the R/V Alliance that covered the water column from 20.5 to 82.5 m with a 2 m interelement spacing. Figures 4(b) and 4(c) show the TR focusing and the BS transmission measured by the VRA, respectively. The sharp focus in Fig. 4(b) occurs at an apparent depth of 63 m (although the PS depth was 60 m, the difference may be due to a slight tilting of the VRA). The intensity decreases rapidly, moving away from the focal depth, and at 80 m (still 32 m above the bottom) the level is 15 dB lower than that of the focus. In contrast, the BS transmission (c) shows the acoustic field spreading over the water column and a high level appears at the lower elements of the VRA, suggesting that high levels also are interacting with the ocean bottom below the VRA.

The returning backscatter from these transmissions was recorded monostatically by the SRA. Figure 5 shows the measured reverberation field: (a) BS and (b) TR transmission. The ambient noise level also is shown in (c) as a reference. During this measurement, three of the SRA transducers (Channels 11, 17, and 21 from the bottom) were not operating correctly and were disabled (thick blue horizontal lines). Figure 5(b) indicates that a lower level appears between 6 to 6.5 s in the upper channel elements (8–29) of the SRA, which is close to the ambient noise level in Fig. 5(c). This structure becomes clear in Fig. 5(d), where the corresponding reverberation levels are superimposed: BS (blue), TR (red), and ambient noise (green). The levels are obtained by incoherent averaging of the upper 19 channels (8–29). The reverberation is strong initially and decreases about 15–20 dB in 7 s. Note that the BS level is about 5 dB higher than that of the TR due to the difference in the transmitted level, as described in Sec. II. In this case, approximately 5 dB less energy was transmitted by the SRA for the TR transmission compared to the conventional BS transmission. The existence of a reverberation notch approximately 400 ms wide (red arrow) and about 3 dB deep is evident. The 3 dB depth is smaller than the 5 dB predicted by numerical modeling in Sec. II, which likely is due to the simplified two-dimensional nature of the simulation.

B. Echo enhancement

In the previous section, we confirmed reduced reverberation using the time-reversal process. The measurements in Fig. 4 also showed an increased level of ensonification at the probe source location (by approximately 5 dB) for the TR transmission over that of the BS transmission.

Here, we present an experimental result including an

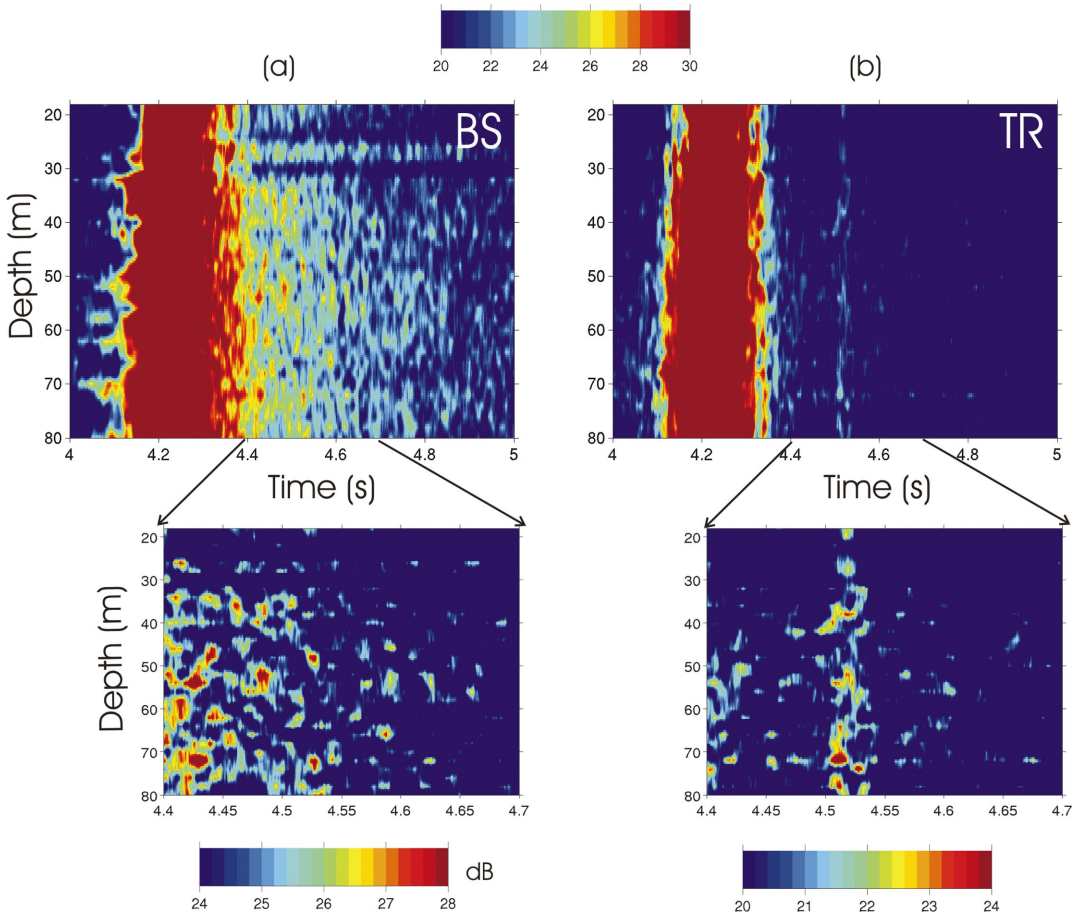


FIG. 7. Experimental results for bistatic scattering observed by the VRA: (a) BS and (b) TR. A signal presumed to be the target echo is seen in (b) at the expected time (4.52 s) corresponding to the SRA-Manning-Alliance range (6.84 km) with the time-reversal transmission after the direct arrival (4.2 s). The target echo is not detectable in (a) due to the high background reverberation. The pulse was a 1 s long LFM (3–4 kHz), and the received signals were matched filtered. Note that the dynamic ranges for the lower two displays are different.

artificial target showing improved target detectability with the time-reversal method. Figure 6 shows the schematic for a bistatic scattering experiment. An air-filled tube was used as an artificial target.[16] The tube was 30 m long and 0.076 m in diameter. It then was folded seven times to be a 4 m long fat target with unknown target strength. The artificial target was deployed at depth of 60 m from the R/V Manning. First, a 1 s long LFM (3–4 kHz) probe source pulse was transmitted near the target. After receiving the probe source pulse, the SRA located 3.52 km away from the R/V Manning retransmitted the time-reversed signal. At the same time, the R/V Alliance deployed the VRA to monitor the target echo bistatically 6.25 km away from the SRA and 3.32 km away from the R/V Manning. For comparison purposes, the SRA also transmitted a broadside ping.

Figures 7(a) and 7(b) show the results of the BS and TR transmission, respectively. Note that the received signals were matched filtered by the original pulse. The broad area around 4.2 s is the direct arrival from the SRA to the VRA. After this, the BS still produces high reverberation up to 4.8 s while the TR shows a relatively low-level background level during the same time. An echo is visible around 4.52 s in the TR (b), which is the expected time of the target arrival corresponding to a range of 6.84 km (the SRA-Manning-Alliance distance). The enhanced detectability of the target echo in Fig. 7 is believed to be due to the combined effect of focusing on the target and reduced reverberation near the target. Unfortunately, the SRA was experiencing a self-noise problem at the time and the relatively low-level backscattered signal (target echo and reverberation) could not be measured monostatically.

IV. CONCLUSIONS

Echo-to-reverberation enhancement using time reversal was demonstrated experimentally using monostatic and bistatic configurations in the 3–4 kHz band in a mildly range-dependent shallow water environment. The monostatic experiment showed a reverberation notch of about 3 dB depth at the probe source range with a time-reversal transmission. This was in addition to an overall reverberation reduction of approximately 5 dB and an enhancement in the ensonification level at the probe source location of approximately 5 dB. These experimental results are consistent qualitatively with numerical simulation results. The bistatic experiment with an artificial target also verified improved target detectability using the time-reversal method.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research under Grant No. N00014-94-1-0458 and Contract No. N00014-01-D-0043-D06.

[1] A. Parvulescu and C. S. Clay, ''Reproducibility of signal transmissions in the ocean,'' Radio Electron. Eng. **29**, 223–228 (1965).
[2] M. Fink, ''Time-reversed acoustics,'' Sci. American, pages 91–97, November 1999.
[3] D. R. Jackson and D. R. Dowling, ''Phase conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181 (1991).
[4] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. Jackson, ''Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror,'' J. Acoust. Soc. Am. **102**, 25–40 (1998).
[5] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, ''A long range and variable focus phase conjugation experiment in shallow water,'' J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[6] P. Roux and M. Fink, ''Time reversal in a waveguide: Study of the temporal and spatial focusing,'' J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[7] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, ''A time-reversal mirror with variable range focusing,'' J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[8] H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, C. Ferla, and D. R. Jackson, ''Iterative time reversal in the ocean,'' J. Acoust. Soc. Am. **105**, 3176–3184 (1999).
[9] S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, ''Spatial resolution of time-reversal arrays in shallow water,'' J. Acoust. Soc. Am. **110**, 820–829 (2001).
[10] H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and P. Guerrini, ''Demonstration of a high frequency acoustic barrier with a time reversal mirror,'' IEEE J. Ocean. Eng. **28**, 246–249 (2003).
[11] B. H. Tracey and H. Schmidt, ''Seismo-acoustic field statistics in shallow water,'' IEEE J. Ocean. Eng. **22**, 317–331 (1997).
[12] J. F. Lingevitch, H. C. Song, and W. A. Kuperman, ''Time reversed reverberation focusing in a waveguide,'' J. Acoust. Soc. Am. **111**, 2609–2614 (2002).
[13] J. A. Goff and T. Jordan, ''Stochastic modeling of seafloor morphology: Inversion of sea beam data for second-order statistics,'' J. Geophys. Res. **93**, 589–608 (1988).
[14] G. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, ''An initial demonstration of underwater acoustic communication using time reversal mirror,'' IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[15] W. S. Hodgkiss, J. D. Skinner, G. E. Edmonds, R. A. Harriss, and D. E. Ensberg, ''A high frequency phase conjugation array,'' in *Proceedings of Oceans 2001*, 2001, pp. 1581–1585.
[16] C. I. Malme, ''Development of a high target strength passive acoustic reflector for low-frequency sonar application,'' IEEE J. Ocean. Eng. **19**, 438–448 (1994).

Experimental investigation of ultrasound propagation in turbulent, diffractive media

T. A. Andreeva[a] and W. W. Durgin
Mechanical Engineering Department, Worcester Polytechnic Institute, Worcester, Massachusetts 01609

(Received 30 March 2003; revised 23 December 2003; accepted 5 January 2004)

Experimental study of the propagation of high-frequency acoustic waves in grid-generated turbulence in laboratory conditions by means of ultrasound technique is considered. The travel time and log-amplitude variance of acoustic waves are investigated as functions of travel distance and mean velocity. Experimental data were interpreted using theoretical analysis based on ray acoustics and Rytov approximation. The travel time shift in acoustic wave propagation in turbulent and undisturbed media is observed. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1687733]

PACS numbers: 43.35.Ae, 43.35.Yb, 43.35.Zc [YHB] Pages: 1532–1536

I. INTRODUCTION

The classical theory of wave propagation in turbulent media considers wave propagation in isotropic or locally isotropic and homogeneous random media and based on statistical representation of the turbulence.[1–3] The statistical moments of phase and log-amplitude fluctuations of a sound wave propagating in the turbulent atmosphere have been calculated by Tatarskii[4] using the ray approximation and by the Rytov method. Ray acoustics has been a standard approach for rigorous consideration of sound wave parameters for outdoor experiments. However, when the propagation distance is too large for diffraction to be neglected the Rytov approximation enables us to obtain, for this class of problems, a broader range of results than geometrical optics does. Moreover, for example, in statistical tomography in seismic media, all scales of heterogeneities are present and finer scale velocity structures also have an impact on the travel times and the imaging process, which leads to failure of the geometrical optics.[5,6]

The modern theory of sound propagation in a moving random medium has been developing intensively since the mid-1980s.[7] The statistical characteristics of sound waves propagating in a moving random medium with an arbitrary state equation have been calculated by the ray acoustics, Rytov, and parabolic-equation methods. All these new results are systematically described by Ostashev.[8] Efforts focusing on the stochastic Helmholtz equation and its parabolic approximation are especially of interest. In this approach the small perturbation method is the most frequently used solution to the parabolic approximation.[9,6] In this method the influence of turbulence on ultrasonic wave propagation is expressed in terms of travel time and amplitude variations.

Experimental outdoor studies of time of flight variance have been performed infrequently despite their significance in understanding the properties of wave propagation in a turbulent atmosphere. Laboratory measurements are also rare due to technological limitations of measuring devices. Some preliminary outdoor investigations on the influence of atmospheric turbulence on sound propagation have been performed by Otrezov and Chunchuzov,[10] Daigle *et al.*,[11,12] Rasmussen,[13] Bass *et al.*,[14] Ostashev *et al.*,[15] Wilson and Thomson.[16,17] Laboratory investigations were done by Ho and Kovasznay,[18] who made measurements across an airjet over an extremely short propagation distance. Blanc-Benon[19] generated plane acoustic waves with a pistonlike sound source and measured the acoustic phase variance as a function of different frequencies.

Another rapidly developing approach, besides analytical and experimental, is the numerical one. The effect of turbulence on sound propagation through numerical simulations was investigated in works by Karweit *et al.* (Ref. 19 and references therein), Blanc-Benon,[20,21] Chevret,[22] and Ioos *et al.*[23] analytically and numerically studied the high frequency propagation of acoustic plane and spherical waves in random media. Using ray acoustics and a perturbation approach they obtained the travel time variance at second order and demonstrated nonlinear behavior of travel time variance at large propagation distances.

In the late 1980s so-called velocity shift was discovered. This effect corresponds to the situation when apparent velocity of the wave is higher than the velocity corresponding to that in the background homogeneous medium; the difference is called ''velocity shift.'' This is in accordance with Fermat's principle: the wave path minimizes the travel time of the wave.[23,6] In other words we can consider the travel time shift in turbulent media. This effect has been studied numerically and theoretically.[6]

The main objective of this paper is to apply the ultrasonic technique to investigate the logamplitude and travel time variations as well as average travel time as functions of travel distance and mean velocity under laboratory conditions. In the experiment, the conditions for ray acoustics are violated and slight diffraction effects have to be taken into account. Therefore, the theoretical analysis of the present work is based on the parabolic equation with Chernov approximation and consequent application of the Rytov method, written in terms of log-amplitude and phase variation. Analytical statistical moments of phase and log-

[a]Electronic mail: tatiana@wpi.edu

 0001-4966/2004/115(4)/1532/5/$20.00

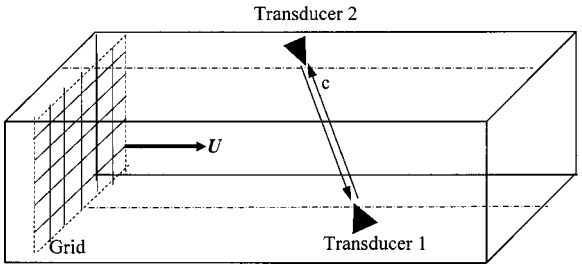


FIG. 1. Sketch of the wind-tunnel test section with ultrasonic transducers functioning as a transmitter and a receiver.

amplitude for Kolmogorov and Gaussian spectra are compared to experimental data.

In Sec. II we describe the experimental setup, which follows an earlier experimental design described in Andreva and Durgin.[24] In Sec. III we summarize the theoretical results of Rytov and Ostashev to make direct comparison with our experimental observations. Mathematical results are listed without derivations since complete discussion can be found in references. In Sec. IV we discuss experimental results and make comparisons with analytical predictions.

II. EXPERIMENTAL APPROACH

The experiments were carried out in a wind tunnel of 107 cm length with a 29.8 cm×29.5 cm rectangular test section. Turbulence was produced by a bi-planar grid consisting of a square mesh of aluminum round rods with diameter 0.635 cm positioned 2.54 cm between centers. The mesh size, M, was therefore 2.54 cm, and the grid solidity was 0.36. At a sufficient distance downstream from the grid the flow becomes locally isotropic and a power law is applicable.[25] Studies and criteria, which relate to identification of the downstream position, where the flow becomes nearly homogeneous, isotropic, and locally isotropic are described in detail in the literature.[26–28] Experimental conditions were chosen similar to the experiments by previous investigators who studied decay of velocity fluctuations, performed measurements of spectrum of velocity fluctuations, etc., for the specific grid and tunnel geometry as well as mean velocities.[25,29,30] In this paper our primary concern will be application of the ultrasonic technique for investigation of the effect of turbulence on ultrasonic wave propagation. In the experiment we measure travel time and amplitude of ultrasonic waves. In our experiment we used two transducers working at a frequency of 100 kHz, designed for air applications, located on the upper and lower sides of the tunnel, as shown in Fig. 1. Transducer-transmitter is excited by the programmable signal generator and a power amplifier by means of burst of four 100 kHz square waves with 50 mV amplitude. The function generator is initiated by National Instrument Data Acquisition Card (DAQ), which produces 5 V amplitude square waves with a frequency of 500 cycles/s. The ultrasound beam, sent by the first transducer, was received by the second one. The analog data, then, from the second transducer was transported to a CompuScope 82G DAQ with large acquisition memory and wide analog bandwidth, that transformed analog data to digital data and transferred data from CompuScope 82G card to the PC memory

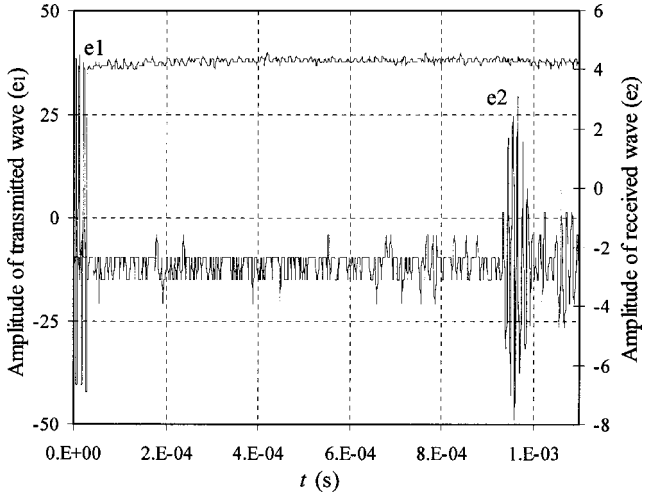


FIG. 2. Transmitted (burst of square waves generated by the function generator) and received (sinusoidal waves) signal as they appear from digital data acquisition system CompuScope 82.

with the resolution of 5×10^{-9} s. Both DAQ cards were installed inside the PC. The major challenge in this class of problems is to measure the propagation time with a high precision. The National Instrument DAQ allows one to detect the burst departure time extremely accurately and, finally, digital representation of the experimental data provided by CompuScope 82G DAQ allowed determination of the travel time very precisely. Figure 2 demonstrates a typical data representation obtained from CompuScope 82G DAQ, transferred to the PC and processed in EXCEL. The acquisition rate was 5×10^{7} samples/s. The first signal e_1 corresponds to the burst of square waves produced by the function generated that initiated the transducer-transmitter. The second signal e_2 is the signal received by the second transducer. The block diagram of analog and digital processing is shown in Fig. 3. The measuring time for each single run was 45 s (approximately 15 Mb of measured data). For each measurement the travel time was averaged over more then 700 realizations.

III. THEORETICAL BACKGROUND

There are three primary spectral subranges in turbulence: the energy-containing, the inertial, and the dissipation.[3] The integral scale L_0 is the characteristic length of the largest motions while the Kolmogorov microscale l_0 is the smallest, most dissipative of the motions. There is no existing model to accurately describe the entire turbulence spectrum. Consequently, it is important to know which portion of the spectrum contributes most significantly in a given acoustic propagation problem. The large-scale, energetic motions drive acoustic phase fluctuations, while smaller-scale motions drive the amplitude fluctuations.[31] The energy-containing subrange plays the primary role in the experiment reported here. For our experimental conditions the smallest, Kolmogorov, scale is $l_0\approx5.6\times10^{-4}$ m, while the largest, integral, scale is the size of the grid spacing, $L_0=2.54\times10^{-2}$ m. The maximum length of the travel path of acous-

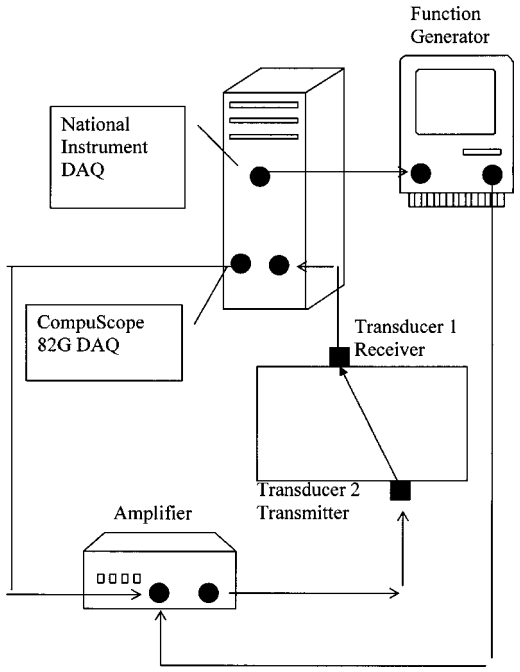


FIG. 3. Block diagram of data acquisition system.

tic wave is 0.45 m. Consequently, according to the geometry of the experiment our experimental domain falls into the diffraction region.

The equations for the variances of the log-amplitude and phase fluctuations of a plane wave in a moving random medium derived from parabolic equation by means of Rytov method[9] have the following form:[8]

$$\langle \chi^2 \rangle = \frac{\pi^2 k^2 x}{2} \int_0^\infty K \left[1 - \frac{K_F^2}{K^2} \sin \frac{K^2}{K_F^2} \right] \Phi_{\text{eff}}(0,K)\, dK, \quad (1)$$

$$\langle \phi^2 \rangle = \frac{\pi^2 k^2 x}{2} \int_0^\infty K \left[1 + \frac{K_F^2}{K^2} \sin \frac{K^2}{K_F^2} \right] \Phi_{\text{eff}}(0,K)\, dK. \quad (2)$$

Here $K_F^2 = k/x$, where k is a wave number and x is a travel distance. Note that $K_F = \sqrt{2\pi}/L_F$, where $L_F = \sqrt{\lambda x}$ is the scale of the first Fresnel zone and λ is a wavelength. Φ_{eff} is the three-dimensional spectral density of the random field and is called an effective function. The equations for statistical moments of ϕ and χ are valid for an arbitrary spectra of inhomogeneities in moving random medium. For our purpose we use Kolmogorov and Gaussian spectra of medium inhomogeneities. For the Kolmogorov spectrum the Markov approximation for the effective spectral density is[8]

$$\Phi_{\text{eff}}(0,K) = A \left(C_\epsilon^2 + \frac{22}{3} \frac{C_v^2}{c_0^2} \right)^2 K^{-11/3}, \quad (3)$$

where A is a numerical coefficient, C_ϵ^2 and C_v^2 are the structure parameters of the random field.[8] Substituting Eq. (3) into Eq. (1) and calculating the integrals over t and K^2 results in a formula for the variance of log-amplitude fluctuations for a plane wave:

$$\langle \chi^2 \rangle = 0.077 \left(C_\epsilon^2 + \frac{22}{3} \frac{C_v^2}{c_0^2} \right) k^{7/6} x^{11/3}. \quad (4)$$

Substitution of Eq. (3) for the Kolmogorov spectrum into the integral for phase fluctuations results in a divergent integral. Therefore, $\langle \phi^2 \rangle$ can be determined neither for a plane nor for a spherical wave.[8] The formula for an effective spectral density for the Gaussian spectrum is

$$\Phi_{\text{eff}}(0,K) = \frac{l^3}{8\pi^{3/2}} \left(\sigma_\epsilon^2 + \frac{\sigma_v^2 K^2 l^2}{c_0^2} \right) \exp(-K^2 l^2/4), \quad (5)$$

where scale l is greater than the inner scale of the turbulence, l_0, and smaller than the outer scale, L_0: $L_0 > l > l_0$. For plane wave propagation, the variances of the log-amplitude and phase fluctuations are[8]

$$\langle \phi^2 \rangle = \frac{\sqrt{\pi} k^2 l x}{8} \left[\left(1 + \frac{\arctan D}{D} \right) \sigma_\epsilon^2 + \left(1 + \frac{\arctan D}{1 + D^2} \right) \frac{4\sigma_v^2}{c_0^2} \right], \quad (6)$$

$$\langle \chi^2 \rangle = \frac{\sqrt{\pi} k^2 l x}{8} \left[\left(1 - \frac{\arctan D}{D} \right) \sigma_\epsilon^2 + \left(1 - \frac{\arctan D}{1 + D^2} \right) \frac{4\sigma_v^2}{c_0^2} \right]. \quad (7)$$

In Eqs. (6) and (7) $D = 4x/(kl^2)$ is the wave parameter used in the theory of waves in random media[1] and σ_v is a variance of velocity fluctuations. Equations (6) and (7) in their original version were given for $\sigma_v^2 = 0$.[9]

IV. EXPERIMENTAL RESULTS AND DISCUSSION

We consider a locally isotropic turbulent flow, generated by a grid at room temperature. The purpose is to investigate the log-amplitude and travel time of acoustic waves as functions of travel distance as well as a mean velocity. Two different experimental setups are presented in Figs. 4(a) and (b). The first setup shown in Fig. 4(a) serves for investigation of acoustic wave parameters as a function of a mean velocity U, that changes from 1 to 10 m/s, so that the Re number based on the grid space size changes from 4016 to 20 080. The path length stayed unchanged. The second setup shown in Fig. 4(b) serves to study the influence of the travel distance that changes from 0.33 to 0.45 m. The mean velocity was unchanged, 3.5 m/s.

In the first set of experiments, acoustic waves were sent upstream and downstream with respect to the mean flow. Travel time for both cases is plotted in Fig. 5 along with the theoretical estimates for the travel times in the homogeneous uniformly moving medium. The effect of the travel-time shift Δt between $\langle t \rangle$ and t_0, where t_0 is travel time in undisturbed media, is observed. Average travel time $\langle t \rangle$ is defined as $\Sigma_t^N t_i/N$. Due to the fast path effect, the effective velocity is higher than the mean velocity of the medium. Indeed, $\langle t \rangle \leqslant t_0$ and $X/\langle t \rangle \geqslant X/t_0$.[23] The effect is in accordance with Fermat's principle and supported by experimental evidence in the area of statistical tomography reported in recent works.[6,32,33]

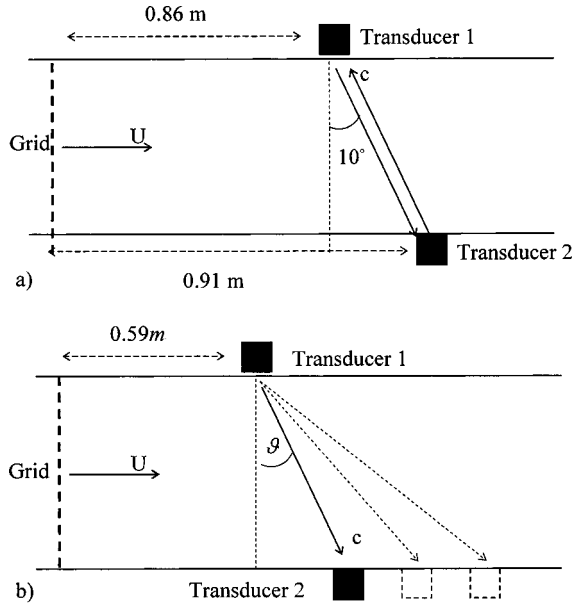


FIG. 4. Diagrams of two experimental setups. Setup (a) serves to investigate the acoustic wave parameters as a function of a mean velocity U, setup (b) serves to stuffy the influence of the travel distance, L.

Another interesting effect of acoustic wave propagation is that in accordance with ray acoustic theory there is a linear increase of travel-time variance with distance.[1] However, nonlinear effects at a certain propagation distance were observed in numerical experiments by Karweit *et al.*[19] and in the work by Iooss *et al.*[23] In our experiment due to the limited size of the wind tunnel we did not reach the distances where these nonlinear effects could be observed. In Fig. 6 we compare our experimental data first with theoretical results obtained by Iooss *et al.*[23] In their work the authors were investigating travel time using geometrical optics approach, which neglects all diffraction phenomena. They developed a theoretical model for the second-order travel-time variance for the plane waves. Second, we compare our results with

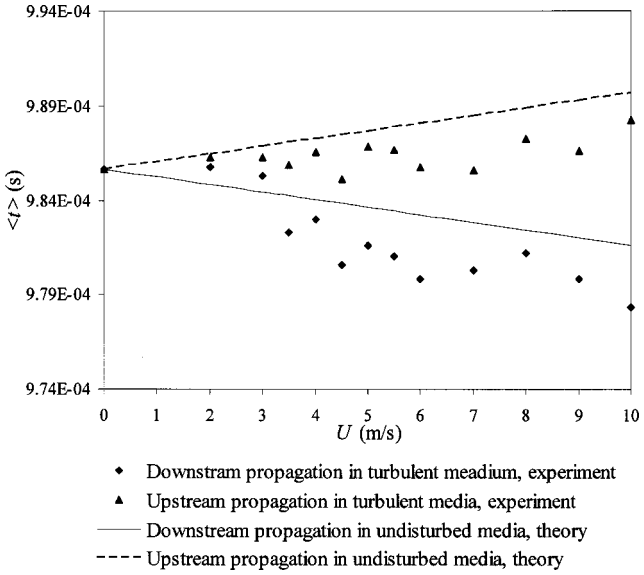


FIG. 5. Experimental data for mean travel time as a function of mean velocity for upstream and downstream propagation plotted along with theoretical estimates for the travel times in undisturbed medium.

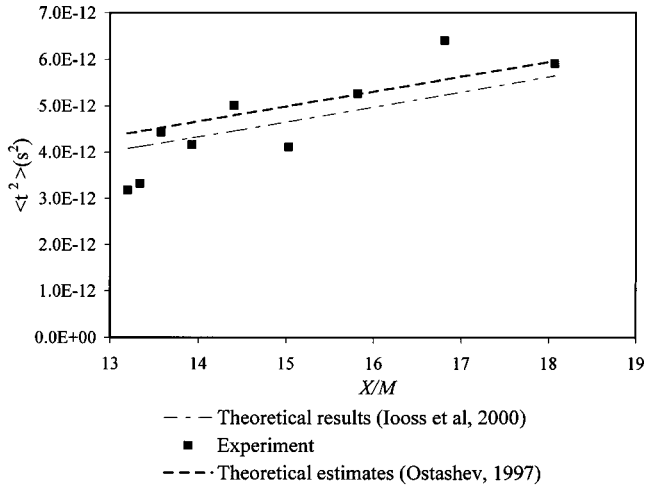


FIG. 6. Experimental data for the travel-time variance vs normalized travel distance x/M. Rytov solution and theoretical model by Iooss *et al.* are plotted for qualitative comparison.

solution of the parabolic equation for the travel-time variance of a plane wave in a moving random media, derived by means of the Rytov method and Markov approximation for the Gaussian spectrum of medium inhomogeneities,[9] Eq. (6). During the experiment we did not have a possibility to measure all flow parameters entering Eq. (6). Consequently, for comparison, we simply reproduce the arctangent behavior of the travel time variance, namely, $\langle \tau^2 \rangle \equiv \langle (t-\langle t \rangle)^2 \rangle \sim (D + \arctan D)$. The scale l entering the parameter D was estimated to be approximately $\sim 5\times 10^{-3}$ m.[25] For demonstration purposes, all the analytical lines start from the same point. We observe that nonlinear effects of second-order travel-time variance do not appear at such short distances. Moreover, comparison of the travel time variance obtained using the Rytov method and ray acoustic approach reveals that some of the results of geometric acoustics are acceptable even beyond the area of the validity of the approach. It has been shown by Rytov[9] that the ray acoustics is accurate enough for phase difference calculations, since accounting for diffraction effects matters only in numerical coefficients.

Figure 7 shows experimental data for log-amplitude variations plotted along with theoretical results, Eq. (4) for Kolmogorov spectra, Eq. (7) for Gaussian spectra and its

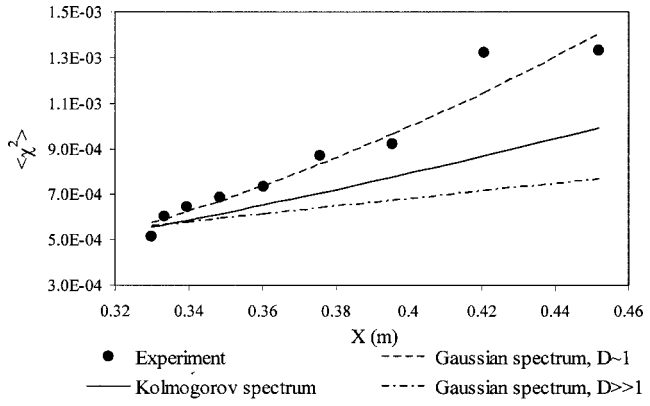


FIG. 7. Experimental data for the log-amplitude variance as a function of travel distance. Rytov solution for Kolmogorov spectra, Gaussian spectra, and Frauhofer diffraction are plotted for qualitative comparison.

limiting case, Frauhofer diffraction for $D \gg 1$. In the intermediate case, which is the case in our experiment (D is close to unity), the nonlinearity due to arctangent is well established. The general form of the log-amplitude variance in the case of Gaussian spectra, which takes into consideration small diffraction effects, is the best fit for our data. For demonstration purposes, all the analytical lines start from the same point. For the present paper we did not intend to achieve quantitative agreement, but rather qualitative accordance.[33]

V. CONCLUSIONS

The new result of this work is the experimental data, collected under the laboratory conditions using high accuracy equipment. Such experiments are rarely performed despite their crucial importance in understanding the described phenomenon.

Using experimental data of travel time the effect of turbulence on acoustic wave propagation has been demonstrated in terms of travel time and log-amplitude variations. Experimental data have been interpreted using theoretical analysis based on Rytov's approach.

From the travel-time measurements performed in both turbulent and nonturbulent media, Fermat's principle has been validated.

It is clear from experimental results for travel-time variance that in the presence of small diffraction effects, the ray acoustic approach is valid, so the area of ray acoustic approach is broader than the rigorous sufficient conditions defined, at least for travel-time fluctuations. The experimental data confirm the qualitative considerations, expressed in Rytov[9] in that amplitude variation is more influenced by diffraction effects than the travel time variation.

Overall, experimental data demonstrate good agreement with analytical estimations. It should be emphasized that such experiments are possible only due to Hi-Tech equipment, available today and capable of collecting enormous amounts of experimental data at high rate and analyzing it. This opens a new perspective in elucidating the properties of wave propagation in random media.

[1] L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw–Hill, New York, 1960).
[2] V. I. Tatarski, *The Effect of the Turbulent Atmosphere on Waves Propagation* (Israel Program for Scientific Translation, Jerusalem, 1971).
[3] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
[4] V. I. Tatarskii, *Wave Propagation in Turbulent Media* (Nauka, Moscow, 1967), in Russian.
[5] B. Iooss and A. Galli, "Statistical tomography for seismic reflection data," *Proceedings of the Sixth International Geostatistics Congress*, Cape Town, South Africa, April 2000.
[6] Y. Samuelides, "Velocity shift using the Rytov approximation," J. Acoust. Soc. Am. **104**, 2596–2603 (1998).
[7] V. E. Ostashev, "Sound propagation and scattering in media with random inhomogeneities of sound speed, density and medium velocity," Waves Random Media **4**, 403–428 (1994).
[8] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN SPON, London, 1997).
[9] S. M. Rytov, Yu. A. Kravcov, and V. I. Tatarskii, *Elements of Random Process Theory: Principles of Statistical Radiophysics* (Springer, Berlin, 1987), Vol. 4.
[10] A. I. Otrezov and I. P. Chunchuzov, "Propagation of low-frequency sound waves in the near-earth layer of the atmosphere," Izv. Akad. Nauk SSSR. Fiz. Atmos. Okeana **22**, 358–362 (1986).
[11] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, "Line-of-site propagation through atmospheric turbulence near the ground," J. Acoust. Soc. Am. **74**, 1505–1513 (1983).
[12] G. A. Daigle, T. F. W. Embleton, and J. E. Piercy, "Propagation of sound in the presence of gradients and turbulence near the ground," J. Acoust. Soc. Am. **79**, 613–627 (1986).
[13] K. B. Rasmussen, "Outdoor sound propagation under the influence of wind and temperature gradients," J. Sound Vib. **104**, 321–335 (1986).
[14] H. E. Bass, L. N. Bolen, R. Raspet, W. McBride, and J. Noble, "Acoustic propagation through a turbulent atmosphere: Experimental characterization," J. Acoust. Soc. Am. **90**, 3307–3313 (1991).
[15] V. E. Ostashev, E. M. Salomons, S. F. Clifford, R. J. Lataitis, D. K. Wilson, Ph. Blanc-Benon, and D. Juvé, "Sound propagation in a turbulent atmosphere near the ground: A parabolic equation approach," J. Acoust. Soc. Am. **109**, 1894–1906 (2001).
[16] D. K. Wilson and D. W. Thomson, J. Acoust. Soc. Am. **96**, 1080–1095 (1994).
[17] D. K. Wilson and D. W. Thomson, Acustica **87**, 677–684 (2001).
[18] C. M. Ho and L. S. G. Kovasznay, "Modulation of an acoustic wave by turbulent shear flow," *U.S. Air Force Office of Scientific Research, Interim Tech. Rep.* No. F44-620-69-C-0023, 1974.
[19] M. Karweit, Ph. Blanc-Benon, D. Juvé, and G. Comte-Bellot, "Simulation of the propagation of an acoustic wave through a turbulent velocity field: A study of phase variance," J. Acoust. Soc. Am. **89**, 52–62 (1991).
[20] Ph. Blanc-Benon, D. Juvé, and G. Comte-Bellot, "Occurrence of caustics for high-frequency caustic waves propagating through turbulent field," Theor. Comput. Fluid Dyn. **2**, 271–278 (1991).
[21] Ph. Blanc-Benon, D. Juvé, and P. Chevret, "On the influence of the turbulence modeling for atmospheric sound propagation," J. Acoust. Soc. Am. **98**, 2924 (1995).
[22] P. Chevret, Ph. Blanc-Benon, and D. Juvé, "A numerical model for sound propagation through a turbulent atmosphere near the ground," J. Acoust. Soc. Am. **100**, 3587–3599 (1996).
[23] B. Iooss, Ph. Blanc-Benon, and C. Lhuillier, "Statistical moments of travel times at second order in isotropic and anisotropic random media," Waves Random Media **10**, 381–394 (2000).
[24] T. A. Andreeva and W. W. Durgin, "Ultrasound technique for prediction of statistical characteristics of grid-generated turbulence," AIAA J. **41**(8), 1438–1443 (2002).
[25] K. R. Sreenivasan, S. Tavoularis, R. Henry, and S. Corrsin, "Temperature fluctuations and scales in grid-generated turbulence," J. Fluid Mech. **100**, 597–621 (1980).
[26] S. Corrsin, *Encyclopedia of Physics* (1963), Vol. 8, p. 568.
[27] M. S. Mohamed and J. C. LaRue, "The decay power law in grid-generated turbulence," J. Fluid Mech. **219**, 195–214 (1990).
[28] M. S. Uberoi and S. Wallis, "Effect of grid geometry on turbulence decay," Phys. Fluids **10**, 1216–1230 (1967).
[29] G. Comte-Bellot and S. Corrsin, "Simple Eulerian time correlation of full- and narrow-band velocity signals in grid-generated, 'isotropic' turbulence," J. Fluid Mech. **48**, 273–337 (1971).
[30] T. T. Yeh and C. W. van Atta, "Spectral transfer of scalar and velocity fields in heated-grid turbulence," J. Fluid Mech. **58**, 233–261 (1973).
[31] D. K. Wilson, "A turbulence spectral model for sound propagation in the atmosphere that incorporates shear and buoyancy forcing," J. Acoust. Soc. Am. **108**, 2021–2038 (2000).
[32] T. Mukerji, G. Mavko, D. Mujica, and N. Lucet, "Scale-dependent seismic velocity in heterogeneous media," Geophysics **60**, 1222–1233 (1995).
[33] M. Roth, G. Muller, and R. Snider, "Velocity shift in random media," Geophys. J. Int. **115**, 552–563 (1993).

Acoustic waves generated by a laser line pulse in cylinders; Application to the elastic constants measurement

Y. Pan[a)]
Institute of Acoustics, Tongji University, 200092, Shanghai, People's Republic of China

C. Rossignol and B. Audoin
Laboratoire de Mécanique Physique, UMR CNRS 8469, Université Bordeaux 1, 33405 Talence, France

(Received 5 August 2003; revised 8 January 2004; accepted 9 January 2004)

A model is proposed to predict acoustic waves generated in a transversely isotropic cylinder by a laser line pulse extended in beamwidth and time duration, and an application to elastic constants measurement is presented. Documented good agreements are observed in the comparison of experimental and theoretical normal displacements for aluminum cylinders under either ablation or thermoelastic generation. Bulk waves are identified and processed for the elastic constants measurement. The effects of source beamwidth and time duration on wave forms and on the elastic constants measurement are predicted by numerical simulations. For nondestructive evaluation applications using bulk waves, a radius of 0.3 mm appears as a minimum limit for the sample size using a laser source of 0.1 mm beamwidth and 20 ns time duration. Elastic constants of aluminum rods are experimentally measured with very good accuracy. [DOI: 10.1121/1.1651191]

PACS numbers: 43.35.Cg, 43.35.Zc, 43.20.Bi. [YHB] Pages: 1537–1545

I. INTRODUCTION

Cylindrical parts often acting as rotating axes of a machine play important roles. Nowadays, they are in a particular favor of new materials with enhanced mechanical properties such as fiber-enhanced composite materials, which have high performance in strength and durability. Many manufactured rods, fibers, and other cylindrical structures have transverse isotropy. There is an increasing demand of measuring nondestructively their elastic constants, which are often directly related to their mechanical performances. Additionally, a cylinder with its unique geometry is a basic target for the acoustic wave propagation research.

Understanding the wave propagation in a cylinder is a necessary step before considering any possible application. Surface acoustic or Rayleigh wave propagating on an isotropic and homogeneous cylinder was considered in 1927 by Sezawa,[1] who obtained its characteristic equation and calculated the dependence of its velocity on the radius of the cylinder. More detailed studies were reported in 1967 by Viktorov,[2] who calculated its dispersive curve and made a comparison to experimental data detected by conventional piezoelectric transducers. Higher Rayleigh-type waves were termed as ''Whispering-gallery modes'' by Uberall in 1973.[3] Due to the coupling difficulty of such transducers, few experimental reports on the wave propagation for a curved medium had been published until the development of the laser ultrasonic technique,[4] in which ultrasonic waves are both generated and detected by lasers. With the remarkable features of noncontact, high spatial and temporal resolutions introduced by this technique, various studies on materials of curved surfaces have been carried out. As an example, Rayleigh wave propagating on a sphere was observed experimentally in 1988,[5] and a further detailed study has been reported later.[6] Rayleigh wave propagating on a cylinder was also studied by this technique.[7,8] The finite element method has been used to predict the bulk and surface wave propagations when laser beam was focused by a cylindrical lens.[9] Very recently, authors have published a model[10] to successfully predict the bulk and surface wave propagations in a transversely isotropic cylinder under either ablation or thermoelastic generation. In this model, the laser pulse was represented by a sudden normal force with delta function in time for ablation generation or by a dipolar force with a Heavside step function in time for thermoelastic generation.[10] Since laser pulses have certain beamwidth and time duration, these representations can only hold true when the cylinder is large enough to neglect the influence of these physical dimensions. Therefore, it is necessary to take account of the influence when the radius of the cylinder turns to small values.

There has been considerable study of the acoustic scattering characteristics of a transversely isotropic cylinder for the purpose of determining its elastic properties. For a typical example, resonance acoustic spectroscopy, the study of resonance effects present in acoustic echoes of an elastic target, was proposed to evaluate the elastic constants of the cylindrical samples.[11] For this technique, the sample must be immersed in a fluid like water or another material. This limits its capability when the sample is not allowed to do so, such as the measurement at elevated temperatures. Actually, many new materials are intended for use at a high temperature. Another resonance technique, the measurement of longitudinal and torsional resonance frequencies of samples of known geometry, has determined nondestructively the elastic constants of various crystals at high temperature.[12] Free vibrations of a transversely isotropic finite cylindrical rod has been analyzed for the purpose of elastic constants

a)Electronic mail: y.pan@lmp.u-bordeaux1.fr

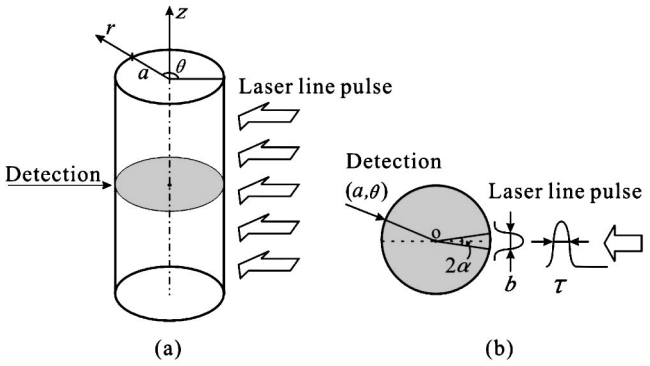


FIG. 1. Problem geometry.

determination.[13] But this technique requires proper machining of samples and fine identification of the resonance peaks. Moreover, noncontact measurement is not possible with this technique. Fortunately, the above limitations are circumvent with the laser ultrasonic technique, in which the ultrasonic waves are generated and detected at a distance, without any contact to the sample.

Recently, many studies have focused on applications of the surface acoustic waves by the laser ultrasonic technique. The velocities of the Rayleigh wave of steel rods receiving different heat treatments were measured to characterize the surface residual stress of the cylindrical parts.[14] The acoustic waves propagation in a sphere has been analyzed to establish a guideline for the design of a ball surface acoustic wave device.[6] The surface wave propagation on a cylinder has been studied to detect surface break defects.[15] According to the author's knowledge, no study on applications of bulk waves in either the sphere or cylinder has been reported using this technique.

In order to address cylinders of all possible sizes for both forward and inverse problems, the published model[10] is generalized in this paper to predict the acoustic waves generated in a transversely isotropic cylinder by a laser line pulse extended in beamwidth and time duration. The application to the measurement of the elastic constants is further considered. At first, the general problem is formulated in Sec. II. Transformed solutions of the generalized model and an inverse transform scheme are presented in Sec. III. To validate the modeling, the calculated normal displacements of aluminum cylinders under either ablation or thermoelastic generation are compared to the experimental signals detected by the laser ultrasonic technique, in Sec. IV. The effects of the source beamwidth and time duration are fully analyzed in Sec. V. Finally, the elastic constants measured on both calculated and experimental wave forms under either generation regime are presented and discussed in Sec. VI.

II. GENERAL FORMULATION

Let us consider a homogeneous and transversely isotropic cylinder of infinite length, radius a, and density ρ. As shown in Fig. 1(a), the symmetrical axis of the cylinder is assumed to coincide with the z axis of its cylindrical coordinates (r, θ, z). A pulsed laser is used to generate acoustic waves in this material. Here, ''detection'' is to denote an experimental technique, generally an optical detection technique,[4] which is applied to measure these acoustic waves. If a cylindrical lens is used to focus the laser beam, this laser can be modeled as a line-like acoustic source, which lies at the boundary of the cylinder, $r=a$, and extends along the z direction. As shown in Fig. 1(b), the laser source is assumed to have a beamwidth b and a time duration τ, and it impacts on the cylindrical surface of an area with a cylindrical angle 2α. To denote the detection position on the cylinder, cylindrical coordinates (a, θ) are chosen, considering $\theta=0$ for the source position.

Let c_{11} and c_{12} denote the two independent elastic constants related to the isotropic plane perpendicular to the z axis. Owing to the symmetry imposed by the source shape, this problem shows invariance along the z direction. The nonzero components of the displacement vector depend on two spatial variables r, θ and on time t. These components, denoted u_r and u_θ can be written as[16]

$$u_r(r,\theta,t)=\frac{\partial\varphi}{\partial r}+\frac{1}{r}\frac{\partial\chi}{\partial\theta},$$

$$u_\theta(r,\theta,t)=\frac{1}{r}\frac{\partial\varphi}{\partial\theta}-\frac{\partial\chi}{\partial r}, \tag{1}$$

where the two scalar potentials φ and χ are governed by the following wave motion equations:

$$\frac{\partial^2\varphi}{\partial r^2}+\frac{1}{r}\frac{\partial\varphi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\varphi}{\partial\theta^2}=\frac{\rho}{c_{11}}\frac{\partial^2\varphi}{\partial t^2},$$

$$\frac{\partial^2\chi}{\partial r^2}+\frac{1}{r}\frac{\partial\chi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\chi}{\partial\theta^2}=\frac{2\rho}{(c_{11}-c_{12})}\frac{\partial^2\chi}{\partial t^2}. \tag{2}$$

Before turning our attention to the boundary conditions, let us define $q(t)$ as the normalized function that represents the source time dependency. Normalization means that the integral of $q(t)$ over its variable equals unity. This function is chosen for the pulsed laser source with a rise time $\tau/2$, i.e., pulse duration τ as[17]

$$q(t)=\frac{4t}{\tau^2}e^{-2t/\tau}, \quad \text{for } t\geqslant 0. \tag{3}$$

In the remainder, $q(t)$ will be changed in to the delta function $\delta(t)$ when a source duration $\tau=0$ is considered. Reciprocally, in the space domain, since ablation occurs when the density of the laser light intensity overpasses a certain threshold,[18] the acoustic source distribution is considered as

$$g_A(\theta)=\begin{cases}\dfrac{1}{2\alpha}, & -\alpha\leqslant\theta\leqslant\alpha,\\ 0, & \text{other},\end{cases} \tag{4}$$

for this generation. Additionally, the distribution for the thermoelastic generation is assumed to be proportional to the light intensity.[19] Therefore a normalized Gaussian function,

$$g_T(\theta)=\frac{1}{2\alpha\sqrt{\pi}}e^{\frac{-\vartheta^2}{4\alpha^2}}, \tag{5}$$

represents the acoustic source in this generation. In Eqs. (4)–(5), the angle α is determined by the beamwidth and the cylindrical radius as $\alpha=\arctan(b/a)$; see Fig. 1(b). Here $g_T(\theta)$

and $g_A(\theta)$ will be changed into the delta function $\delta(\theta)$ when a source width $b=0$ is considered.

Then, two components σ_{rr} and $\sigma_{r\theta}$ of the stress tensor at any point of the surface must comply with the following boundary conditions:

$$\begin{cases} \sigma_{rr}|_{r=a} = -F_0 f(t) \sum_{n=-\infty}^{+\infty} g_A(\theta - 2n\pi), \\ \sigma_{r\theta}|_{r=a} = 0, \end{cases} \tag{6}$$

for the ablation generation,[18] or

$$\begin{cases} \sigma_{rr}|_{r=a} = 0, \\ \sigma_{r\theta}|_{r=a} = -F_0 \int_0^t f(\varepsilon)\, d\varepsilon \sum_{n=-\infty}^{+\infty} \dfrac{\partial g_T(\theta - 2n\pi)}{\partial \theta}, \end{cases} \tag{7}$$

for the thermoelastic generation.[17] In Eqs. (6) and (7), F_0 is a certain loading in N.μs.m^{-1} related to the laser line pulse, and n stands for the number of clockwise ($n>0$) or counterclockwise ($n<0$) roundtrips of the generated acoustic waves. Let us underline that adopting these expressions, a normal loading is considered in the ablation generation, Eq. (6), and a dipolar force is represented in Eq. (7) for the source shape in the thermoelastic generation.[20] Additionally, if the pulse duration is $\tau=0$, the ablative source is impulsive, and it shows a step-like evolution in the time domain for the thermoelastic generation since thermal diffusion is neglected.

III. TRANSFORMED SOLUTIONS AND NUMERICAL INVERSE SCHEME

The two-dimensional Fourier transform of the displacement field over the coordinate θ and time t is now considered, and it is noted as U_i ($i=r$ or θ). On noting $\nu = k_\theta a$, where k_θ is the component of the wave vector $\mathbf{k}$ along the θ direction, the nonzero components of the displacement at a given position and time are then as follows:

$$u_i(r,\theta,t) = \frac{1}{4\pi^2} \int \int_{-\infty}^{+\infty} U_i(r,\nu,\omega) e^{-j(\nu\theta - \omega t)} d\nu\, d\omega. \tag{8}$$

Doing so, the wave motion equations and the boundary equations can be linearized, providing explicit forms for the potentials φ and χ under either generation regime. Components of the displacement are then obtained for ablation generation,

$$\begin{aligned} U_r(r,\nu,\omega) = &- \frac{F_0 a}{(c_{11}-c_{12})D_\nu} \frac{\mathrm{sinc}(\nu\alpha)}{(1+j\omega\tau)^2} \\ &\times \left\{ \left(\nu^2 - \frac{k_T^2 a^2}{2} - B_T \right) B_L \frac{J'_\nu(k_L r)}{J'_\nu(k_L a)} \right. \\ &\left. + \nu^2 (1-B_L) \frac{a}{r} \frac{J_\nu(k_T r)}{J_\nu(k_T a)} \right\} \sum_{n=-\infty}^{+\infty} e^{j\nu 2n\pi}, \end{aligned}$$

$$\begin{aligned} U_\theta(r,\nu,\omega) = &\frac{j\nu F_0 a}{(c_{11}-c_{12})D_\nu} \frac{\mathrm{sinc}(\nu\alpha)}{(1+j\omega\tau)^2} \\ &\times \left\{ \left(\nu^2 - \frac{k_T^2 a^2}{2} - B_T \right) \frac{a}{r} \frac{J_\nu(k_L r)}{J_\nu(k_L a)} \right. \\ &\left. + (1-B_L) B_T \frac{J'_\nu(k_T r)}{J'_\nu(k_T a)} \right\} \sum_{n=-\infty}^{+\infty} e^{j\nu 2n\pi}, \end{aligned} \tag{9}$$

and for thermoelastic generation,

$$\begin{aligned} U_r(r,\nu,\omega) = &\frac{j\nu^2 F_0 a}{\omega(c_{11}-c_{12})D_\nu} \frac{e^{-\alpha^2\nu^2}}{(1+j\omega\tau)^2} \\ &\times \left\{ (1-B_T) B_L \frac{J'_\nu(k_L r)}{J'_\nu(k_L a)} \right. \\ &\left. + \left(\nu^2 - \frac{k_T^2 a^2}{2} - B_L \right) \frac{a}{r} \frac{J_\nu(k_T r)}{J_\nu(k_T a)} \right\} \\ &\times \sum_{n=-\infty}^{+\infty} e^{j\nu 2n\pi}, \end{aligned}$$

$$\begin{aligned} U_\theta(r,\nu,\omega) = &\frac{\nu F_0 a}{\omega(c_{11}-c_{12})D_\nu} \frac{e^{-\alpha^2\nu^2}}{(1+j\omega\tau)^2} \\ &\times \left\{ \nu^2 (1-B_T) \frac{a}{r} \frac{J_\nu(k_L r)}{J_\nu(k_L a)} \right. \\ &\left. + \left(\nu^2 - \frac{k_T^2 a^2}{2} - B_L \right) B_T \frac{J'_\nu(k_T r)}{J'_\nu(k_T a)} \right\} \\ &\times \sum_{n=-\infty}^{+\infty} e^{j\nu 2n\pi}, \end{aligned} \tag{10}$$

where

$$\begin{aligned} D_\nu &= (\nu^2 - k_T^2 a^2/2)^2 - \nu^2 + k_T^2 a^2 (B_L + B_T)/2 \\ &\quad + (1-\nu^2) B_L B_T, \\ B_L &= k_L a J'_\nu(k_L a)/J_\nu(k_L a), \\ B_T &= k_T a J'_\nu(k_T a)/J_\nu(k_T a). \end{aligned} \tag{11}$$

Note that in Eqs. (7)–(9), $k_L = \omega\sqrt{\rho/c_{11}}$ and $k_T = \omega\sqrt{2\rho/(c_{11}-c_{12})}$ are the scalar wave vector of the longitudinal and transverse waves, respectively. Additionally, $\mathrm{sinc}(\nu\alpha) = \sin(\nu\alpha)/(\nu\alpha)$ is the sinc function, and $J'_\nu(x)$ is the derivative of the Bessel function $J_\nu(x)$.

Now, let us focus on the calculation of the integral in Eq. (8). When dealing with an elastic material, the integrand shows discontinuities for particular k_θ values. They correspond to poles associated with the zeros of the dispersion equation,

$$D_\nu = 0, \tag{12}$$

that describe the cylindrical Rayleigh waves[2] and Whispering Gallery waves.[3] The integration thus appears to be not consistent with the Fourier transformation. A suited numerical integration method should, therefore, be applied. For each value of the angular frequency ω, the integral on the

real axis of the variable k_θ is calculated by means of the method suggested by Weaver *et al.*[21] In this scheme, the Fourier transform is generalized by replacing ω by a complex variable $\omega - j\delta$ with a small, constant, and imaginary part δ. With this change of variable, Eq. (8) becomes

$$u_i(r,\theta,t) = \frac{e^{\delta t}}{4\pi^2} \int \int_{-\infty}^{+\infty} U_i(r,\nu,\omega - j\delta) \times e^{-j(\nu\theta - \omega t)} d\nu\, d\omega. \quad (13)$$

The benefit of this method is twofold: (i) it preserves the application of the fast Fourier transform algorithms for the final inversion, and (ii) the integrand is a nonsingular function that may now be integrated numerically. To perform the numerical integration, the value $\delta = 0.01$ rad.μs^{-1} has been chosen for the auxiliary parameter in the following numerical calculations.

IV. EXPERIMENTAL VALIDATION

Using the inverse scheme described in Sec. III, the normal component of the displacement can be calculated for either generation at the surface of various cylinders. The calculated normal displacements are compared to the experimental signals for two aluminum cylinders of radius 4.99 and 2.06 mm. A Nd:YAG laser is used for ultrasonic wave generation in either the ablation or thermoelastic regime. The pulse duration is 20 ns and infrared light emission is obtained at 1064 nm with a maximum burst energy output of 340 mJ. The collimated optical beam is focused by means of a cylindrical lens (the focus length is 150 mm). The line length and width are about 4 cm and 0.1 mm, respectively. Using an optical heterodyne probe with a power output of 100 mW and with a sensitivity[5] of 10^{-14} m.Hz$^{-1/2}$, the normal displacement is measured at the surface. See Ref. 22 for the detail experimental setup. In this paper, the observation angle is the difference between angular coordinates of the line source and point receiver in the cylindrical reference frame.

The experimental and calculated normal displacements generated by a laser line source of 20 ns duration and 0.1 mm beamwidth under ablation generation ($a = 4.99$ mm) and thermoelastic generation ($a = 2.06$ mm) are shown in Fig. 2 and Fig. 3, respectively. The observation angle is 180°. The aluminum density considered for calculations is $\rho = 2690$ Kg.m^{-3}, and the stiffness coefficients are $c_{11}^{\text{ref}} = 111.3$ GPa and $c_{12}^{\text{ref}} = 59.1$ GPa.[23] The calculated wave forms have been scaled vertically by a constant corresponding to the source strength to bring the amplitudes of the two signals into the same scale. In this paper, all the time scales in the figures are dimensionless variables obtained by dividing time by a factor t_L, the arrival time of direct longitudinal wave ($t_L = 2a/V_L$, where V_L is the velocity of longitudinal waves in the studied cylinder). This dimensionless time favors the comparison, since the time scale remains unchanged for the same material with a different radius.

As shown in Figs. 2–3, the calculated and experimental wave forms are in very good agreement for both generations. The time, shape, and relative amplitude of each arrival are identical. Especially, the arrivals of the first (R_1) and the second (R_2) roundtrip of the cylindrical Rayleigh waves and their dispersive behavior (the component of the low-frequency part travels relatively fast) are clearly observed in both wave forms. The arrival of the third (R_3) roundtrip cylindrical Rayleigh wave is also calculated and measured in the thermoelastic generation. It is clear that such a cylindrical Rayleigh wave has the similar shape and dispersive property as the spherical Rayleigh wave studied by Royer *et al.*[5] The experimental and calculated direct longitudinal wave (L) and the reflected transverse wave (TT) are observed under the ablation generation, whereas they are not observable under the thermoelastic generation. This phenomenon can be explained by the different directivities of the two generation.[4] Moreover, this difference of the directivity between ablation and thermoelastic generation can also explain the relative amplitude difference of the reflected longitudinal waves (LL)

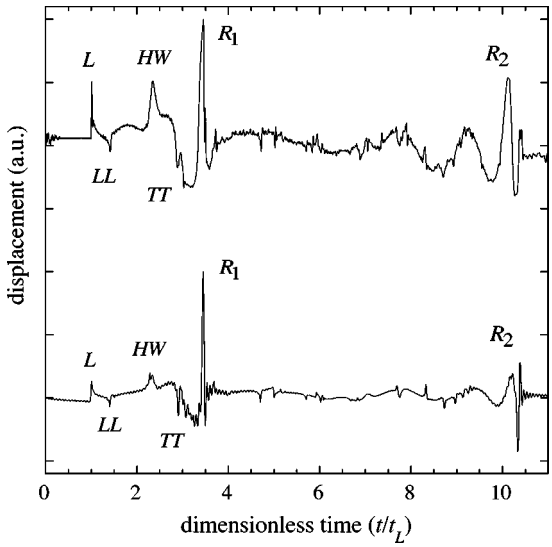


FIG. 2. Experimental (top) and calculated (below) normal displacements generated by a laser line pulse of 20 ns time duration and 0.1 mm beamwidth under the ablation regime at an observation angle of 180° for an aluminum cylinder of 4.99 mm radius.

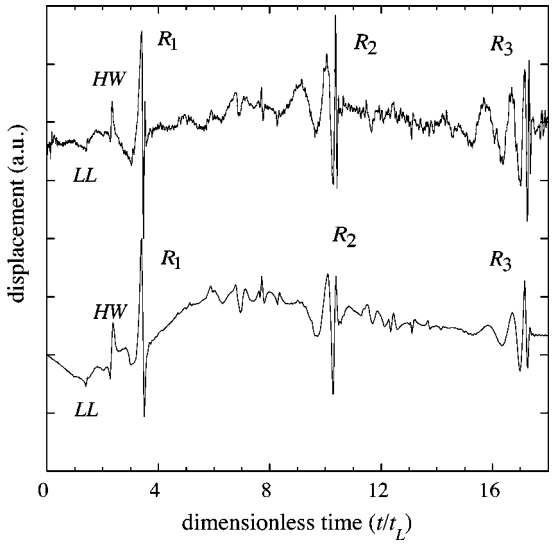


FIG. 3. Experimental (top) and calculated (below) normal displacements generated by a laser line pulse of 20 ns time duration and 0.1 mm beamwidth under the thermoelastic regime at an observation angle of 180° for an aluminum cylinder of 2.06 mm radius.

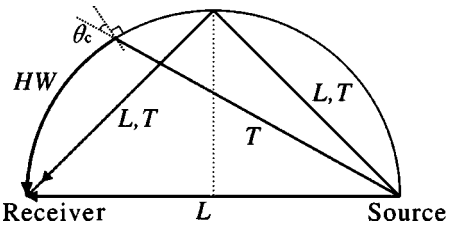


FIG. 4. Ray trajectories of *L*, *LL*, *TT*, and *HW* waves observed in Figs. 2–3. Here θ_c is the critical angle for aluminum rods.

and the head waves (*HW*) that are observed for both regimes. The ray trajectories of *L*, *LL*, *TT*, and *HW* waves are shown in Fig. 4. Following arrivals (not marked in Figs. 2–3) with small amplitude are bulk wave modes resulting from the multiple reflections of the longitudinal and transverse waves within the cylinder.[8] A comparison of the relative arrival times and amplitudes of these bulk waves emphasize the very good agreement between experiment and theory. Additionally, this comparison validates a more general calculation scheme than that already published,[10] since the time duration and beamwidth are taken into account here, whereas the Green's function was considered previously.

V. EFFECTS OF TIME DURATION AND BEAMWIDTH

Analyzing the effects of the source beamwidth and the time duration of the laser is necessary for the application purpose. As presented in Sec. IV, cylindrical Rayleigh waves, various longitudinal and transverse bulk waves are clearly observable in the experimental wave forms for two aluminum cylinders. If further attempting to recover elastic constants by measuring the time arrivals of selected waves, one possibly raises such a question as whether these waves are still observable for a very small cylinder. To respond this question, one must study their effects on these wave modes. In other words, one must understand how the shape and relative amplitude of a selected wave mode change along with the size of the cylinder.

For such an analysis, numerical simulation instead of experiment is preferred. The simulation can be performed on the present model. Even though these effects can be studied by experimenting on cylinders with radii covering a wide range, simulating is relatively direct, complete, and clear. The numerical simulations are carried in two folds: (i) study the effect of the source time duration by varying the radius of the cylinder while excluding the effect of the source beamwidth; and (ii) study the effect of the source beamwidth by varying the radius of the cylinder while excluding the effect of the source time duration.

A. Source time duration

First let us analyze the effect of the source time duration. Since a laser pulse generally has a fixed time duration τ, this effect is analyzed at different τ/t_L ratios by varying the radius of the cylinder. The source beamwidth is assumed to be zero to exclude its effect. The low limit of τ/t_L is zero, and its upper limit is estimated as two if supposing no bulk wave is observed when half the equivalent wavelength $\lambda_\tau = \tau V_L$ is greater than $2a$, the diameter of the cylinder. Applying the

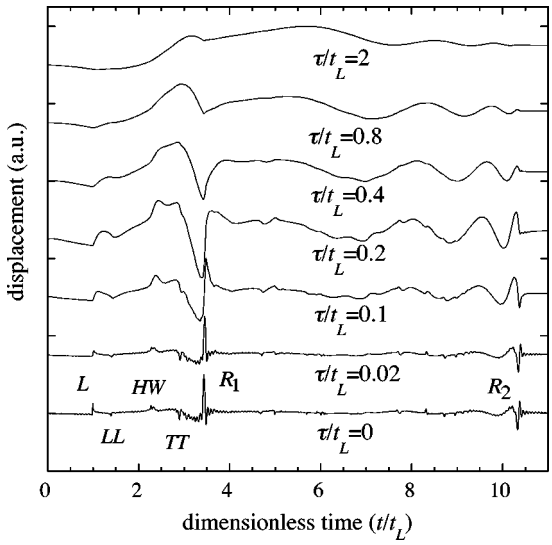


FIG. 5. Calculated normal displacements at an observation angle of 180° for aluminum cylinders under the ablation generation for different ratios of the source time duration τ over t_L the arrival time of the direct longitudinal wave.

numerical inverse scheme described in Sec. III, normal displacements at an observation angle of 180° for aluminum cylinders under either ablation or thermoelastic generation are calculated for τ/t_L ratios of 0, 0.02, 0.1, 0.2, 0.4, 0.8, and 2. Results are shown in Figs. 5–6, where the wave forms are vertically scaled to obtain the same maximum amplitude for each of them. To fix ideas, the ratio between maxima obtained for τ/t_L equal 0 and 2, is 26 and 2 for Fig. 5 and Fig. 6, respectively. Under both regimes, the surface and various bulk waves are losing more of their relatively high-frequency components, and their amplitudes are becoming flatter until they have completely disappeared, when ratios τ/t_L increase from 0 to 2. Since the time duration of the pulse acts as a low-pass filtering with its bandwidth inversely proportional to τ, it cuts off more of their relatively high-frequency parts

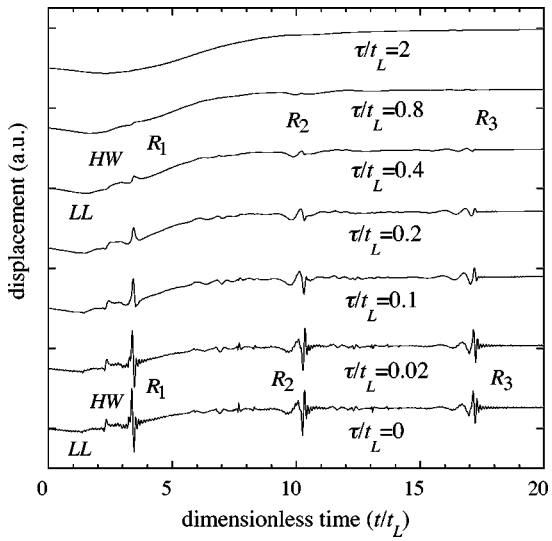


FIG. 6. Calculated normal displacements at an observation angle of 180° for aluminum cylinders under the thermoelastic generation for different ratios of the source time duration τ over t_L the arrival time of the direct longitudinal wave.

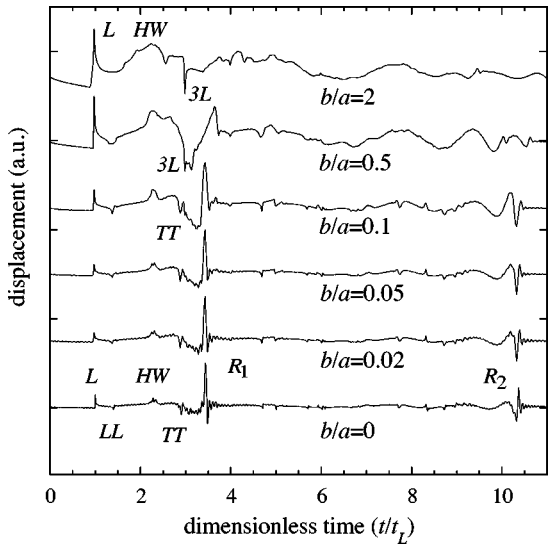


FIG. 7. Calculated normal displacements at an observation angle of 180° for aluminum cylinders under the ablation generation for different ratios of the source beamwith b over the cylinder radius a.

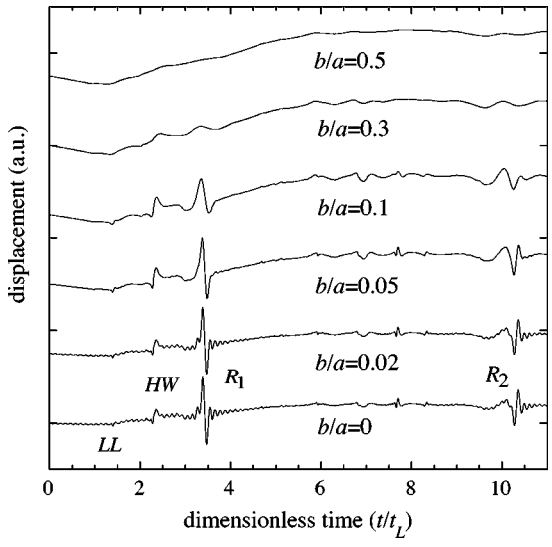


FIG. 8. Calculated normal displacements at an observation angle of 180° for aluminum cylinders under the thermoelastic generation for different ratios of the source beamwith b over the cylinder radius a.

when τ increases. Owing to this filtering, two successive echoes cannot be distinguished if the delay between their arrivals is less than the time duration. Considering the direct longitudinal L and once reflected longitudinal LL (the time arrival is $t_{LL}=\sqrt{2}t_L$) waves, it comes that they cannot be separated if τ/t_L is greater than 0.4. This is in agreement with the evolution of the wave forms plotted in Figs. 5–6. In other words, if time arrivals are required for nondestructive evaluation (NDE) purposes, cylinders with a radius less than $1.25\tau V_L$ should not be chosen. For a laser source of 20 ns time duration, this radius limit is 0.16 mm for an aluminum rod.

B. Source beamwidth

Now let us analyze the effect of the source beamwidth. As a finely focused laser generally has a certain beamwidth b, this effect is analyzed at different b/a ratios by varying the radius of the cylinder. The low limit of b/a ratio is zero, and its upper limit is two. Here the source time duration is assumed to be zero to exclude its influence. As shown in Fig. 7 and Fig. 8, normal displacements at an observation angle of 180° for aluminum cylinders are calculated for b/a ratios of 0, 0.02, 0.05, 0.1, 0.5, and 2 under ablation generation and of 0, 0.02, 0.05, 0.1, 0.3, and 0.5 under thermoelastic generation. The wave forms are also vertically scaled. The ratio between maxima obtained for b/a equal 0 and 2 is 9 for Fig. 7, and this ratio for b/a equal 0 and 0.5 is 3 for Fig. 8. In the ablation generation, Fig. 7, the amplitude of the direct longitudinal wave (L) and that of the directly reflected wave ($3L$) propagating back and forth along the cylinder diameter increase when b/a changes from 0 to 2, and no filtering occurs. Relative amplitudes of other bulk waves and surface waves flatten, until they completely disappear. However, HW arrivals are still observable. Increasing the source width acts as a low-pass filtering in the ν domain. As detailed in Ref. 24, this filtering changes the source directivity, and enlarging the source size favors the generation of longitudinal waves in a direction normal to the interface for this generation. In the thermoelastic generation, Fig. 8, since the source directivities of two generations are different, LL waves are observable whereas L waves are no more distinguishable. Wave forms are therefore more affected by increasing source size in this generation than in the ablation generation. Consequently, bulk LL and HW waves are not observable when b/a reaches 0.5, but they are slightly affected by the source directivity for ratio b/a up to 0.3. For a laser source of 0.1 mm beamwidth, the cylinder radius limit is about 0.3 mm for NDE purposes.

When experimental signals are considered, both effects should be considered simultaneously. Therefore, for NDE application using bulk waves, a radius of 0.3 mm appears as a minimum limit for the sample size with our experimental devices.

VI. APPLICATION TO THE MEASUREMENT OF ELASTIC CONSTANTS

As an inverse problem, it is generally considered to recover the elastic constants by the time arrivals of acoustic waves detected through the laser ultrasonic system. As described in Sec. II, two independent elastic constants, c_{11} and c_{12}, are related to the isotropic plane of a transversely isotropic cylinder, and they link with the longitudinal and transverse wave velocities V_L and V_T by the following equation:

$$c_{11}=\rho V_L^2$$
$$c_{12}=\rho(V_L^2-2V_T^2). \tag{14}$$

Since the mass density ρ and radius a of the tested cylinder can be easily measured, these two velocities V_L and V_T should be obtained at the same time for the measurement. Now let us determine how to measure the two velocities under either ablation or thermoelastic generation.

First let us look at the experimental waveform under ablation generation in Fig. 2. The longitudinal wave velocity V_L can be determined by measuring the arrival time of the direct longitudinal wave (L), or the reflected longitudinal

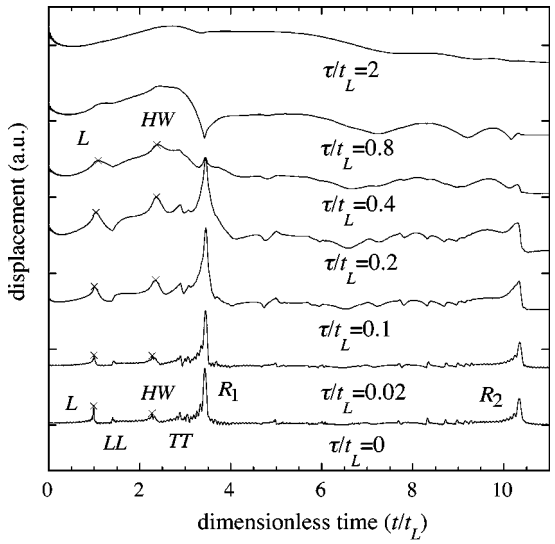


FIG. 9. Moduli of the analytic signals associated to the calculated normal displacements in Fig. 5. Crosses show the positions of the maximum moduli considered for the wave arrivals.

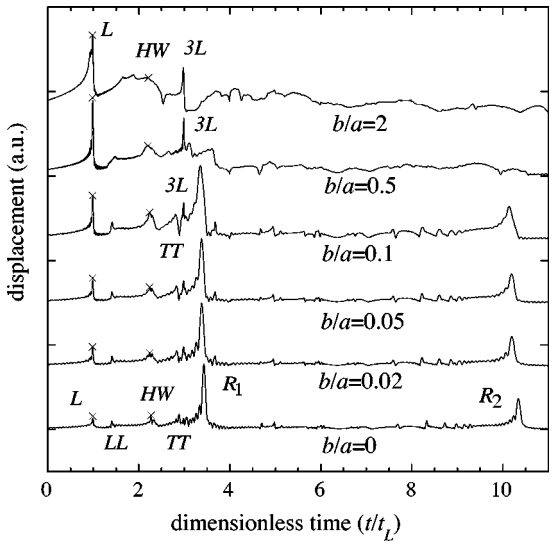


FIG. 10. Moduli of the analytic signals associated to the calculated normal displacements in Fig. 7. Crosses show the positions of the maximum moduli considered for the wave arrivals.

wave (*LL*). The transverse wave velocity V_T can be determined by measuring the arrival time of the reflected transverse wave (*TT*), or the head wave (*HW*) after V_L was determined. The arrival time of head wave t_{HW} links with V_L and V_T by the formulas $t_{HW}=t_L(\beta+\cot\beta)$ and $\beta=\sin^{-1}(V_T/V_L)$. Owing to their relative amplitudes, *L* and *HW* waves are chosen for the measurement. Although the cylinder Rayleigh wave has a high amplitude, it is not considered here since it is dispersive.[5] The following bulk waves are also not considered as they have relative low amplitude. Then let us look at the experimental wave form under thermoelastic generation in Fig. 3. The longitudinal wave velocity V_L can be determined by measuring the arrival time of the reflected longitudinal wave (*LL*). The transverse wave velocity V_T can be determined by measuring the arrival time of the head wave *HW* after V_L was determined. The cylinder Rayleigh wave is also not considered for the same reason as for the ablation generation. Following bulk waves, having undergone several reflections with or without mode conversion at the cylinder interface, could also be used since their amplitude can be high. However, they are not considered here since recovering of their exact paths is not a trivial matter. The above measurement schemes of both regimes only hold true for experiments at an observation angle 180°. If another observation angle is chosen as presented in the experimental setup of Clorennec and Royer,[15] a different scheme can be found after acoustic waves are identified by the published model[10] or the model presented in Sec. II.

It is also direct and complete to analyze the effects of source time duration and beamwidth on the measurement by processing the calculated wave forms. First let us look at the effect of time duration under ablation generation. In order to determine the arrival time of selected waves in the calculated wave forms in Fig. 5 of the last section, a signal processing published by one of the authors[25] is applied to locate the wave arrivals. The moduli of the analytic signals[25] associated to wave forms in Fig. 5 are shown in Fig. 9. Numerical interpolation is further adopted to precise the maximum positions corresponding to the arrivals of *L* and *HW* waves. These positions are marked with crosses in Fig. 9. The two elastic constants are calculated from t_L the arrival time of the *L* wave and t_{HW} the arrival time of the *HW* wave, and they are compared to the reference values in Sec. IV used to calculate the wave forms. As reported in Table I, their deviations increase as τ/t_L turns to 0.4, and it is difficult to determine the elastic constants when τ/t_L is greater than 0.4. Then let us look at the effect of beamwidth under ablation generation. As shown in Fig. 10, applying the same signal processing yields the moduli of the analytic signals associated to wave forms in Fig. 7. The determined elastic constants are

TABLE I. The deviations of the elastic constants determined from the calculated wave forms under ablation (Fig. 5) and thermoelastic (Fig. 6) regimes for various τ/t_L ratios.

	Ablation regime		Thermoelastic regime	
t/t_L	$(c_{11}-c_{11}^{\text{ref}})/c_{11}^{\text{ref}}$	$(c_{12}-c_{12}^{\text{ref}})/c_{12}^{\text{ref}}$	$(c_{11}-c_{11}^{\text{ref}})/c_{11}^{\text{ref}}$	$(c_{12}-c_{12}^{\text{ref}})/c_{12}^{\text{ref}}$
0	0.3%	−2.0%	−0.9%	−1.5%
0.02	0.5%	−5.2%	1.4%	1.5%
0.1	−1.3%	−6.9%	−6.7%	−8.4%
0.2	−7.9%	−13.1%	−9.6%	−13.1%
0.4	−16.0%	−20.8%	−9.7%	−13.5%

TABLE II. The deviations of the elastic constants determined from the calculated wave forms under ablation (Fig. 7) and thermoelastic (Fig. 8) regimes for various *b/a* ratios.

	Ablation regime		Thermoelastic regime	
b/a	$(c_{11}-c_{11}^{\text{ref}})/c_{11}^{\text{ref}}$	$(c_{12}-c_{12}^{\text{ref}})/c_{12}^{\text{ref}}$	$(c_{11}-c_{11}^{\text{ref}})/c_{11}^{\text{ref}}$	$(c_{12}-c_{12}^{\text{ref}})/c_{12}^{\text{ref}}$
0	0.3%	−2.0%	−0.9%	−1.5%
0.02	3.5%	0.8%	1.5%	1.7%
0.05	3.5%	0.7%	1.5%	1.8%
0.1	3.5%	0.6%	1.3%	1.5%
0.3	⋯	⋯	1.6%	−4.9%
0.5	3.5%	−3.9%	⋯	⋯
2	3.3%	−2.6%	⋯	⋯

TABLE III. Elastic constants measured for aluminum rods of various radii.

Regimes	a (mm)	b/a	τ/t_L	c_{11} (GPa)	c_{12} (GPa)
Ablation (Fig. 2)	4.99	0.02	0.013	110	61
Thermoelastic (Fig. 3)	2.06	0.05	0.032	111	58
Thermoelastic	1.00	0.1	0.064	110	61
Thermoelastic	0.50	0.2	0.128	113	62

also compared to the reference values. As listed in Table II, their deviations are of the same level when b/a turns to 2. It could be concluded that the laser beamwidth does not affect the elastic constant measurement for ablation generation. As also reported in Table I, the effect of the source time duration under thermoelastic generation is of the same order of magnitude as under ablation generation, and similar comments can be suggested. But the effect of the source beamwidth under thermoelastic generation is a little different from that under ablation generation. As also listed in Table II, the bulk waves are not measurable when b/a reaches 0.5, because the directivities of two regimes are different.[4]

Finally, let us carry the same signal processing on experimental signals under either regime shown in Figs. 2–3. Additionally, two aluminum rods with radii of 1.00 and 0.50 mm are further chosen for the measurement under a thermoelastic regime. The measured elastic constants are reported in Table III. These values are very close to that in the reference.[23] The deviations for the measured elastic constant c_{11}, estimated from the simulation results in Tables I and II, are 4%, 6%, 7%, and 10% for the aluminum rods of radii 4.99, 2.06, 1.00, and 0.50 mm, respectively. The corresponding deviations for another elastic constant c_{12} are estimated to be 5%, 7%, 8%, and 13%. Here, the relatively large deviation for a pair of τ/t_L and b/a ratios is assumed to take into account of both effects of the source time duration and beamwidth. These deviations remain low, even for the cylinder whose radius is close to the minimum limit estimated in Sec. V. Actually, these deviations will be much smaller if the measured elastic constants are compared to the reference values.

It can be concluded that the waves are correctly identified, and that the method used to measure the time arrivals is suitable for such a purpose of elastic constants recovery.

VII. CONCLUSION

A theoretical model has been presented to predict the acoustic field generated by a laser line pulse extended in the beamwidth and time duration under either the ablation or thermoelastic regime at any point of a homogeneous and transversely isotropic cylinder. Experimental and theoretical normal displacements under either regime were obtained and compared for aluminum cylinders. Very good agreements are observed in the time, shape and relative amplitude (i) of the cylindrical Rayleigh waves with different roundtrips, and (ii) of the various longitudinal and transverse bulk waves propagating through the cylinder or reflected at the free circular surface.

The effects of the source beamwidth and time duration have been analyzed by varying the radius of the cylinder. The source time duration under either regime acts as a low-pass filtering with its bandwidth inversely proportional to its width in time. The source beamwidth affects the directivities of both longitudinal and transverse waves for either regime. Enlarging the source size favors the generation of longitudinal waves in a direction normal to the interface for ablation generation. For NDE applications using bulk waves, a radius of 0.3 mm appears as a minimum limit for the sample size using a laser of 0.1 mm beamwidth and 20 ns time duration.

To recover the elastic constants, the velocity of the longitudinal wave could be first determined by measuring the arrival time of the direct longitudinal wave for ablation generation or that of the reflected longitudinal wave for thermoelastic generation, and then the velocity of the transverse wave could be determined by measuring the arrival time of the head wave for either generation. The source beamwidth slightly affects the elastic constant measurement, whereas the measurement error enlarges as the source time duration increases. Such a measurement scheme including the signal processing is further justified by results on experimental wave forms detected on aluminum rods under ablation and thermoelastic generations.

ACKNOWLEDGMENTS

This work was done at the Laboratoire de Mécanique Physique, Université Bordeaux 1. One of the authors (Y. P.) was supported by the National Center for Scientific Research of France (CNRS) and the China Scholarship Council.

[1] K. Sezawa, ''Dispersion of elastic waves propagated on the surface of stratified bodies and on curved surfaces'' Bull. Earthquake Res. Inst., Univ. Tokyo **3**, 1–18 (1927).
[2] A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[3] H. Uberall, *Physical Acoustics* (Academic, New York, 1973), Vol. 10, p. 1–60.
[4] C. B. Scruby and L. E. Drain, *Laser Ultrasonics: Techniques and Applications* (Adam Hilger, New York, 1990).
[5] D. Royer, E. Dieulesaint, X. Jia, and Y. Shui, ''Optical generation and detection of surface acoustic waves on a sphere,'' Appl. Phys. Lett. **52**, 706–708 (1988).
[6] S. Ishikawa, H. Cho, Y. Tsukahara, N. Nakaso, and K. Yamanaka, ''Analysis of spurious bulk waves in ball surface wave device,'' Ultrasonics **41**, 1–8 (2003).
[7] M. Qian and X. Wu, ''Studies of cylindrical Rayleigh waves with laser ultrasonics,'' Prog. Nat. Sci. **11**, s258–s264 (2001).
[8] D. Clorennec and D. Royer, ''Analysis of surface acoustic wave propagation on a cylinder using laser ultrasonics,'' Appl. Phys. Lett. **82**, 4608–4610 (2003).
[9] X. Wu and M. Qian, ''Simulation of the finite element method on wave propagation in cylinders,'' Prog. Nat. Sci. **11**, s265–s268 (2001).
[10] Y. Pan, C. Rossignol, and B. Audoin, ''Acoustic waves generated by a laser line pulse in a transversely isotropic cylinder,'' Appl. Phys. Lett. **82**, 4379–4381 (2003).
[11] F. Honarvar and A. N. Sinclair, ''Nondestructive evaluation of cylindrical components by resonance acoustic spectroscopy,'' Ultrasonics **36**, 845–854 (1998).

[12] Y. A. Burenkov and S. P. Nikanorov, ''Elastic properties and binding forces in crystals with diamond and sphalerite lattices,'' Sov. Phys. Solid State **26**, 1940–1944 (1984).
[13] C. P. Lusher and W. N. Hardy, ''Axisymmetric free vibrations of a transversely isotropic finite cylindrical rod,'' J. Appl. Mech. **55**, 855–862 (1988).
[14] M. Qian, M. Duquennoy, M. Ouatouh, F. Jenot, and M. Ourak, ''Laser ultrasonic characterization of surface residual stresses in steel rods,'' Rev. Prog. Quant. Nondestr. Eval. **20**, 1487–1493 (2001).
[15] D. Clorennec, D. Royer, and H. Walaszek, ''Nondestructive evaluation of cylindrical parts using laser ultrasonics,'' Ultrasonics **40**, 783–789 (2002).
[16] A. Rahman and F. Ahmad, ''Representation of the displacement in terms of scalar functions for use in transversely isotropic materials,'' J. Acoust. Soc. Am. **104**, 3675–3676 (1998).
[17] L. R. F. Rose, ''Point-source representation for laser-generated ultrasound,'' J. Acoust. Soc. Am. **75**, 723–732 (1984).
[18] D. A. Hutchins, *Physical Acoustics* (Academic, New York, 1988), Vol. 18, pp. 21–123.
[19] Y. H. Berthelot and I. J. Busch-Vishniac, ''Thermoacoustic radiation of sound by a moving laser source'' J. Acoust. Soc. Am. **81**, 317–327 (1987).
[20] C. B. Scruby, R. J. Dewhurst, D. A. Hutchins, and S. B. Palmer, ''Quantitative studies of thermally generated elastic waves in laser-irradiated metals,'' J. Appl. Phys. **51**, 6210–6216 (1980).
[21] R. L. Weaver, W. Sachse, and K. Y. Kim, ''Transient elastic waves in a transversely isotropic plate,'' J. Appl. Mech. **63**, 337–346 (1996).
[22] F. Reverdy and B. Audoin, ''Elastic constants determination of anisotropic materials from phase velocities of acoustic waves generated and detected by lasers,'' J. Acoust. Soc. Am. **109**, 1965–1972 (2001).
[23] O. L. Anderson, *Physical Acoustics* (Academic, New York, 1965), Vol. 3, pp. 43–95.
[24] J. D. Aussel, A. Le Brun, and J. C. Baboux, ''Generating acoustic waves by laser: Theoretical and experimental study of the emission source,'' Ultrasonics **26**, 245–255 (1988).
[25] B. Audoin, C. Bescond, and M. Deschamps, ''Measurement of stiffness coefficients of anisotropic materials from pointlike generation and detection of acoustic waves,'' J. Appl. Phys. **80**, 3760–3771 (1996).

A theoretical model for a finite-size acoustic receiver

Adrian Neild[a] and David A. Hutchins
School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom

(Received 18 April 2003; accepted for publication 1 December 2003)

A new formulation is presented for the prediction of the signals transmitted by a rectangular acoustic source, and detected by a finite-sized rectangular receiver. This problem is encountered in measurements where the receiver is of a finite size compared to the shortest wavelength emitted by the source. The paper examines this problem theoretically, and a new formulation is presented in which the geometrical aspects are considered in some detail. The result is a simplified approach, in which the overall impulse response can be obtained from a combination of certain contributions defined geometrically. Theoretical field predictions are compared to experimental measurements for ultrasonic signals radiated into air, to demonstrate the usefulness of the approach. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1646406]

PACS numbers: 43.35.Yb [JGH] Pages: 1546–1556

I. INTRODUCTION

The standard approach to the characterization of the field radiated by an acoustic source is to scan a miniature detector throughout the field, and to plot spatial variations in the received amplitude. In the ultrasonic case, the receiver can take many different forms, including miniature piezoelectric hydrophones,[1] or by using a thin polymer piezoelectric membrane.[2] More accurate calibration work uses a metal-coated thin membrane and laser interferometry.[3] In all cases, however, the basic assumption is that the receiver diameter is small compared to the shortest wavelength likely to be encountered.

It is often the case that the receiver does not conform to the above criteria. In such cases, the directivity pattern of the receiver will not be that of a point receiver, and hence the measured field from the source will deviate from that which actually exists. The waveform detected at a particular field location will also be recorded incorrectly. It is therefore of interest to be able to predict the effect of a finite-sized receiver on the measured field from a source. These effects have to be taken into account in many acoustic measurements. For instance, a ''diffraction correction''[4] is often applied to attenuation measurements, to account for the effects of the finite size of the transducers used. This can be applied at a specific frequency, for cases where two circular transducers are moved apart axially. Such a situation has been modeled using the Lommel diffraction correction integral.[5,6] In addition, the transient case of coaxially aligned circular transducers has been considered.[7] However, there does not appear to be a time-domain solution in the literature for the general case where a finite-size receiver is moved throughout the field of a source under transient excitation. In this paper, a new formulation for the pressure signal measured by a finite rectangular receiver will thus be derived.

Harris states that it is common practice to equate the received signal to the average pressure over the receiver area.[8] This is taken as the basis for the model presented here. The assumption is made that the scalar impulse response can be integrated over the face of the receiver, and that this total system impulse response can be convolved with the time differential of the piston velocity waveform, to give the pressure waveform. The face and edges of the receiver are assumed to be aligned with those of the source. In addition, it is assumed that the propagation medium is a nondissipative and homogeneous fluid. The source and receiver are assumed to be located within rigid baffles, with multiple reflections being ignored. The present finite-receiver model is developed by deconstructing a model proposed by San Emeterio and Ullate[9] which finds the pressure waveform at a point in space from a rectangular receiver, and readdressing the problem in such a way that the integral over the receiver area can be performed. In the following, the model is verified by using it to produce simulated peak-to-peak pressure fields, which are compared to experimentally obtained results.

II. THE THEORETICAL APPROACH

A. Background and geometrical considerations

The solution to the general problem requires the impulse response from the source to be integrated over the area of the receiver. This is achieved here by splitting the source into a series of arc-limited, edge-limited, and planar integrals. These can be integrated over the area of the receiver, and combined to provide a general integral result. Once the receiver has been segmented so that each segment falls into only one region, the source can then be reconstructed by defining the variables in the general integral correctly, achieved by rotating the coordinate system of the general integral to the edge and point being considered. By stepping in time, and adding the effect of each segment, a scalar impulse response for the source and finite-receiver system can be found. This can be convolved with the time differential of the velocity response of the plane piston source to give a predicted pressure waveform.

The impulse response model is based on the Rayleigh equation,[10] and has been used extensively to model the pressure field from circular[11] and rectangular[9,12,13] plane piston

[a]Electronic mail: adrian.neild@imes.mavt.ethz.ch. Present address: Institute of Mechanical Systems Center of Mechanics, ETH Zürich CH-8092 Zürich, Switzerland.

 0001-4966/2004/115(4)/1546/11/$20.00

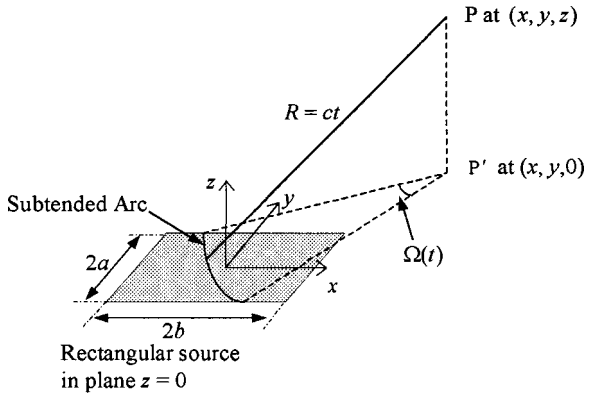


FIG. 1. Arc subtended at the projection of point P(x,y,z) onto the plane of the rectangular piston source.

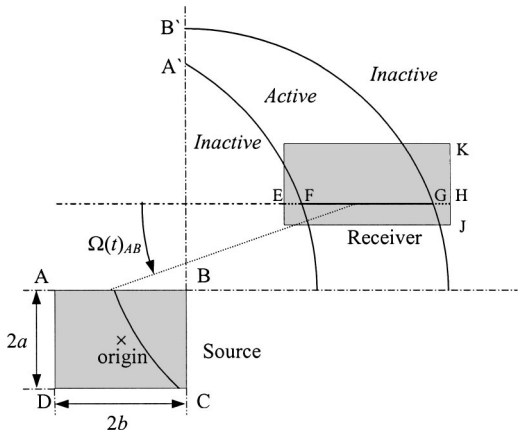


FIG. 2. Illustration of region segmentation for edge AB of the source.

transducers. Harris[8] states that a convolution method can be used to find the pressure at a point in a field, using the equation

$$p(\bar{r},t)=\rho_0\frac{\partial}{\partial t}[v(t)\times h(\bar{r},t)], \tag{1}$$

where $\bar{r}$ defines the position of the field point P(x,y,z), relative to an origin at the center of the rectangular source, $v(t)$ is the velocity waveform of the plane piston, and $h(\bar{r},t)$ is the scalar response due to an impulse. For a source surrounded by an infinite rigid baffle, the Rayleigh integral yields

$$h(\bar{r},t)=\int_s\frac{\delta(t-R/c)}{2\pi R}dS, \tag{2}$$

where R is the distance from the point to the surface element δs on the piston face $(R=ct)$, and c is the speed of sound. San Emeterio and Ullate[9] states that this reduces to

$$h(\bar{r},t)=\frac{c\cdot\Omega(t)}{2\pi}. \tag{3}$$

Here, $\Omega(t)$ is the angle of the arc which intersects the plane piston, as shown in Fig. 1, where the point $P'(x,y,0)$ is the projection of P(x,y,z) onto the plane containing the piston (which is the $z=0$ plane). The origin is at the center of the rectangular piston source, with four corners labeled A–D as shown, of size $2a\times 2b$. Hence, the problem is reduced to one of geometry, where $\Omega(t)$ must be found at sequential time steps. The problem need only be considered in two dimensions, if the time is redefined as an equivalent time, referred to as planar time (t_p) within the $x-y$ plane according to

$$t_p=\frac{\sqrt{c^2t^2-z^2}}{c}. \tag{4}$$

A model has been developed by San Emeterio and Ullate[9] which finds the angle $\Omega(t)$ at any given position of P(x,y,z). This approach treats the source as a whole, and the equations used at any given moment in time depend on the combination of edges, which are crossed by the arc. If the region in Fig. 1, defined by $x\geqslant b$ and $y\geqslant a$, is considered, then depending on the location of P′ within this region different equations to calculate $\Omega(t)$ would apply. However, it is possible to treat each edge separately.[14] This approach has major advantages when applied to a finite-receiver model. The difference in the two approaches is that when the source is treated as a whole, the angle $\Omega(t)$ is found directly for each moment in time using a combination of equations depending on location of P′. By treating each edge separately, a single equation can be used to calculate $\Omega(t)_{\text{edge}}$ for each moment in time, prior to a summation of these strings of values to give $\Omega(t)$. The angle $\Omega(t)_{\text{edge}}$ is defined as the angle from the point of intersection of the arc with the edge to a fixed reference line; this is shown in Fig. 2 for edge AB. The terms *active* and *inactive* will be used to describe certain regions of the plane containing the source (the $z=0$ plane). A reception point is considered to be *active* if, at time point t_p, a section of an arc of radius ct_p centered at that point lies within the source, so that $\Omega(t)_{\text{edge}}\neq 0$ (i.e., there is a contribution to the impulse response). Conversely, for any point lying within an *inactive* area, $\Omega(t)_{\text{edge}}=0$. The source has corners labeled ABCD, and arcs centered at each corner of the source and of radius ct_p are labeled A′ and B′. The rectangular receiver is then split into active and inactive regions. One of the advantages of this approach compared to treating the source as a whole is that, for any point in the active part of the region defined by $x\geqslant b$ and $y\geqslant a$, only one equations applies, and this is true for any of the edges.

The second advantage of using the new approach is that the angle $\Omega(t)_{\text{edge}}$ remains constant for all points within the active region along any line parallel to the edge being considered, so the integration over the area of the receiver can be reduced to an integral in either the x or y direction, depending on which edge is being considered. This is again shown at a particular point in time (t_p) in Fig. 2 for edge AB, where it can be seen that the required angle $\Omega(t)_{AB}$ is the same for all points on the line FG. Similarly for edge AD, the angle would be constant for each point along any line parallel to JK.

B. The method of integration

In order to find the net pressure waveform arriving at a finite receiver, it is now necessary to integrate over the receiver area. This proposition becomes possible when each

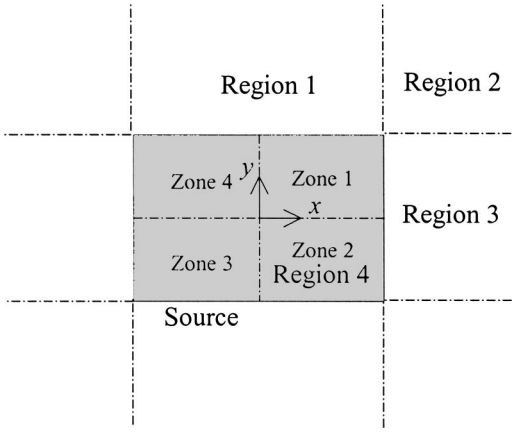


FIG. 3. Definition of regions with respect to the rectangular source.

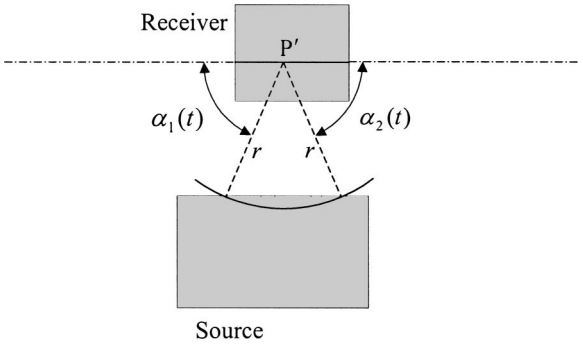


FIG. 4. Demonstration of the need to define two angles in an arc-limited integral.

angle relating to a particular edge is considered separately. Figure 2 illustrated how these angles remain constant along lines parallel to the edge under consideration (in this case AB). Hence, integration of $\Omega(t)_{AB}$ over the receiver area, in the example shown in Fig. 2, would correspond to an integral of $[\Omega(t)_{AB} \times \mathrm{FG}]$ over the distance from J to K.

The problem, however, has to be considered in a wider context, in that the integration across the receiver varies according to the position of the receiver relative to the source. Figure 3 shows a diagram of the plane $z=0$, containing the source. The area outside the source has been subdivided as shown into four primary regions, noting that the unlabeled areas are equivalents of regions 1–4. When the receiver geometrically is entirely within each region, a particular approach to integration is taken. These are described below. In the more usual case where the receiver overlaps more than one region, the integration is performed separately for that part of the receiver within each region, and the resultant components summed to get the total integral.

Consider first a receiver totally within region 2. In the example shown earlier in Fig. 2, the receiver is crossed by both arcs A′ and B′. The procedure is to integrate for arcs B′ and A′ separately, and subtract the two, remembering that only the active area between the two arcs actually contributes to the integral. We call this process "arc-limited integration." Various geometrical factors can be used to perform the integration. These are given in detail in the Appendix.

For a receiver completely within region 1 or region 3 of Fig. 3, a different approach has to be considered. The projection of the receiver onto the $z=0$ plane will be alongside two edges of the source, as shown in Fig. 4. In this case, in order to find the effect of such an edge it will be necessary to define the "edge-limited integral." If the point P′ is considered, then the arc centered at P′ crosses the nearest edge twice at the particular instance in time shown. Hence, it is now necessary to consider two angles, α_1 and α_2, which can be dealt with separately. The angles have been labeled with the letter α as they are with respect to a line parallel to the edge, and depending on which edge is being considered this is not necessarily the same as the reference line (in which case the angle is labeled Ω). The method of calculation for these edge-limited integrals is also included in the Appendix. Also described is a final integration, required when the projection of the receiver onto the $z=0$ plane containing the source is such that the two overlap, i.e., when the receiver projection intrudes onto the source in one of the four zones shown earlier in Fig. 3. In this case, an additional form of integration (the form of which we call the "planar integral") must be considered, as again described in the Appendix.

C. The general solution

As stated earlier, the final method of solution for a particular size and location of the receiver consists of working out the projection of the receiver onto the $z=0$ plane, and then performing the integration. The precise form of integration needed will depend on which of the three types of integral described above are required. Figure 5 shows a typical example, where in this case the receiver overlaps the source when the former is projected onto the $z=0$ plane. The location of the receiver is expressed as the coordinate x_r, y_r, and the sizes of the source and receiver are $2b \times 2a$ and $2e \times 2d$, respectively. In the example shown in Fig. 5(a), the receiver falls into four regions (regions 1 to 4), and hence is divided into four segments shown in Figs. 5(b)–(e) [shown at half the scale of Fig. 5(a)] which can then be considered separately. The four edges of the source have been labeled edge 1 to edge 4, respectively.

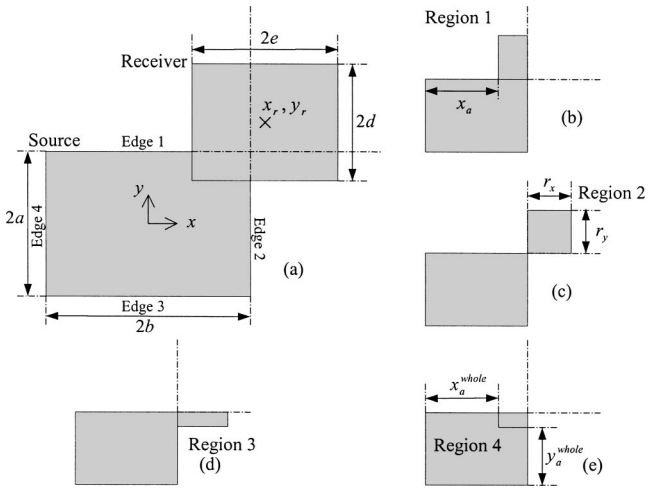


FIG. 5. An illustration of segmentation of the receiver. In (a), the whole receiver is shown overlapping with the source. (b), (c), (d), and (e) then demonstrate how the problem is considered as four separate parts.

After segmentation, the parts of the receiver are considered separately for each sector. By using the same terms to describe the geometry of the receiver segment in each sector, any computer program written to numerically perform this model is simplified, as it could loop through the sectors simply changing the values of these terms.[14] The size of the receiver segment is defined using the terms r_x in the x direction, and r_y in the y direction. The location of the segment is defined using x_a and x_b as, respectively, the distance (in the x direction) from edge 4 and edge 2 of the source to the nearest edge of the receiver. Similarly, y_a and y_b are, respectively, the distance (in the y direction) from edge 3 and edge 1 of the source to the nearest edge of the receiver. Hence, for the example given in Fig. 5(b), x_a is shown, y_a is $2a$, x_b and y_b are 0.

With knowledge of the geometrical variables above for a particular case, a general solution to the problem can now be formed. This requires integration of the scalar impulse response in terms of the $\Omega(t)$ [given earlier in Eq. (3)] over the area of the receiver. In order to average the impulse response, $h(\bar{r},t)$, over the area of the receiver, Eq. (3) must be amended so that the subtended arc, $\Omega(t)$, is integrated over the area of the receiver for each successive time point and by dividing by the area of the receiver, giving

$$h(\bar{r},t)=\frac{\mathrm{c}}{2\pi\cdot R_{DA}\cdot R_{DC}}\cdot\int\int\Omega(\bar{r},t)dv_1\,du_1. \tag{5}$$

By considering the angle to each edge separately, the result for the arc-limited and edge-limited integrals of the Appendix is reduced to a line integral. The result will be termed $h_{\mathrm{general}}(\bar{r},t)$, and is given at any given t_p in the form

$$h_{\mathrm{general}}(\bar{r},t)=\frac{\mathrm{c}}{2\pi\cdot R_{DA}\cdot R_{DC}}\cdot\int_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}\Omega_{\mathrm{edge}}(t)\cdot v_1\,du_1. \tag{6}$$

It can be seen from Fig. 14(b) in the Appendix that an angle relating to an edge can be found using $\alpha_{\mathrm{edge}}(t)=\sin^{-1}(u_1/r)$, where $r=ct_p$, and this angle must be with reference to a fixed line so that the angles can be added together without error. Consequently, the constants C_a and C_b are used, to relate $\Omega_{\mathrm{edge}}(t)$ to $\alpha_{\mathrm{edge}}(t)$, giving

$$\Omega_{\mathrm{edge}}(t)=C_a+C_b\sin^{-1}(u_1/r). \tag{7}$$

In addition, a way must be found to express v_1 in Eq. (6) for any case, and to this end the constants C_c and C_d are introduced. The generalized form becomes

$$h_{\mathrm{general}}(\bar{r},t)=\frac{\mathrm{c}}{2\pi\cdot R_{DA}\cdot R_{DC}}\cdot\int_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}(C_a+C_b\cdot\sin^{-1}\times(u_1/r))\cdot(C_c\cdot\sqrt{r^2-u_1^2}+C_d)du_1, \tag{8}$$

and hence one integral will be valid for both the corner and edge contributions, over all time intervals, provided the correct values of C_a, C_b, C_c, C_d are used. This becomes more manageable by multiplying out the brackets, and separating, such that

$$h_{\mathrm{general}}(\bar{r},t)=\frac{\mathrm{c}}{2\pi\cdot R_{DA}\cdot R_{DC}}(A+B+C), \tag{9}$$

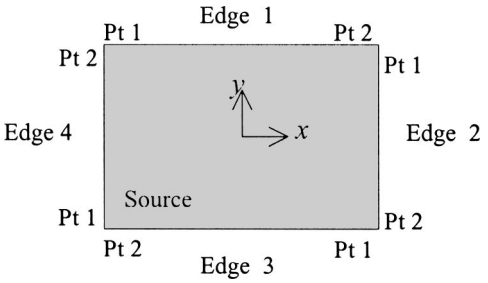


FIG. 6. Definitions of points and edges for a rectangular source.

where

$$A=C_b\cdot C_d\cdot\int_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}\sin^{-1}(u_1/r)du_1=C_b\cdot C_d\cdot[u_1\cdot\sin^{-1}(u_1/r)+\sqrt{r^2-u_1^2}]_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}, \tag{10}$$

and

$$B=C_b\cdot C_c\cdot\int_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}\sqrt{r^2-u_1^2}\cdot\sin^{-1}(u_1/r)du_1=C_b\cdot C_c\cdot\left[\frac{r^2}{4}\cdot(\sin^{-1}(u_1/r))^2+\frac{u_1}{2}\cdot\sin^{-1}(u_1/r)\cdot\sqrt{r^2-u_1^2}+\frac{r^2}{8}-\frac{u_1^2}{4}\right]_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}, \tag{11}$$

and finally

$$C=C_a\cdot\int_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}C_c\cdot\sqrt{r^2-u_1^2}+C_d\,du_1=C_a\left[C_c\cdot\left\{\frac{u_1}{2}\cdot\sqrt{r^2-u_1^2}-\frac{r^2}{2}\cdot\cos^{-1}(u_1/r)\right\}+C_d\cdot u_1\right]_{u_1^{\mathrm{lower}}}^{u_1^{\mathrm{upper}}}. \tag{12}$$

Equations (9)–(12) give solutions for the arc-limited and edge-limited integrals. The planar integral can be solved with reference to Eq. (6). Because the angle $\Omega(t)$ is constant at 2π over the whole area of the receiver, the result always has the constant value c. The three definite integrals in Eqs. (9)–(12) are well suited to a computer program, which would solve the equations in sequential time steps.

It is now possible to reconstruct the source from the arc-limited, edge-limited, and planar integrals, within each region as defined earlier in Fig. 3. This is done by finding the correct value for v_d, u_d, R_{DA}, R_{DC}, C_a, C_b, C_c, C_d, and t_p^{end} in each case, and is performed for both corners of each edge in turn. The labeling of these points is shown more clearly in Fig. 6, where it should be noted that each corner of the source has two different labels depending on the edge being considered. This allows a simplification, as when the contributions are summed, all those from Pt $1s$ are positive, and all those from Pt $2s$ are negative.

TABLE I. Values for location and size constants with reference to Figs. 5 and 6.

Edge	R_{DA}	R_{DC}	u_d	v_d Pt 1	v_d Pt 2
1	r_y	r_x	y_b	x_a	x_b
2	r_x	r_y	x_b	y_b	y_a
3	r_y	r_x	y_a	x_b	x_a
4	r_x	r_y	x_a	y_a	y_b

First, it has been stated that the angle $\Omega_{\text{edge}}(t)$ must be defined consistently for the summation of different integrals to be valid. For regions 1 to 3, the reference line used is defined as the line passing through the point on the receiver being considered in the negative x direction along a line parallel to the x axis. The angle runs from $-\pi$ to $+\pi$ measured clockwise; in these three regions the discontinuity across the line which passes from $-\pi$ to $+\pi$ is not crossed. However, this is not the case for region 4. Consequently, region 4 is subdivided into 4 zones, and the receiver further segmented and reflected such that each segment lies in zone 1. The angle used is defined as from 0 to 2π from the negative x direction, with the exception of edge 4 Pt 1, which is measured as the angle from the negative x direction plus 2π; in this way the discontinuity is not crossed.

The values for each of the variables in Eqs. (9)–(12) can now be evaluated. Table I gives the values of v_d, u_d, R_{DA}, and R_{DC}, in terms of the actual distances for the particular problem under investigation. For each region, the values for the constants, C_a, C_b, C_c, C_d can also be defined. The first two of the generalized integral constants, C_a and C_b, are used to describe the angle with respect to a reference line; these are the same for both edge-limited and planar integrals for any given edge and point. They are dictated by the rotation required to superimpose the angle $\Omega(t)$ used in the generalized solution [Fig. 14(b) in the Appendix] on the edge and point under consideration, and so vary for each point and edge. The values of C_a and C_b are given in Table II. The other two constants used in the generalized integral, namely C_c and C_d, are used to define the length of the receiver parallel to the edge which is active, v_1, as a function of time and perpendicular distance from the edge, u_1. These values are constant for all regions and edges. They are also constant for the edge-limited integral, where C_c is 0 and C_d is R_{DC}. The values for C_c and C_d only vary with the time interval (as defined in Table V of the Appendix) for the arc-limited integral. For the arc-limited integral they are given in Table III.

Finally, the time at which Ω returns to 0 has to be considered, and is termed t_p^{end}. When the edge under consideration lies parallel to the receiver, it is treated as an edge-limited integral minus a corner-limited integral for each end (Pt 1 and Pt 2) of the edge, as described in the Appendix; in this case Ω returns to 0 when the corner-limited integral is equal to the edge-limited integral, which will occur when the arc of radius $r=ct_p$ cast from the point on the edge encompasses the whole of the receiver, so t_p^{end} is t_p^{stable} for both Pt 1 and Pt 2. When the edge is not alongside the receiver, as is always the case in region 2, the edge is treated as a corner-limited integral from the nearest edge minus a corner-limited integral from the furthest point, as described in the Appendix; in this case Ω returns to 0, when the arc cast from the furthest point encompasses the whole of the receiver, so for both points t_p^{end} is the same as t_p^{stable} for the furthest point. So, t_p^{end} is equal to t_p^{stable} (meaning that time interval 3 in Table V does not occur), except for edges where there is no edge contribution when t_p^{end} for the nearest point is as given in Table IV.

The value of t_p^{end} must also be found for the planar integral. The planar integral is used when the receiver overlaps the source; that is, some part of the receiver lies within region 4. When this occurs the receiver is further segmented and reflected such that each part lies within zone 1, as defined in Fig. 5. The planar integral attributes a value of 2π across the whole area of the segment. The arc- and edge-limited integrals are then used to reduce this to the correct value of the angle subtended at each point by the arc lying within the source, across the receiver. Consider point Q in Fig. 7; the arc from this points crosses two edges, so it remains within the active area for these edges; however, the arc of the same radius centered at point P crosses no edges so lies within no active area. What this means is that for point Q the planar integral should be active, while for point P it should not. However, as the planar integral applies the value

TABLE II. Values for constants C_a and C_b with reference to Figs. 5 and 6.

	Region 1		Region 2		Region 3		Region 4	
	C_a	C_b	C_a	C_b	C_a	C_b	C_a	C_b
Edge 1 Pt 1	0	1	0	1	0	-1	2π	-1
Edge 1 Pt 2	π	-1					π	1
Edge 2 Pt 1	$\pi/2$	1	$\pi/2$	-1	$-\pi/2$	1	$3\pi/2$	-1
Edge 1 Pt 2					$\pi/2$	1	$\pi/2$	1
Edge 3 Pt 1	π	-1	0	1	0	1	π	-1
Edge 1 Pt 2	0	1					0	1
Edge 4 Pt 1	$\pi/2$	-1	$\pi/2$	-1	$\pi/2$	-1	$5\pi/2$	-1
Edge 1 Pt 2					$-\pi/2$	1	$3\pi/2$	1

TABLE III. Values of C_c and C_d for arc-limited integrals.

Time interval	C_c	C_d
1	1	$-v_d$
2(i)	0	R_{DC}
2(ii)	1	$-v_d$
3	0	R_{DC}

of 2π to the whole receiver, this is not possible. It is therefore necessary to introduce a planar-amend term; this is an arc-limited integral from edge 4 Pt 1, where the only change is that the angle is set as 2π, so $C_a=2\pi$ and $C_b=0$. With this in place the value of t_p^{end} for the planar integral is equated to that of edge 4 Pt 1.

The total value of the impulse response for each segment can be found using

$$h(\bar{r},t)=h^{\text{plane}}(\bar{r},t)-h^{\text{plane\_amend}}(\bar{r},t)-\{h_{\text{edge1 pt1}}(\bar{r},t)-h_{\text{edge1 pt2}}(\bar{r},t)+h_{\text{edge2 pt1}}(\bar{r},t)-h_{\text{edge2 pt2}}(\bar{r},t)+h_{\text{edge3 pt1}}(\bar{r},t)-h_{\text{edge3 pt2}}(\bar{r},t)+h_{\text{edge4 pt1}}(\bar{r},t)-h_{\text{edge4 pt2}}(\bar{r},t)\}, \quad (13)$$

where the superscript refers to the type of integral and each $h_{\text{edgeXptY}}(\bar{r},t)$ refers to the edge X and point Y and consists of a corner and edge limited integral, such that

$$h_{\text{edgeXptY}}(\bar{r},t)=h^{\text{edge}}_{\text{edgeXptY}}(\bar{r},t)-h^{\text{corner}}_{\text{edgeXptY}}(\bar{r},t). \quad (14)$$

When Eq. (13) and Eq. (14) are combined it can again be seen that the corner-limited integral for Pt 1 on all edges is positive, and Pt 2 negative, and that the edge-limited integral for any given point has the opposite sign as the corner-limited integral.

III. INITIAL PREDICTIONS OF THE FINITE RECEIVER MODEL

The use of a finite receiver for a particular measurement results in a change to the received signal at a given field position over that from a point detector. To illustrate this, the theoretically predicted peak-to-peak pressure obtained when finite-sized receivers of three sizes and a point receiver are scanned along the x axis, for constant values of $z=20$ and $y=0$ mm, are shown in Fig. 8. The source dimensions were 2.6 mm in the x direction and 5.2 mm in the y direction (corresponding to a source size of approximately $4\lambda\times8\lambda$). It was assumed to radiate into air, and to be driven with a 500-kHz continuous-wave signal. The three different-sized receivers have the same area, but varying aspect ratios, with dimensions of $4\times1\lambda$, $2\times2\lambda$, and $1\times4\lambda$, expressed as width (x)$\times$length (y). The amplitude received in the point receiver

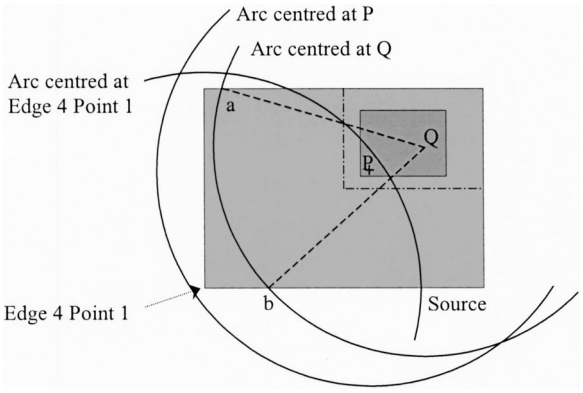


FIG. 7. The arcs generated at certain points on an overlapping source and receiver, to illustrate the need to amend the plane integral contribution.

case is normalized, and then the data for the other receivers are scaled using the same factor. It can be seen that the effect of the use of a finite receiver is to smooth sidelobes, and reduce the average pressure measure over the area. It can be seen that the degree to which this occurs is dependent on the aspect ratio. Of the three receivers used, the maximum pressure amplitude is received by the square ($2\times2\lambda$) transducer, for which the difference in the maximum and minimum time of flight from any point on the source to any point on the receiver is the smallest. The $4\times1\lambda$-sized receiver has the second largest on-axis peak-to-peak pressure, with the minimum being measured by the $1\times4\lambda$ receiver. It can be seen that the effect of the use of a finite receiver increases as the detection point moves away from the transducer axis (at $x=0$). This is due to the directivity of the receiver causing sidelobes to be smoothed at greater subtended angles to the source.

Figure 9 presents variations in the peak-to-peak pressure amplitude along the axis joining the source and receiver (the z-axis) as the source and receiver are moved apart. This is shown for various receiver sizes. The source dimensions and drive waveform used were the same as for Fig. 8, and the receiver dimensions were 0.068, 0.34, 0.68, 1.02, and 1.36 mm square, corresponding to 0.1λ, 0.5λ, 1λ, 1.5λ, and 2λ. The data are plotted in a normalized fashion, relative to the axial maximum for the smallest receiver, which has the largest received amplitude after correction for the relative areas of the receivers. It can be seen that the greatest effect of increasing the size of the receiver is within the near field, noting that the approximate near-field/far-field boundary is at a z value of 12 mm. This is due to the larger relative differences in the various possible paths from source to receiver for the larger receiver sizes. Note that, even in the far field, the maximum amplitude is lower for the larger receiver, once

TABLE IV. Exceptions for t_p^{end}.

Region	Edge	Point	t_p^{end}
2 and 3	1 3	2 1	$t_p^{\text{end}}=\sqrt{(v_d+R_{DC}+2b)^2+(u_d+R_{DA})^2}/c$
1 and 2	2 4	1 2	$t_p^{\text{end}}=\sqrt{(v_d+R_{DC}+2a)^2+(u_d+R_{DA})^2}/c$

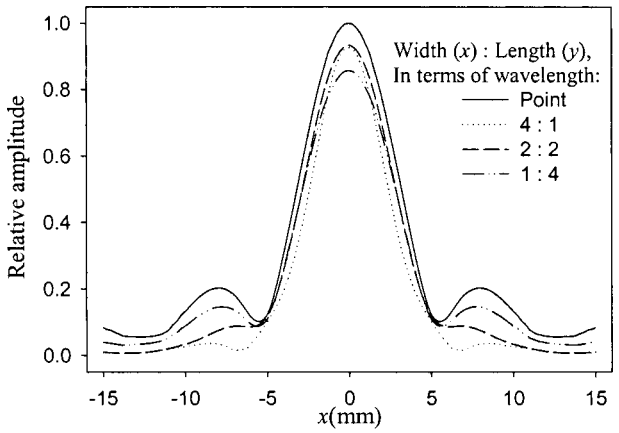


FIG. 8. Predicted variations in the peak-to-peak sound-pressure amplitude that would be detected by four different sizes of rectangular receiver. The rectangular source was of size 2.6×5.2 mm, radiating into air. The plot is along the x axis at z=20 mm and $y=0$ mm, for a 500-kHz drive signal.

the area is compensated for. It is evident that this modeling approach could be used for diffraction correction.

IV. COMPARISON OF THEORY AND EXPERIMENT

In the previous section, the effects of the receiver shape and overall surface area were discussed. This will now be studied by comparing the theoretical field plots of maximum pressure amplitude to experimental data, obtained in air.

A computer-controlled scanning stage was used to move the receiver with respect to the source, in the x, z plane, using the apparatus shown in Fig. 10. The source was driven using tone bursts from a Wavetek 191 signal generator, with an output of 30 V pk-pk. The source was a capacitive transducer, of a type that has been described elsewhere in the literature.[15,16] This is known to have a response that approximated to a plane piston when driven in air.[17] It contained a rigid micromachined silicon backplate, and a 5-μm-thick metallized Mylar membrane. Application of varying voltages caused the flexible membrane to vibrate, hence resulting in acoustic radiation. A dc bias was applied to the source via a decoupler circuit, to help attract the membrane to the backplate. The transducer had a peak response at approximately 500 kHz, and hence tone bursts at this frequency were used,

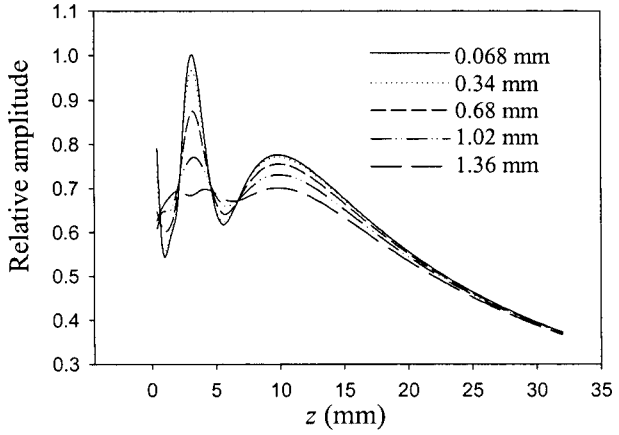


FIG. 9. Relative variation in axial peak-to-peak pressure level, for various sizes of square receivers. The source conditions are the same as in Fig. 8.

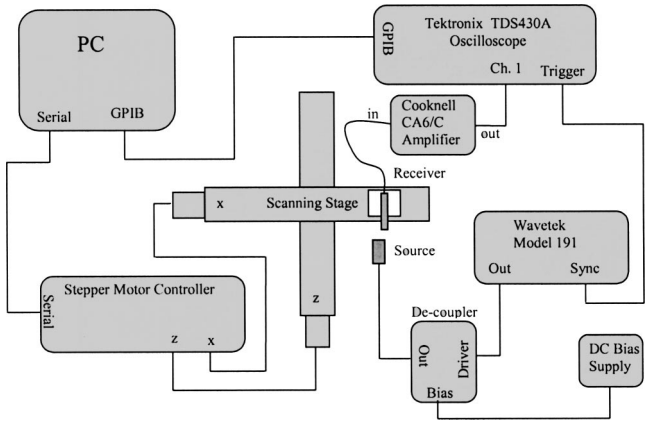


FIG. 10. Experimental apparatus for scanning the radiated fields of rectangular sources in air with a finite-size rectangular receiver.

where the wavelength in air was ~0.67 mm. The source was nominally circular, with a 10-mm diameter. However, a paper aperture was manufactured and placed over the front surface of the source, hence producing a 5-mm square radiating area. Note that paper absorbs ultrasound at this frequency very well, but is also reasonably well-matched in acoustic impedance to air, hence reducing the possibility of multiple reflections.

A 4-mm-square finite-sized receiver was constructed in a similar way to the source. The receiver was connected to a Cooknell charge amplifier, which also supplied a dc polarization voltage (in a similar way to a conventional condenser microphone). This provided a voltage time waveform output, which was digitized at each receiver location using the Tektronix TDS430A oscilloscope, before transferring the data to a PC. The stepper motor controller was also connected to the same PC. Scans could thus be performed by moving the receiver relative to the source, and recording the received time waveform at intervals of 0.5 mm in the x direction, and 1 mm in the z direction. The results could then be plotted in terms of spatial variations in the received peak-to-peak sound-pressure level amplitude.

It was thought necessary to try and use as small a receiver as possible, for comparative purposes. Hence, a fully micromachined 0.5-mm-square capacitive receiver was also used. This again has been described elsewhere,[18,19] and used silicon substrate as the backplate and a silicon nitride membrane, manufactured in a CMOS-compatible microfabrication facility. It used the same detection electronics as that described above. This was used, as the accurate manufacture of paper apertures of this small size was found to be difficult.

The scans compared the effect of the two different sizes of square receivers (i.e., receivers with the same aspect ratio of 1:1), of dimensions 0.5-mm square and 4-mm square. The normalized maximum pressure amplitude obtained experimentally for the two receivers is compared to theoretical predictions, using the finite-receiver model, for the smaller case in Figs. 11(a) and (b), respectively. Figures 11(c) and (d) make the same comparison for the larger receiver size. Figure 11(e) shows the theoretical field, from a source of the same size, driven under the same conditions, for a point receiver. The data are displayed as a gray-scale image, where a

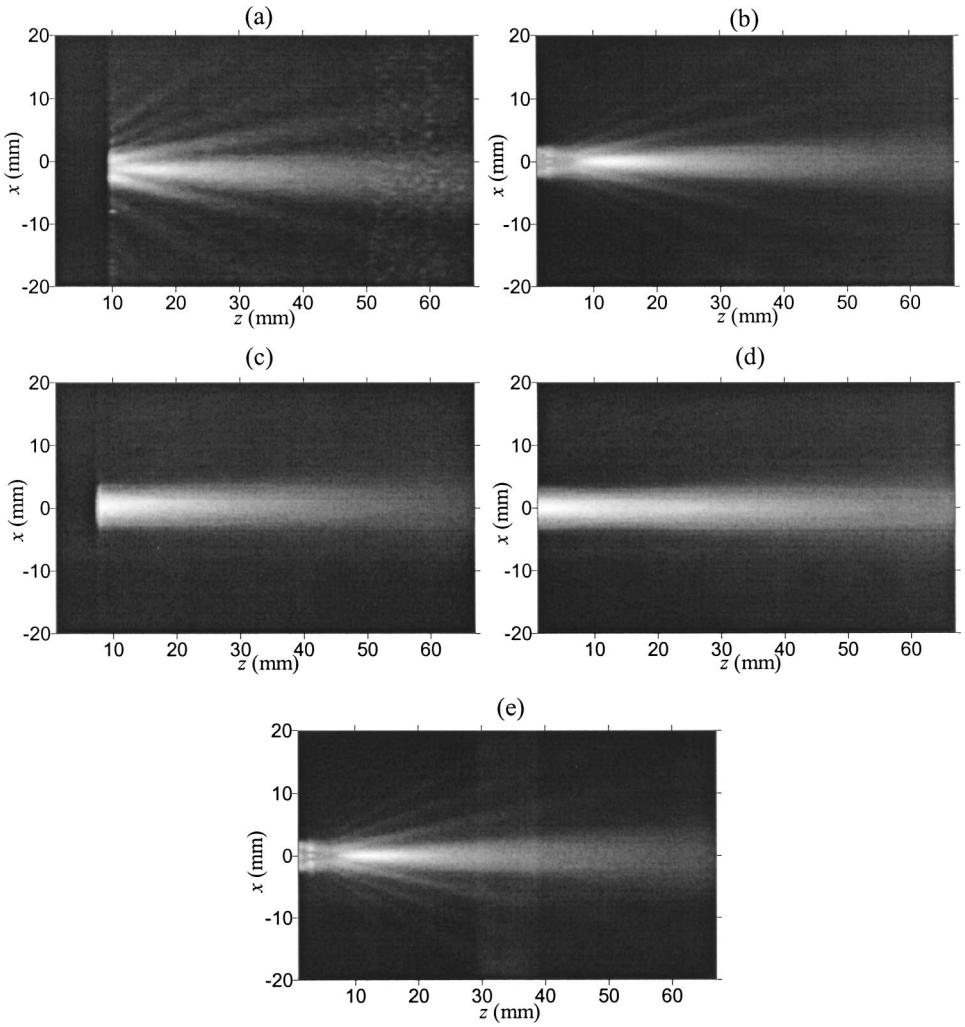


FIG. 11. (a), (c) Experimental and (b), (d) theoretical beam plots for a 0.5-mm and 4-mm square receiver in air, respectively. (e) shows the theoretical pressure field for a point receiver. The source was 5-mm square, driven by a tone burst at 500 kHz.

lighter shade indicates a higher amplitude. Note that experimental data could not be plotted within the first 10 mm of the source, due to electrical cross coupling and the finite size of the source and receiver casings. It is thus shown as a black area. It can, however, be seen that the use of a larger, more directional receiver has the effect of suppressing the sidelobes, as was also seen in the previous section. A comparison of Figs. 11(b) and (e) indicates that the 0.5-mm-square receiver has a negligible effect on the measured pressure field, showing that the receiver is now sufficiently small that the effect of the finite-receiver aperture has not affected the measured field unduly.

V. CONCLUSIONS

A finite-receiver model has been developed, based on the impulse response approach, which is capable of predicting the pressure transmitted between two rectangular transducers. This has been achieved by integrating the scalar impulse response over the face of the receiver. To obtain such an integral, an alternative method of calculation was developed, in which the receiver is treated as a series of components, and the integrals required for calculation simplified. The overall response could then be obtained by combining the results from these various geometrical elements.

Further validation of the model was presented by comparing the predicted peak-to-peak pressure amplitude field patterns to data experimentally obtained in air. These results showed a good correlation between theory and experiment, and were discussed in relation to the size and aspect ratio of the receiver and source transducers.

This model can be used to assess the effect of various-sized receivers on the measured pressure field and if required to determine the size needed such that the effect on the measured radiated field is negligible. It can also be used to account for diffraction correction, as two transducers are moved apart on-axis in an analogous way to that calculated for circular pistons.[8] It is felt that this approach is thus of interest to workers who need to scan the fields of transduc-

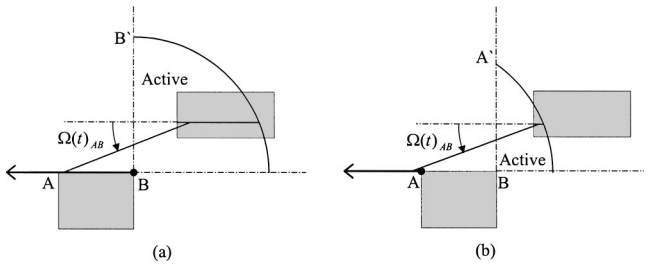


FIG. 12. Corner-limited integration split into two parts for edge AB.

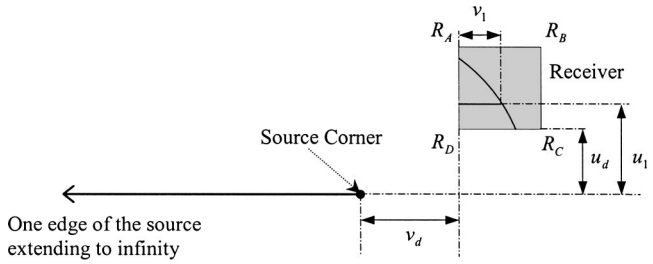


FIG. 13. Dimensions used in deriving arc-limited integral.

ers, who perform calibration work, or who need to correct received amplitudes for the effects of diffraction.

APPENDIX: CALCULATION OF INTEGRALS

1. Calculation of the arc-limited integral

It is possible to simplify the process of calculating this integral, for a receiver in region 2 (see Fig. 3), by considering the integral of the product of EG and $\Omega(t)_{AB}$, minus the integral of product of EF and $\Omega(t)_{AB}$, over the same range. First, consider the integral involving EG. To alter what is seen in Fig. 2, to an integral over the length EG, it would be necessary to remove arc A′; this can be done notionally by moving point A away from B towards infinity. In such a scenario the arc B′ would be the only boundary to the active area. Second, consider the integral involving EF. For the line EF to be in the active area, then a notional edge extending from A towards infinity away from B, would need to be considered, as this would result in the active area being limited only by arc A′. Consequently, the problem is changed to that shown in Fig. 12, the result required being the result from Fig. 12(a) minus that from Fig. 12(b). It can be seen that the problem posed by both of these cases is essentially the same. This conceptual construct of the integral of $\Omega(t)_{\text{edge}}$ over the area of the receiver limited by an arc centered at the end of a line stretching to infinity is termed the arc-limited integral, and a generalized solution will be found. The dimensions used to solve the arc-limited integral are shown in Fig. 13. These are u_1, which is the length perpendicular to the edge to the line of the receiver being considered; v_1, which is the length of the receiver at u_1 which falls within the active part of the field at any given t_p; u_d, which is the perpendicular distance from the considered edge to the receiver; and v_d, which is the length from the corner of the source (that is the tip of the infinite line) to the receiver parallel to the edge being considered. In Fig. 13, the arc shown is centered on the end of the line. In addition, it is necessary to define R_{DC} as the distance between the receiver

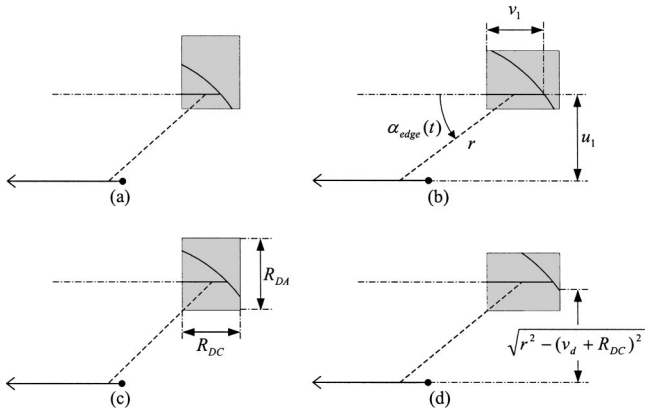


FIG. 14. Receiver at various time intervals.

corners R_D and R_C, and R_{DA} the distance between R_D and R_A. The integration will be performed over the area of the receiver within the arc of radius ct_p, centered at the tip of the infinite line, labeled source corner. In Fig. 13, the arc is shown intersecting the two nearest edges of the receiver, although in general there are four combinations of receiver edges, which could be crossed at a point in time, shown in Fig. 14. To describe these it is also necessary to define four planar times: t_p^{start}, the time at which the leading corner of the receiver becomes active; t_p^{switch}, the time at which $v_1 = R_{DC}$ at $u_1 = u_d$; t_p^{stable}, the time at which the whole of the receiver is active [given in Eq. (A1)], and finally t_p^{end}, which will be covered later when the edges are pieced together. The first three of these are given by

$$t_p^{\text{start}} = \sqrt{u_d^2 + v_d^2}/c,$$
$$t_p^{\text{switch}} = \sqrt{(v_d + R_{DC})^2 + u_d^2}/c, \tag{A1}$$
$$t_p^{\text{stable}} = \sqrt{(v_d + R_{DC})^2 + (u_d + R_{DA})^2}/c.$$

The four planar times can be used to define three time intervals in each of which the limits of u_1 can be found, and v_1 can be found as a function of u_1 within these limits. These are found with reference to r, which is defined as ct_p. They are listed in Table V. It can be seen from Table V that between t_p^{switch} and t_p^{stable}, v_1 must be expressed in two different ways: (i) and (ii). The reason for this can be seen in Fig. 14(d). For $u_1 < \sqrt{r^2 - u_1^2} - v_d$, the length v_1 is equal to the length of the receiver, while for higher values of u_1 the length of v_1 is limited by the arc so the same equation applies as that used for Figs. 14(a) and (b).

Examples of arcs across a receiver during each of the time intervals listed in Table V, with the exception of time interval 3, when the whole of the receiver is active, are given

TABLE V. Dimensions within each time period.

Time Intervals 1:3	Limits of u_1	$v_1 = \text{fns}(u_1)$
(1) $t_p^{\text{start}} \leqslant t_p \leqslant t_p^{\text{switch}}$	$u_d : \min(u_d + R_{DA}, \sqrt{r^2 - v_d^2})$	$\sqrt{r^2 - u_1^2} - v_d$
(2) $t_p^{\text{switch}} \leqslant t_p \leqslant t_p^{\text{stable}}$ (i)	$u_d : \sqrt{r^2 - (v_d + R_{DC})^2}$	R_{DC}
(ii)	$\sqrt{r^2 - (v_d + R_{DC})^2} : \min(u_d + R_{DA}, \sqrt{r^2 - v_d^2})$	$\sqrt{r^2 - u_1^2} - v_d$
(3) $t_p^{\text{stable}} \leqslant t_p \leqslant t_p^{\text{end}}$	$u_d : u_d + R_{DA}$	R_{DC}

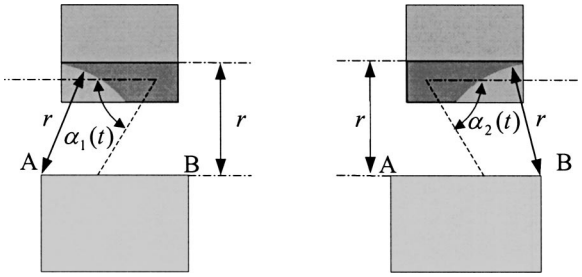


FIG. 15. Active area for an edge-limited integral.

in Figs. 14(a) to (d). Figures 14(a) and (b) are during time interval 1, and (c) and (d) occur after t_p^{switch}, and hence are in time interval 2. The upper limit of u_1 is expressed as a minimum of two expressions in Table V, depending on whether or not the arc crosses the more distant edge of the receiver parallel to the infinite line. In (a) and (c) it is the second of these two expressions and in (b) and (d) it is the first. The angle $\alpha_{\text{edge}}(t)$ as shown in Fig. 14 can be expressed as a function of u_1, using

$$\alpha_{\text{edge}}(t) = \sin^{-1}(u_1/r). \tag{A2}$$

The contribution due to a finite edge can then be found using the corner-limited integral from the furthest corner of the edge subtracted from that of the nearest corner, the only difference in the expressions being the value of v_d, t_p^{start}, t_p^{switch}, t_p^{stable}, and t_p^{end}.

2. Calculation of the edge-limited integral

As stated in the main text, and illustrated in Fig. 4, this integral involves the use of two angles [$\alpha_1(t)$ and $\alpha_2(t)$]. The angles must be integrated over the active section of the receiver. These are shown in Fig. 15, for each of the two angles, as the darker areas of the receiver. For angle $\alpha_1(t)$, the active area is limited by being within the distance r measured normally to the edge of under consideration (here AB), and by being more distant than r away from point A. In this situation then, the integral of $\alpha_1(t)$ over the active area of the receiver can be equated to the integral $\alpha_1(t)$ over the area within a distance of r from the edge of the source, minus a corner-limited integral centered on A. So, it is necessary to find the first of these, termed an edge-limited integral, which is equivalent to the case when the edge of the source is infinitely long, as then there would be no intersecting arc. It is possible to perform the integral of $\alpha_2(t)$ in a similar manner.

The edge-limited integral is defined as the integral limited by an infinite line alongside the receiver. It can be seen from Fig. 4 that the angles $\alpha_1(t)$ and $\alpha_2(t)$ are equal in magnitude. However, they must be considered separately, because when these different integrals are combined all the angles $\Omega_{\text{edge}}(t)$ must be with reference to the same line, meaning that the values of C_a and C_b relating $\Omega_{\text{edge}}(t)$ to $\alpha_{\text{edge}}(t)$ will differ [see Eq. (7)].

The variables used are given in Fig. 16, which shows a receiver alongside an infinite line. It can be seen that the length v_1 is equal to the length of the receiver in the direction parallel to the infinite edge, and so equates to R_{DC}, and the integral limits of u_1 in this illustration are u_d:r.

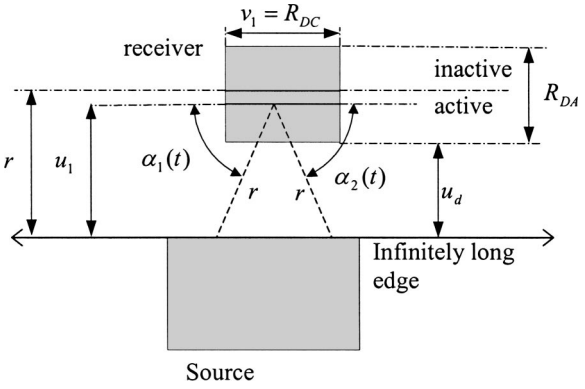


FIG. 16. Dimensions used to derive the edge-limited integral.

There is just one time interval to consider defined as being between t_p^{edge} and t_p^{end}. The value of t_p^{edge} can be found using

$$t_p^{\text{edge}} = u_d/c. \tag{A3}$$

The distance v_1 is defined in the same way as for the corner limited integral, as the length of the receiver which is active at a given u_1, between the limits of u_1,

Time interval	Limits of u_1	v_1
$t_p^{\text{edge}} \leqslant t_p \leqslant t_p^{\text{end}}$	u_d:min(u_d+R_{DA}, r)	R_{DC}

(A4)

3. Calculation of the planar integral

When the projection of the receiver onto the $z=0$ plane containing the source is such that they overlap, the planar integral must be considered. If a point is considered which falls into this region, then it will subtend an arc, $\Omega(t) = 2\pi$, until r is such that the arc crosses the nearest edge. As with the corner and edge-limited integrals, it is useful to consider an extension of the source, so that is covers an infinite plane. This would mean that during the time interval $t_p^{\text{plane}} \leqslant t_p \leqslant t_p^{\text{end}}$, the whole of the area of the receiver, $R_{DA} \times R_{DC}$, subtends an arc of $\Omega(t) = 2\pi$, where

$$t_p^{\text{plane}} = 0, \tag{A5}$$

and as previously, t_p^{end} will be covered later when these contributions are pieced together to form a rectangular source.

[1] P. A. Lewin, G. Lypacewicz, R. Bautista, and V. Devaraju, "Sensitivity of ultrasonic hydrophone probes below 1 MHz," Ultrasonics **38**, 135–139 (2000).

[2] H. L. W. Chan, S. T. Lau, K. W. Kwok, Q. Q. Zhang, Q. F. Zhou, and C. L. Choy, "Nanocomposite ultrasonic hydrophones," Sens. Actuators A **75**, 252–256 (1999).

[3] D. R. Bacon, "Primary calibration of ultrasonic hydrophones using optical interferometry," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 152–161 (1988).

[4] A. S. Khimunin, "Numerical calculation of the diffraction corrections for the precise measurements of ultrasonic absorption," Acustica **27**, 173–181 (1972).

[5] C. J. Daly and N. A. H. K. Rao, "The arcos and Lommel formulations—Approximate closed-form diffraction corrections," J. Acoust. Soc. Am. **105**, 3067–3077 (1999).

[6] P. H. Rogers and A. L. V. Buren, "An exact expression for the Lommel diffraction correction integral," J. Acoust. Soc. Am. **55**, 724–728 (1974).

[7] E. F. C. M. R. Layton, H. D. Hardy, and J. A. Bucaro, "Effects of diffrac-

tion on stress pulse propogation,'' J. Acoust. Soc. Am. **64**, 250–256 (1978).
[8] G. R. Harris, ''Review of transient field-theory for a baffled planar piston,'' J. Acoust. Soc. Am. **70**, 10–20 (1981).
[9] J. L. San Emeterio and L. G. Ullate, ''Diffraction impulse-response of rectangular transducers,'' J. Acoust. Soc. Am. **92**, 651–662 (1992).
[10] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 1.
[11] P. R. Stepanishen, ''Transient radiation from pistons in an infinite planar baffle,'' J. Acoust. Soc. Am. **49**, 1629–1638 (1971).
[12] J. C. Lockwood and J. G. Willette, ''High-speed method for computing the exact solution for the pressure variations in the nearfield of a baffled piston,'' J. Acoust. Soc. Am. **53**, 735–741 (1973).
[13] A. Freedman, ''Sound field of a rectangular piston,'' J. Acoust. Soc. Am. **32**, 197–209 (1960).
[14] A. Neild, ''Ultrasonic air-coupled capacitive arrays,'' Ph.D. thesis, University of Warwick, 2003.
[15] D. W. Schindel and D. A. Hutchins, ''Applications of micromachined capacitance transducers in air-coupled ultrasonics and nondestructive evaluation,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 51–57 (1995).
[16] D. W. Schindel, D. A. Hutchins, and W. A. Grandia, ''Capacitive and piezoelectric air-coupled transducers for resonant ultrasonic inspection,'' Ultrasonics **34**, 621–627 (1996).
[17] A. G. Bashford, D. A. Hutchins, and D. W. Schindel, ''Radiated fields of an air-coupled ultrasonic transducer,'' Ultrasonics **34**, 169–172 (1996).
[18] R. A. Noble, A. D. R. Jones, T. J. Robertson, D. A. Hutchins, and D. R. Billson, ''Novel, wide bandwidth, micromachined ultrasonic transducers,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1495–1507 (2001).
[19] J. S. McIntosh, ''Advances in Capacitive Micromachined Ultrasonic Transducers (CMUTS),'' Ph.D. thesis submitted at the University of Warwick, 2002.

Air-coupled ultrasonic sensing of grass-covered vibrating surfaces; qualitative comparisons with laser Doppler vibrometry

Andi G. Petculescu[a] and James M. Sabatier
National Center for Physical Acoustics, 1 Coliseum Drive, University, Mississippi 38677

(Received 18 April 2003; accepted for publication 5 January 2004)

The paper addresses several sensitive issues concerning the use of air-coupled ultrasound to probe small vibrations of surfaces covered with low-lying vegetation such as grass. The operation of the ultrasonic sensor is compared to that of a laser Doppler vibrometer, in various contexts. It is shown that ambient air motion affects either system, albeit differently. As air speed increases, the acoustic sensor detects a progressively richer turbulent spectrum, which reduces its sensitivity. In turn, optical sensors are prone to tremendous signal losses when probing moving vegetation, due to randomly varying speckle patterns. The work was supported by the Office of Naval Research.
 [DOI: 10.1121/1.1650329]

PACS numbers: 43.35.Yb, 43.28.Gq [YHB] Pages: 1557–1564

I. INTRODUCTION

Laser Doppler vibrometry (LDV) used in conjunction with acoustic-to-seismic coupling[1–3] has proven to be a reliable way to detect compliant and rigid buried objects.[4–7] Another technique based on the Doppler effect employs microwave Doppler radar (MDR).[8] Some applications require the use of arrays of a large number of sensors; as a result, the costs associated with implementing LDV or MDR systems become substantial. A major source of signal degradation in the LDV technique consists of a loss of useful optical power when probing grass-covered surfaces. An acoustic approach, on the other hand, is expected to be less sensitive given the relative ease of tuning the transmitter to frequencies able to penetrate grass or other scattering centers on the surface. Also, air-coupled ultrasonics is able to provide the same range of wavelengths as MDR does, at frequencies several orders of magnitude smaller. Beside costs, this precludes the limitations associated with GHz electronics and signal processing. Several works have tackled the development of air-coupled ultrasonic vibration sensors to detect targets buried in sand in the presence of seismic surface waves;[9,10] however, to the best of the authors' knowledge, the influence of ambient air motion on ultrasonic sensor operation, as well as a direct comparison with optical sensors, has not been addressed, in the context of Doppler vibrometry.

Ultrasonic Doppler sensors can be used in tandem with microwave radar, through phase coincidence. This concept explores the fact that the Doppler shifts generated by a target on the two types of sensors are coherent, despite external degrading factors that affect them differently.[11,12]

An acoustic field incident on a vibrating surface is *not* a noncontact technique *per se*; the fluid that supports wave propagation acts as the coupling medium, in contact with both the sensor and the surface. In addition to the (surface-constrained) linear kinematic transformation represented by the Doppler effect, the acoustic energy radiated by the surface may couple nonlinearly with the probe field (mixing). However, the strength of this interaction depends primarily on medium nonlinearities. The nonlinear mixing increases with the interaction volume whereas the Doppler effect is restricted to the vibrating boundary.[13] Casula and Royer[14] and Royer and Casula[15] describe the operation of a water-coupled, normal-incidence, focused-ultrasound technique for measuring surface velocity transients. They show that, for water, the parametric interaction dominates after several surface vibration wavelengths. Barriere and Royer[16] use parametric interactions to measure the nonlinearity parameter of water and ethanol; they also develop a theoretical model of the diffraction effects on parametric interactions.

The main phenomenology of practical acoustic vibration sensors has been investigated thoroughly by various authors[17–19] and has made the subject of patented systems.[20] It has also been the center of much debate. For instance, two contending sides were Censor,[21–23] and Piquette and Van Buren.[24–26] Censor tackles the problem kinematically, expressing the linearized wave equation via a Galilean coordinate transformation, within the quasistatic approximation (slow surface oscillation); this approach embodies the pure Doppler effect. In Ref. 23, Censor considers the effects of surface *and* medium motion, within the linear approximation. He argues that, in order for mass continuity to be observed, both cases must be included in the problem. Piquette and Van Buren embrace this idea for water as the propagation medium. They contend that, while laser vibrometers are based on the pure Doppler effect, since the governing (Maxwell's) equations are inherently linear, the acoustic case is inherently nonlinear due to the nature of the Navier–Stokes equation and should be treated as such. Moreover, they argue that, in liquid media, a reduction of the ultrasonic probe power causes commensurate drops in the (linear) Doppler and the nonlinear parametric effects. Due to the scope of the present work, these views will not be discussed further.

This work is motivated by the ongoing effort to develop alternatives to optical sensors for measuring vibrations of rough surfaces such as soils and sands through overlying vegetation (grass, leaves, etc.). Through a series of compara-

[a]Electronic mail: apetculescu@northwestern.edu

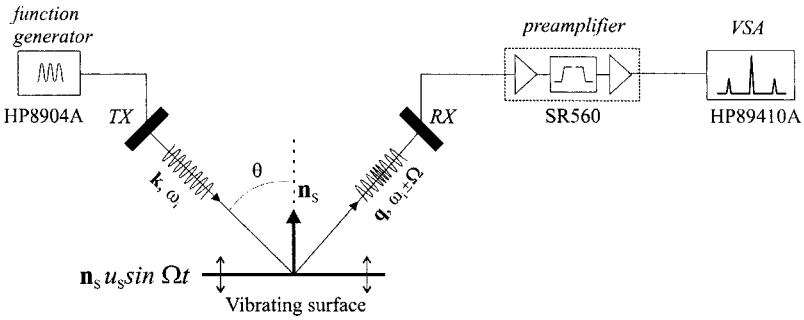


FIG. 1. Sensing geometry and experimental arrangement.

tive experiments, the authors present cases where ultrasonic sensors can be advantageous over laser vibrometers.

II. DOPPLER VIBROMETRY AND ANGLE MODULATION

In this section, the operation of an acoustic vibration sensor is described, based solely upon simple kinematic arguments. The line of thought is constructed in such a manner as to provide a connection with experimental data. It is by no means intended as a thorough study of ultrasound Doppler sensors; see, for instance, the referenced works by Censor and Piquette and Van Buren for attempts to quantify the phenomenology involved. Nevertheless, Censor's solution of the linearized wave equation (Ref. 23) yields results that reduce to those derived at the end of this section.

Consider a plane wave emitted by a transmitter in the laboratory system S, reflected by a target moving at constant velocity $\mathbf{V}$ (system S'), and detected back in S. If $\mathbf{k}$ and $\mathbf{q}$ are the incident (i) and reflected (r) wave vectors, respectively, the frequency detected by the receiver in S is given by[27]

$$\omega_r = \omega_i + (\mathbf{q}-\mathbf{k})\cdot\mathbf{V}. \tag{1}$$

The quantity $\Delta\omega_D \equiv (\mathbf{q}-\mathbf{k})\cdot\mathbf{V}$ is the Doppler shift. Equivalently, Eq. (1) states that during a small time interval dt, the phase of the incident field is changed, upon reflection, by the amount $(\mathbf{q}-\mathbf{k})\cdot\mathbf{V}\,\mathrm{d}t$.

In the case of Doppler vibration sensors, the "target" velocity is time dependent. Consider a surface vibrating sinusoidally in the normal direction $\mathbf{n}_s$ with angular frequency Ω and instantaneous velocity $\mathbf{V}(t) = \mathbf{n}_s u_s \cos\Omega t$, u_s being the velocity amplitude (Fig. 1). The instantaneous displacement of the surface is $\xi_s(t) = \mathbf{n}_s(u_s/\Omega)\sin\Omega t$.

Although the present study considers the simple bistatic geometry shown in Fig. 1, it is useful to note the general character of the displacement amplitude $\xi_S(t)$: beside the case of translation of the surface parallel to itself, it can also represent tilt or rocking motion. The change in frequency upon reflection from the oscillating boundary is

$$\omega_r(t) = \omega_i + (\mathbf{q}-\mathbf{k})\cdot\mathbf{V}(t). \tag{2}$$

The temporal phase of the reflected field is determined from the time integral of Eq. (2):

$$\Phi_r(t) = \omega_i t + (\mathbf{q}-\mathbf{k})\cdot\int_0^t \mathbf{V}(\tau)\,d\tau = \omega_i t + (\mathbf{q}-\mathbf{k})\cdot\xi_s(t). \tag{3}$$

Equations (2) and (3) embody the *instantaneous frequency* and *instantaneous phase*, within the concept of angle modulation. The propagation medium constitutes the "communication channel" while the surface vibration assumes the role of the modulator. In the case of ultrasonic sensors, assuming plane waves, the received pressure field is

$$p_r(r,t) = |p_r| e^{i\mathbf{q}\cdot\mathbf{r}} e^{-i[\omega_i t + (\mathbf{q}-\mathbf{k})\cdot\mathbf{n}_s(u_s/\Omega)\sin\Omega t]}, \tag{4}$$

where $|p_r|$ is its magnitude. The right-hand exponent term in the brackets represents a sinusoidal modulation of the signal phase. Frequency or phase demodulation of the received signal generates a voltage proportional to the instantaneous frequency $\omega_r(t)$ or phase $\Phi_r(t)$, respectively,[28] and thus retrieves the vibration information. Although the two types of angle modulation are mathematically inseparable, a fine line is drawn between them within the context of practical vibrometer design. Equations (2) and (3) state that frequency demodulation yields a direct measure of the vibration *velocity* amplitude, while phase demodulation is directly related to the *displacement* amplitude. If the medium has sizable acoustic nonlinearities, the phase in Eq. (3) is supplemented by nonlinear mixing terms.

In keeping with electronic communications nomenclature, a *phase modulation index* is introduced as

$$\beta \equiv (\mathbf{q}-\mathbf{k})\cdot\mathbf{n}_s\frac{u_s}{\Omega}. \tag{5}$$

For ultrasonic sensors, assuming $u_s \ll c_0$ (c_0 being the speed of sound in the medium), β becomes[29]

$$\beta \approx \frac{2\omega_i}{c_0}\frac{u_s}{\Omega}\cos\theta = \frac{2\omega_i}{c_0}\xi_s\cos\theta. \tag{6}$$

$\xi_s \equiv u_s/\Omega$ denotes the displacement amplitude and θ is the incidence angle, as in Fig. 1. The presence of the surface vibration angular frequency Ω in the denominator indicates an analogy to narrowband angle modulation by a single tone, corresponding to weak modulation amplitudes.[28] Using the generating function for Bessel functions as $e^{iz\sin\theta} = \sum_{n=-\infty}^{\infty} J_n(z)e^{in\theta}$, Eq. (4) becomes

$$p_r(r,t) = \sum_{n=-\infty}^{\infty} |p_r| J_n(\beta) e^{-i(\omega_i + n\Omega)t} e^{i\mathbf{q}\cdot\mathbf{r}}, \tag{7}$$

where J_n is the Bessel function of the first kind. Equation (7) is essential to investigating the operation of ultrasonic vibrometers. The energy in the received ultrasonic field is divided between the central "carrier" ($n=0$) and an infinite number of symmetric sidebands, of frequencies $\omega_i \pm n\Omega$ and amplitudes $|p_r|J_n(\beta)$. In the world of electronic communications, this is typical of angle modulation: the PM (or FM) side-

bands siphon energy out of the primary (carrier) signal. The sidebands contain all the information pertinent to the surface vibration. Taking the analogy to electronic communications further, it is conceivable to think about the possibility of increasing the modulation index up to the first zero of J_0, that is $\beta=2.405$, in which case the detected carrier amplitude is virtually zero and all the energy is transferred into the sidebands. However, for practical purposes, this corresponds to an increase of, say, the surface displacement amplitude by several orders of magnitude.

For very small modulation indices, the only surviving orders in the Bessel series are $n=0, \pm 1$; this is equivalent to the received signal having a bandwidth of approximately Ω/π. The Bessel functions satisfy $J_{-1}(\beta)=-J_1(\beta)$ and, given the small-argument behavior of Bessel functions, $J_0(\beta)\approx 1$ and $J_1(\beta)\approx\beta/2$. Thus, the vibration sidebands are out of phase, with amplitudes (called SB, for simplicity), given by

$$\mathrm{SB}\approx \pm |p_r| \frac{u_s}{\Omega} \frac{\omega_i}{c_0} \cos\theta. \tag{8}$$

Assuming weak dissipation, the magnitude of the received pressure is close to that of the incident field, $|p_r|\approx|p_i|$. Equations (6)–(8) are central to understanding the sensitivity issues of ultrasonic vibration sensors. It is clear that the sensitivity can be increased if (i) the vibration displacement (or velocity) amplitude is increased, (ii) the probe frequency is larger, or (iii) the probe strength is increased. A local reduction of the sound speed will also lead to increased detection sensitivity; this could be attained, e.g., by enclosing the sensor head into a ''bubble'' containing a ''slow'' medium, in contact with the probed surface. However, such an approach may give rise to problems associated to standing waves, losses, etc., which have to be investigated further.

Any relative motion between the sensor head and the surface generates additional Doppler peaks that may lie inside the measurement bandwidth. In addition, sensor oscillations, whether coherent with the surface vibrations or due to nearby equipment, will affect operation. Piquette and Van Buren[25] address this scenario in the context of underwater vibration sensing. Proper vibration insulation along with a judicious choice of scanning speed will help alleviate these factors. An interesting technique for factoring out any additional target-sensor relative motion developed for laser vibrometry is explained in detail in Ref. 30.

III. SIGNAL DEGRADATION

The spectral purity of the measurement signals is perhaps the most basic requirement of sensor development. In many cases, quadrature Doppler systems suffer from imbalances in amplitude and phase.[31] These may result in signals larger than the measured vibration levels; if they are close to the vibration frequencies, they can introduce sizable errors in the demodulation process.

In addition to electronic noise incurred in detection, a major source of signal degradation for an air-coupled acoustic sensor is vortical or turbulent air motion (e.g., vortex shedding induced by sensor motion or wind). By analogy with electronic communications, it can be said that the role of wind is that of ''channel noise'' (albeit not ''white,'' but intermittent). As the speed of air motion increases, the sensor detects a progressively richer (turbulent) spectrum that can ultimately overwhelm the information-bearing sidebands. Many authors have developed models aimed at quantifying acoustic scattering in turbulent media.[32,33] The farfield expression for the scattered acoustic pressure per unit volume derived in Ref. 33 is easily amenable to experiments. The scattering mechanism can be described qualitatively as follows: eddies with specific length scales, excited by the incident field of frequency $\omega_i/2\pi$ on time scales smaller than the typical turbulent time scales, become acoustic sources radiating at different frequencies $\omega/2\pi$. Contreras and Lund[34] arrived at an equivalent expression for scattering by thermal turbulence. Various authors[35–37] have done intensive experimental studies of ultrasound scattering by turbulence.

Additionally, wind blowing over grass-covered surfaces induces a complex multicomponent motion of the grass blades, which is detected by the vibrometer. Common laser vibrometers relying on backscattered light tend to be susceptible to this phenomenon: in-plane motion of a grass blade randomly alters the spatial characteristics of the speckle pattern inherent in coherent scattering from surface irregularities on the order of the optical wavelength. This, in turn, causes significant drops in the demodulated signal.[38,39] In the ultrasonic case, the noticeably larger ''acoustic speckle'' precludes any such tendencies. Fiber-coupled laser Doppler sensors are also affected by fiber vibrations induced by wind or other sources. A very clear and systematic theoretical/experimental investigation of environmental effects on fiber-coupled systems is given in Ref. 40.

In the absence of wind, signal degradation also appears as a result of scattering of the incident field from surface inhomogeneities (grains, pebbles, etc.) and rooted or loose vegetation (grass, fallen leaves, pine needles, moss, etc.) with characteristic length scales. Conceptually, this type of scattering is expected to impose lesser constraints on acoustic sensors due to the relative ease of transducer tunability. However, decreasing the probe frequencies in order to ''see through'' various scattering agents comes at the price of reduced Doppler sensitivity [Eqs. (6) and (8)].

IV. EXPERIMENT, RESULTS, AND DISCUSSIONS

The experimental arrangement is shown in Fig. 1. Two similar custom-made piezoelectric transducers (center frequency≈120 kHz, bandwidth≈10 kHz, total divergence ≈9°) acting as transmitter (TX) and receiver (RX) form the sensor. The angle between each transducer axis and the normal is 13°; the average sensor elevation was 14 cm, resulting in a total pathlength (transmitter to receiver) of 28 cm. The modulated carrier is filtered and amplified by a low-noise preamplifier (SR560) and subsequently fed to a HP89410A vector signal analyzer (VSA), capable of real-time demodulation. An HP8904A signal generator supplies the transducer excitation for most measurements. Commercial shakers are used to excite the vibration of the probed surface. Two types of shakers are used: a clamped-plate type (AURA model AST-2B-4, $f_{\mathrm{res}}=40$ Hz) and a piston type (LDS model V203,

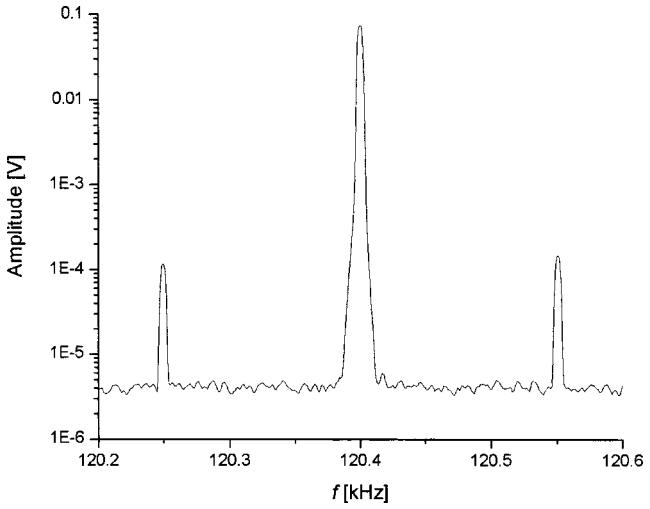


FIG. 2. Typical spectrum of ultrasonic Doppler signal, showing the carrier and the first-order sidebands.

broadband). The surface displacement amplitude is monitored with a Polytec PDV 100 laser Doppler vibrometer (LDV). In the next sections, the acronym UDV will be used for ultrasonic Doppler vibrometry. V_{TX} is the voltage applied to the ultrasonic transmitter, proportional to the incident pressure magnitude. This choice is warranted by the fact that absolute, quantitative measurements are not the primary goal of the present study. The vibration levels are low enough so that the linear regime holds and only the first-order sidebands are observed.

A. Energy balance

A typical spectrum of the Doppler signal received from an Al plate attached rigidly to an LDS shaker is shown in Fig. 2. The received acoustic energy is split between the carrier and the sidebands. Increasing the vibration level leads to an increase of the sideband amplitude accompanied by a mild decrease in the carrier amplitude. The importance of this interplay is didactic in that it embodies a core principle of Doppler vibrometry. Moreover, ultrasonic vibrometers enable one to look directly at *both* the carrier and the sidebands; in the optical case, one must down-convert the carrier frequency by many orders of magnitude before even attempting to monitor it. In the present configuration, for the relatively low transmitter excitation used (<1 W), the minimum displacement detected was approximately 200 nm without averaging and 25 nm with averaging; these figures can be improved if larger drive levels are used. Lower excitation levels were preferred in order to preclude sensor saturation effects.

B. Measurements on a sand surface

The aim of this investigation is to assess the ability of an ultrasonic sensor to follow the output of a commercial laser Doppler vibrometer, in a more realistic scenario. The AURA shaker, excited at 150 Hz, was buried in sand; the probe frequency was 120.4 kHz. The vibrations of the surface above the shaker were probed in parallel with the UDV and the LDV (the latter was aimed at an incidence angle $\theta_{\mathrm{LDV}} = 17.5°$. The results are shown in Fig. 3. The dataset in case (a) corresponds to the shaker surface being flush with that of the sand. Cases (b) and (c) were taken with the shaker buried approximately 2.5 cm beneath the surface, one day apart from each other. The upper plots show the sideband amplitude together with the pointwise surface displacement measured with the LDV, for increasing shaker excitation, V_S. The energy dynamics between the carrier and the sidebands is evident in the lower graphs showing the carrier behavior. The vibration information obtained ultrasonically has a structure similar to that obtained with the LDV. As an aside, it is interesting to note the behavior of the sand surface during the process. The UDV/LDV vibration response is not linear; nor should one expect it to be (e.g., complex sand response to low-frequency excitation, inertial effects on shaker operation due to a nonrigid base, etc.). Moreover, the sideband and carrier amplitudes have an interesting interplay in all three cases. Case (c) is unarguably the most intriguing: the onset of the ''vibration saturation'' regime at $V_S \approx 2.5$ V (apparent in both UDV and LDV data) may indicate a particular dynamic state of the grains between the shaker and the surface. In all three cases, it is possible that a correlation exists between the carrier curve and the local slope of the sideband curve. These issues remain open to further investigation.

C. The influence of air motion and surface vegetation

Surface scattering centers as well as wind affect both laser and ultrasonic Doppler vibrometers. The scattering strength depends primarily upon the wavelength-to-obstacle-dimension ratio. This ratio is more easily controllable for an UDV than it is for a LDV. Simple experiments have been performed in order to assess the influence of various scenarios involving surface scatterers and ambient air motion. One case involves scattering from a bed of randomly oriented pine needles strewn on top of the shaker plate ($f_{\mathrm{probe}} = 120.4$ kHz, $f_{\mathrm{surface}} = 150$ Hz). The average width of the pine needles is 1.5 mm, which represents an approximate fraction of 0.52 of the ultrasonic probe wavelength. Figure 4 shows waterfall displays of real-time LDV and UDV demodulated spectra. $\Phi_{\mathrm{UDV}}(\omega)$ is the FFT of the demodulated phase of the ultrasonic signal. The time index entries represent times separated by about 0.33 s. The UDV signal keeps a fairly constant signal-to-noise ratio (SNR) over significantly longer time spans than does the LDV. The plots in Fig. 5 show time averages of such data.

In order to study the combined effects of wind and surface vegetation, a ''rooted-grass simulator'' was built from 4 mm wide paper strips with an average length of 30 mm inserted in a closed-cell foam substrate. The average ''grass'' density was approximately 1.77 blades/cm^2. The ''grass simulator'' was placed on the shaker plate. To be able to get a definite assessment, the shaker displacement was deliberately set to a relatively high value (approx. 550 nm). A relatively weak jet of nitrogen was used to set the ''grass'' blades in oscillatory motion, independent of the surface vibration. The jet was positioned as close to the foam surface as possible. The LDV beam was focused on a moving blade that was also inside the ultrasonic spot size. Figure 6 shows explicitly an instance where the LDV signal is completely over-

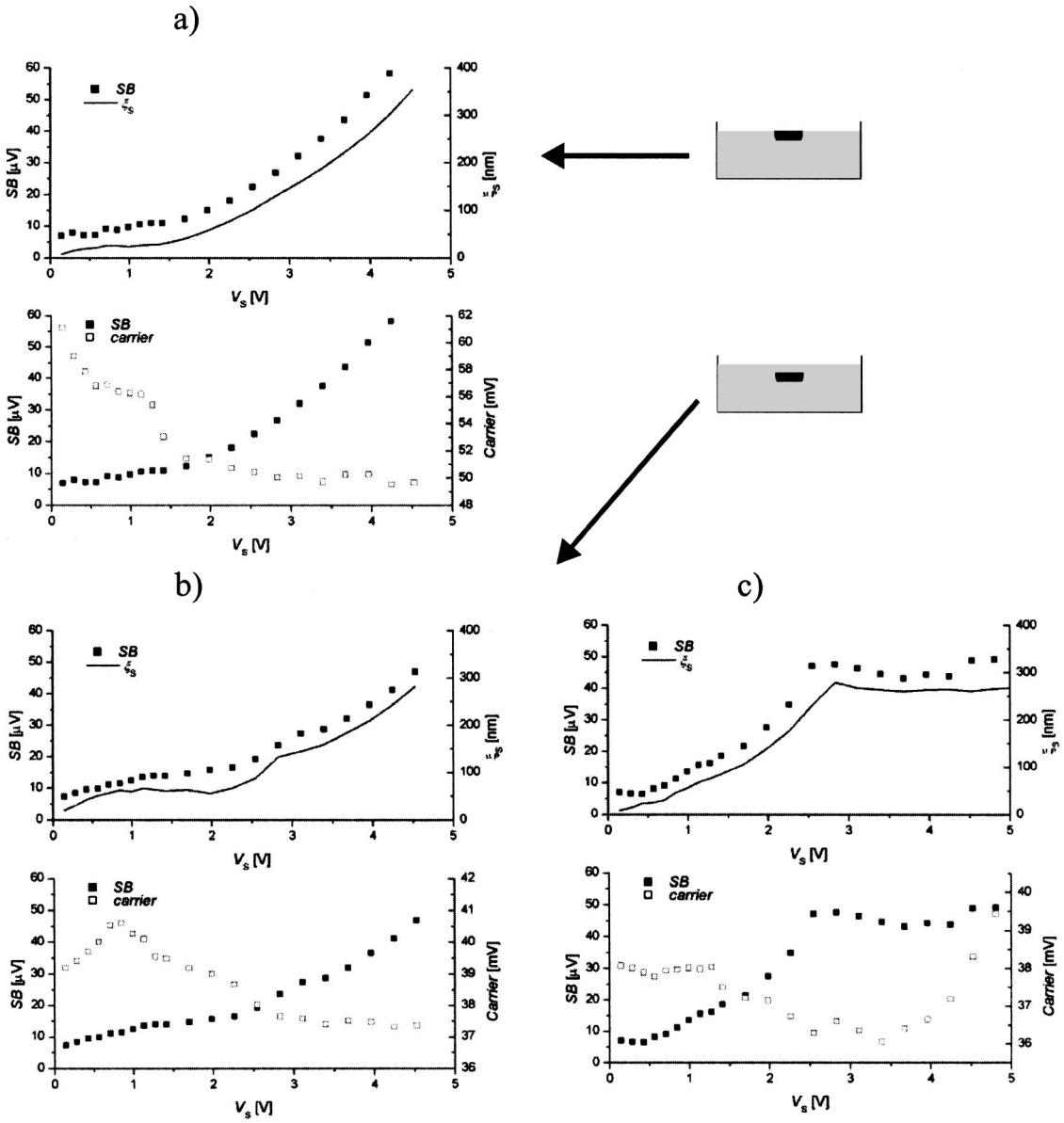


FIG. 3. Measurements in sand. (a) Shaker flush with sand surface. (b) and (c) Shaker buried approx. 2.5 cm deep. Upper plots: ■=sideband (UDV); —=surface displacement (LDV). Lower plots: □=carrier (UDV).

whelmed, while the UDV still produces a useful output. A major component of the unsteady motion of the ''grass'' blades is indicated by the closest satellite peak structure around the carrier. The onset of ''grass'' motion is discussed in more detail next.

A series of measurements was performed with the UDV in order to separate out the effects of the ''wind structure'' itself and that of the grass motion. Three datasets were taken with the jet entering the sensing region at three different heights. The heights were chosen such that the air motion interacts gradually with the grass. Figure 7 shows the UDV signal as the jet is lowered from a condition of virtually no grass motion (a) to relatively strong grass motion (c); the 150 Hz vibration sidebands are clearly seen. In case (a), only the jet ''structure'' (e.g., eddies) scatters the acoustic field. In (b), the jet barely ''touches'' the tips of the ''grass'' blades, setting some of them in unsteady motion (as observed by visual inspection). In this case, the overall shape of the turbulent Doppler ''hump'' is similar to that from case (a), with the exception of the incipient structure at approximately 17 Hz to the right and left of the carrier frequency. This can be explained by the fact that the vorticity field has a greater effect than the blade motion. The satellite structure develops as the jet is lowered further, to case (c), when it hits the blades ''head-on'' and sets a large number of them in motion. This time, the spectrum differs more visibly from that of (a); the pronounced peaks around the carrier indicate the onset of jet-induced grass vibrations. A similar structure appears in the lower plot of Fig. 6(a).

Note: Choosing jets in order to deconstruct the effects of wind may seem a naive choice; yet it is not so far-fetched: although the structure of the wind close to the ground is the subject of continuous debate, it is believed to have a Kolmogorov spectrum, characterized by eddies of many length scales.[41] The turbulent scattering mechanism depends on the relationship between the incident wavelength and eddy size.

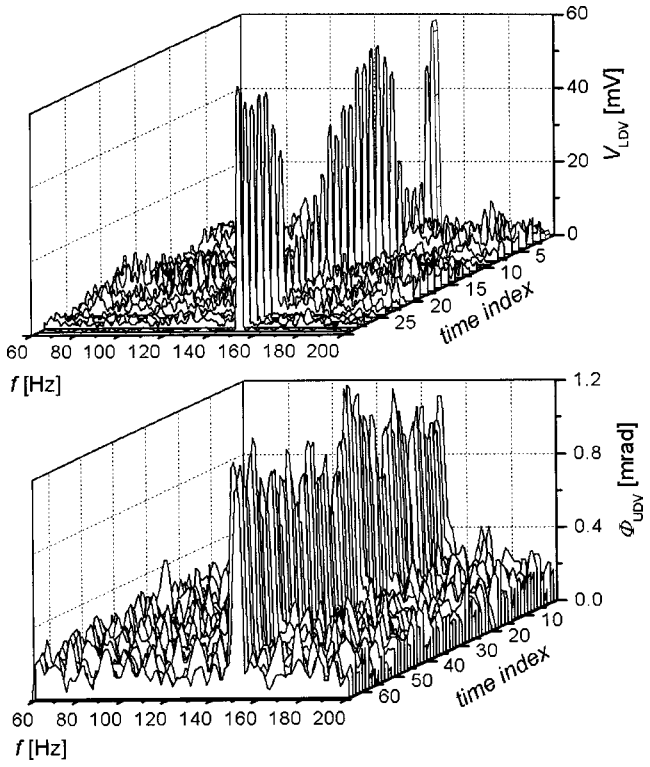


FIG. 4. Real-time waterfall scans of LDV (upper) and UDV (lower) spectra ($f_{surface} = 150$ Hz). The LDV signal loses strength considerably many times while the UDV peak is relatively self-consistent (note the almost double time span of the UDV record).

The very fact that the jet influences the ultrasonic field as seen in Fig. 7 indicates that there exist eddy scales commensurate with the probe wavelength.

V. CONCLUSIONS

A number of topics pertinent to the use of air-coupled ultrasound to probe ground vibrations have been discussed. It has been shown that the operation of a UDV system can be adequately described by the frequency or phase modulation:

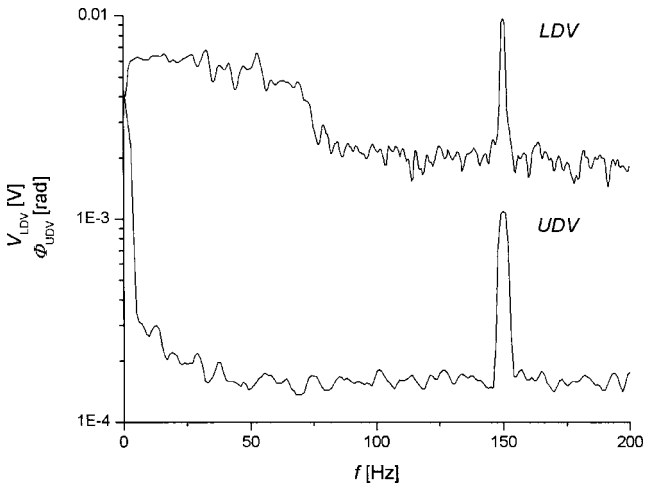


FIG. 5. Averaged LDV and UDV responses from a bed of pine needles. For SNR comparison, the data are shown on the same scale (although the units differ). The SNR of the UDV is comparable with or slightly better than that of the LDV.

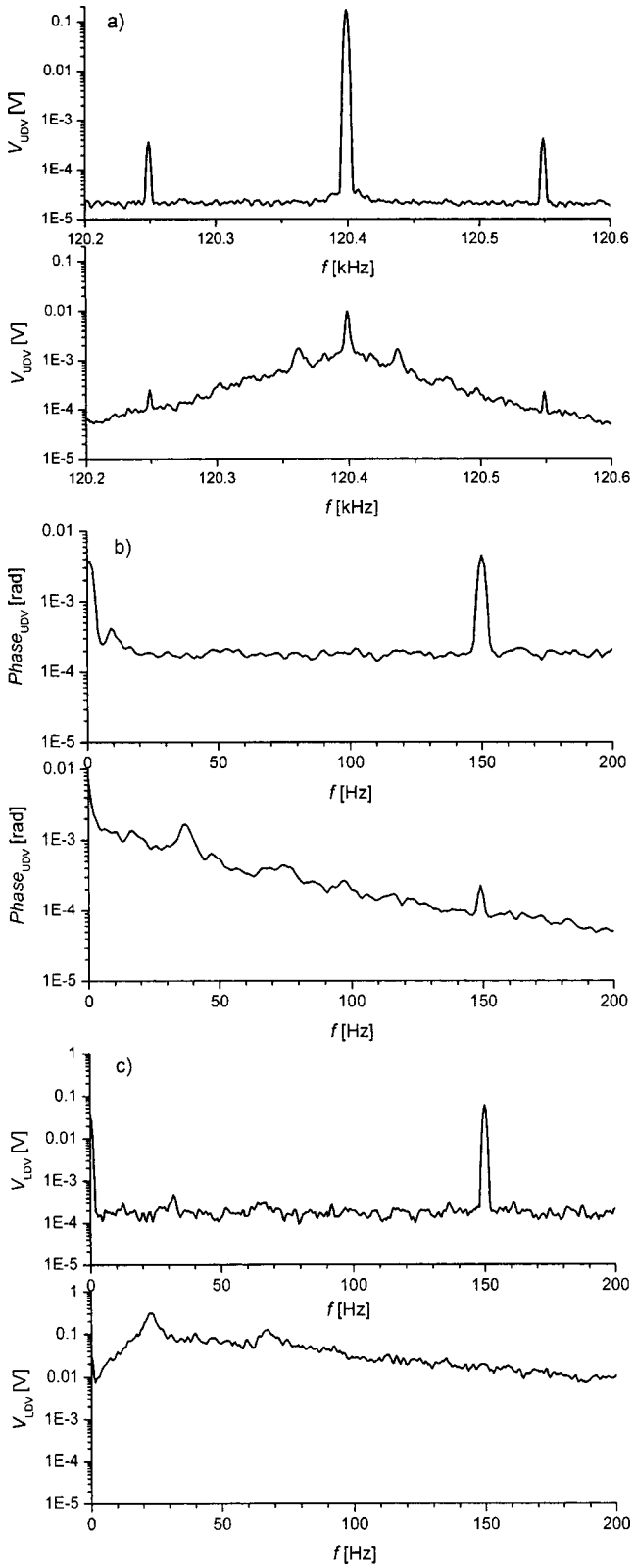


FIG. 6. Wind effects on a grass-covered surface. (a) Received UDV signal; (b) demodulated UDV signal; (c) LDV signal. Upper plots—without wind; lower plots—with wind.

part of the energy of the incident field is transferred from the carrier to the information-bearing sidebands. On a nonrigid, granular surface (sand), the performance of the UDV prototype is comparable to that of a commercial LDV. Moving one step further, to the scenario of vegetation-covered sur-

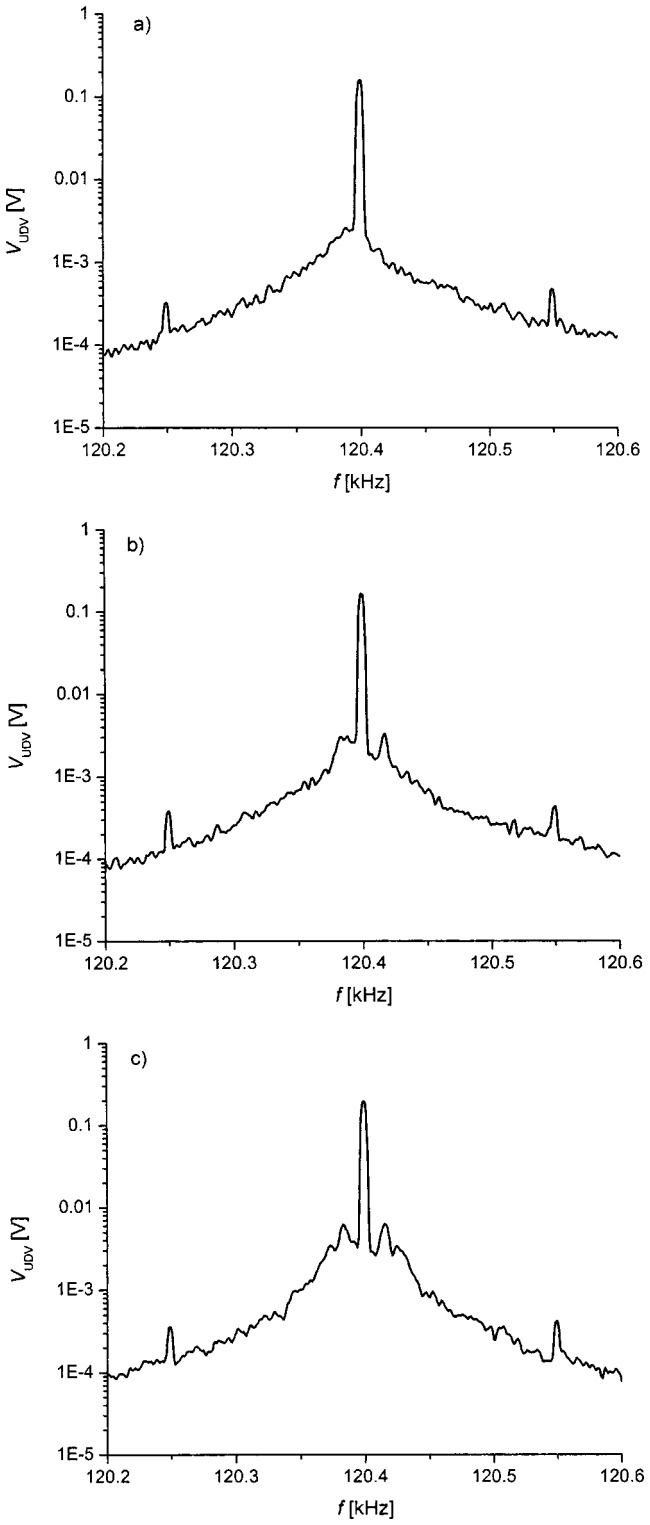


FIG. 7. Onset of grass motion as a weak jet (between sensor and surface) is lowered from location (a)—no grass motion; to (b)—weak grass motion; and (c)—strong grass motion.

faces, it has been shown that a moving pine needle or a grass blade may be the Achilles' heel of optical sensors due to speckle-induced signal loss. Of course, a UDV is expected to experience similar difficulties when the scatterers become greater than the wavelength. An investigation of ''grass blade dynamics'' may prove useful in understanding these effects. For instance, studies of (i) flexural (and other) modes excited in the blade as a result of periodic forcing applied at its root and (ii) complex vibrational modes excited by unsteady air motion, will be useful.

Another intriguing aspect is the behavior of the ultrasonic Doppler sensor in the presence of wind. However detrimental wind can be in the detection of small vibrations, it nevertheless brings a whole new dimension into the problem: that of scattering by turbulence. The better one understands its mechanism, the more likely one is to develop reliable techniques for wind alleviation in air-coupled acoustic vibrometry. Passive wind noise reduction may be as simple as using a shield around the sensing volume. An alternative option consists of the ''homogenization'' of the air turbulence by making the shield walls out of reticulated materials; the choice of the ''cell'' size is dictated, for example, by the requirement that the average eddy size be smaller than the ultrasonic wavelength, to inhibit scattering. Active techniques may involve ways to manipulate the received wind spectrum. If an absorption line at the vibration frequency could be burned into the turbulent spectrum detected by the UDV, then the SNR could be improved locally, at the vibration sidebands.

ACKNOWLEDGMENTS

The authors extend their gratitude to Professor Christoffe Baudet, Roger Waxler, Barton Smith, and Ning Xiang, for fruitful discussions. The work is supported by a grant from the Office of Naval Research.

[1] J. M. Sabatier, H. E. Bass, L. E. Bolen, and K. Attenborough, ''Acoustically induced seismic waves,'' J. Acoust. Soc. Am. **80**, 646–649 (1986).
[2] J. M. Sabatier, H. E. Bass, and G. R. Elliott, ''On the location of frequencies of maximum acoustic-to-seismic coupling,'' J. Acoust. Soc. Am. **80**, 1200–1202 (1986).
[3] K. Attenborough, H. E. Bass, and J. M. Sabatier, ''Elastic properties of the Earth's surface,'' Chapter 15 in *Handbook of Elastic Properties of Solids, Liquids, and Gases, Vol. III: Elastic Properties of Solids: Biological and Organic Materials, Earth and Marine Sciences*, edited by Levy, Bass, and Stern (Academic, New York, 2001).
[4] W. Patrick Arnott and J. M. Sabatier, ''Laser-Doppler vibrometer measurements of acoustic to seismic coupling,'' Appl. Acoust. **30**, 279–291 (1990).
[5] N. Xiang and J. Sabatier, ''Pre-differentiated M-sequences for laser-Doppler-based identification of vibrational systems,'' J. Acoust. Soc. Am. **106**, 2215(A) (1999).
[6] J. M. Sabatier and N. Xiang, ''Laser-Doppler-based acoustic-to-seismic detection of buried mines,'' Proc. SPIE **3710**, 215–222 (1999).
[7] C. J. Hickey, J. M. Sabatier, and T. M. McGee, ''A method for the detection of shallow buried objects,'' Ann. Geofis. **43**, 1225–1233 (19xx).
[8] W. R. Scott, Jr., C. Schroeder, and J. S. Martin, ''An acousto-electromagnetic sensor for locating land mines,'' *Detection and Remediation Technologies for Mines and Minelike Targets III*, Proc. SPIE **3392**, April 1998, pp. 176–186.
[9] J. S. Martin *et al.*, ''Ultrasonic displacement sensor for the seismic detection of buried landmines,'' *Proc. SPIE: 2002 Annual International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, FL, April 2002.
[10] C. T. Schroeder and W. R. Scott, Jr., ''On the complex conjugate roots of the Rayleigh equation: The leaky surface wave,'' J. Acoust. Soc. Am. **110**, 2867–2877 (2001).
[11] H. Russer and V. Magori, ''Phase coincidence between ultrasound and microwaves—a powerful and flexible multisensor principle,'' *Eurosensors XIV*, 28–30 August, 2000, Copenhagen, Denmark.
[12] H. Russer and V. Magori, ''Sweep linearization of a microwave FMCW Doppler sensor by an ultrasonic reference,'' *IEEE International Frequency Control Symposium*, Orlando, Fl, 1997, pp. 201–206.

[13] K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge, University Press Cambridge, 1998), Chap. 5.
[14] O. Casula and D. Royer, ''Transient surface velocity measurements in a liquid by an active ultrasonic probe,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 760–767 (1998).
[15] D. Royer and O. Casula, ''A sensitive ultrasonic method for measuring transient motions of a surface,'' Appl. Phys. Lett. **67**, 1706–1715 (2001).
[16] C. Barriere and D. Royer, ''Diffraction effects in the parametric interaction of acoustic waves: application to measurements of the nonlinearity parameter B/A in liquids,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 760–767 (1998).
[17] M. Cox and P. H. Rogers, ''Ultrasonic measurements of the vibrational amplitudes of auditory organs in fish,'' Adv. Test Measurement **23**, 337–350 (1986).
[18] S.-R. Huang, R. M. Lerner, and K. J. Parker, ''On estimating the amplitude of harmonic vibration from the Doppler spectrum of reflected signals,'' J. Acoust. Soc. Am. **88**, 2702–2712 (1990).
[19] S.-R. Huang, R. M. Lerner, and K. J. Parker, ''Time domain Doppler estimators of the amplitude of vibrating targets,'' J. Acoust. Soc. Am. **91**, 965–974 (1992).
[20] P. H. Rogers and M. Cox, ''Noninvasive vibration measurement system,'' U.S. Patent No. 4,819,649, 11 April, 1989.
[21] D. Censor, ''Harmonic and transient scattering from time-varying obstacles,'' J. Acoust. Soc. Am. **76**, 1527–1534 (1984).
[22] D. Censor, ''Scattering by time-varying obstacles,'' J. Sound Vib. **25**, 101–110 (1972).
[23] D. Censor, ''Acoustical Doppler effect analysis—Is it a valid method,'' J. Acoust. Soc. Am. **83**, 1223–1230 (1988).
[24] J. C. Piquette and A. L. Van Buren, ''Comments on 'Harmonic and transient scattering from time-varying obstacles','' J. Acoust. Soc. Am. **76**, 1527–1534 (1984).
[25] J. C. Piquette and A. L. Van Buren, ''Some further remarks regarding scattering of an acoustic wave by a vibrating surface,'' J. Acoust. Soc. Am. **80**, 1533–1536 (1986).
[26] J. C. Piquette and A. L. Van Buren, ''Nonlinear scattering of acoustic waves by vibrating surfaces,'' J. Acoust. Soc. Am. **76**, 880–889 (1984).
[27] J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975).
[28] B. P. Lathi, *Modern Digital and Analog Communication Systems* (Oxford University Press, New York, 1998), Chap. 5.
[29] A. D. Pierce, ''Acoustics,'' *An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), p. 456.
[30] A. Hocknell, R. Jones, and S. J. Rothberg, ''Remote vibration measurements: compensation of waveform distortion due to whole body translations,'' J. Sound Vib. **214**, 285–307 (1998).
[31] R. Moraes and D. H. Evans, ''Compensation for phase and amplitude imbalance in quadrature Doppler signals,'' Ultrasound Med. Biol. **22**, 129–137 (1996).
[32] R. H. Kraichnan, ''The scattering of sound in a turbulent medium,'' J. Acoust. Soc. Am. **25**, 1096–1104 (1953).
[33] F. Lund and C. Rojas, ''Ultrasound as probe of turbulence,'' Physica D **37**, 508–514 (1998).
[34] H. Contreras and F. Lund, ''Ultrasound as a probe for turbulence II, Temperature inhomogeneities,'' Phys. Lett. A **149**, 127–130 (1990).
[35] N. Schaeffer, ''Diffusion d'ultrasons: Mesures dans un jet d'air turbulent,'' M.Sc. thesis, École Normale Supérieure de Lyon, Lyon, France, 2001.
[36] C. Baudet, S. Ciliberto, and J. F. Pinton, ''Spectral analysis of the von Karman flow using ultrasound scattering,'' Phys. Rev. Lett. **67**, 193–195 (1991).
[37] C. Baudet, O. Michel, and W. J. Williams, ''Detection of coherent vorticity structures using time-scale resolved acoustic spectroscopy,'' Physica D **128**, 1–17 (1999).
[38] C. B. Scruby and L. E. Drain, *Laser Ultrasonics. Techniques and Applications* (Hilger, Bristol, 1990), pp. 89–97.
[39] S. J. Rothberg, J. R. Baker, and N. A. Halliwell, ''Laser vibrometry: pseudo-vibrations,'' J. Sound Vib. **135**, 516–522 (1989).
[40] C. N. Pannell, J. D. C. Jones, and D. A. Jackson, ''The effect of environmental acoustic noise on optical fibre based velocity and vibration sensor systems,'' Meas. Sci. Technol. **5**, 412–417 (1994).
[41] R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer Academic, Dordrecht, 2001), pp. 390–393.

Three-dimensional ray tracing of laser ultrasound for weld penetration sensing

Bao Mi and I. Charles Ume[a)]
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 2 September 2003; revised 29 December 2003; accepted 2 January 2004)

The time-of-flight (TOF) method is an ultrasonic nondestructive testing (NDT) technique. The TOF of an ultrasonic wave can be correlated to weld penetration depth, and hence weld quality. Changes in material properties due to temperature gradients will cause ultrasonic speed to vary during welding, which causes a curved propagation path. A ray tracing algorithm is required in order to study how ultrasound propagates within a weld sample. In this paper, a three-dimensional (3-D) ray tracing algorithm based on Fermat's principle is presented. First, ray equations are derived using the calculus of variation. Then, a numerical algorithm is developed to solve the derived ray equations and obtain the curved propagation path. This algorithm includes finite element analysis (FEA) to obtain the transient temperature distribution during the welding and shooting method to solve the boundary value problem. After the curved ray path is obtained, the TOF can be found by integrating the time variable along the ray path. An analytical relationship between the TOF and penetration depth can be established by repeating the ray tracing algorithm for different penetration depths. Experimental measurements of TOF have been performed, and this data is to be used to validate the numerical results. [DOI: 10.1121/1.1649942]

PACS numbers: 43.35.Zc, 43.20.El, 43.20.Dk [YHB] Pages: 1565–1571

I. INTRODUCTION

The laser ultrasonic technique is a noncontact means of generating ultrasound and it serves as one of the few candidates for real-time process control, where contact transducers are not applicable.[1] Such applications include the sensing and control of welding and solidification,[2,3] paper inspection,[4,5] solder joint inspection,[6] etc. The TOF method is an ultrasonic NDT technique that finds its applications in ultrasound speed measurement,[7] liquid–solid interface reconstruction,[8] and monitoring weld quality.[9] When used for weld quality control, this technique relies on diffracted signals at the flaw tip or weld root. The TOF measurement corresponding to the signal diffracted from the weld root can be correlated to weld penetration depth, and hence weld quality.[10] Figure 1 shows how this technique can be used for weld penetration depth monitoring. Ultrasound is generated by a laser phased array source on one side of the weld sample, diffracted at the weld root [shown in Fig. 1(b)], and picked up by a receiver on the other side. When a shear wave is used, the first received ultrasonic signal corresponds to the diffracted wave from the weld root because a shear wave does not propagate through liquid.

The TOF is determined by the ultrasound propagation path (which is related to the penetration depth) as in Eq. (1). During real-time welding, changes in material properties due to a temperature gradient cause ultrasonic speed to vary. The propagation path of ultrasound will be curved. A ray tracing algorithm is needed in order to study how ultrasound propagates within the specimen:

$$\mathrm{TOF}=\int_{s_1}\frac{ds_1}{\nu(x,y,z)}+\int_{s_2}\frac{ds_2}{\nu(x,y,z)}. \tag{1}$$

The ray tracing problem has been studied in geophysics for seismic modeling.[11,12] Ray tracing algorithms developed so far in geophysics cannot be directly applied to the welding case. Ray tracing analysis has been used to analyze 2-D wave propagation in the diametral plane of model cylindrical solid–liquid interfaces.[8,13] 2-D ray tracing has been studied and applied to ultrasonic wave propagation in temperature gradients based on Snell's law.[14] In this paper, a 3-D ray tracing algorithm based on Fermat's principle is presented. First of all, the temperature field within the weld sample is obtained by FEA. Then, a set of ray equations is derived using the calculus of variations. Unlike Snell's law, the derivation directly from Fermat's principle does not require the assumption of discrete plane layers and works for both 2-D and 3-D cases. The Runge–Kutta method is applied to solve the initial value ray equations and the shooting method is applied to solve the boundary value problem in order to obtain the ray path specified by the source and destination. After the curved ray path is obtained, the TOF can be found by integrating the variable of time along the ray path. Assisted by geometric analysis, an analytical relationship between TOF and penetration depth can be established by repeating the ray tracing algorithm for different penetration depths. Experimental measurements of TOF have been performed to validate the numerical results.

The ultimate goal of this research is to monitor and control the weld penetration depth in real-time by using laser generated ultrasound. The weld penetration depth is one of the most important geometric parameters that define weld quality. The analytical relationship between the TOF and penetration depth established by the ray tracing algorithm provides insight into ultrasound propagation in high-temperature gradient media, and offers valuable guidance on the sensing system design.

a)Electronic mail: charles.ume@me.gatech.edu

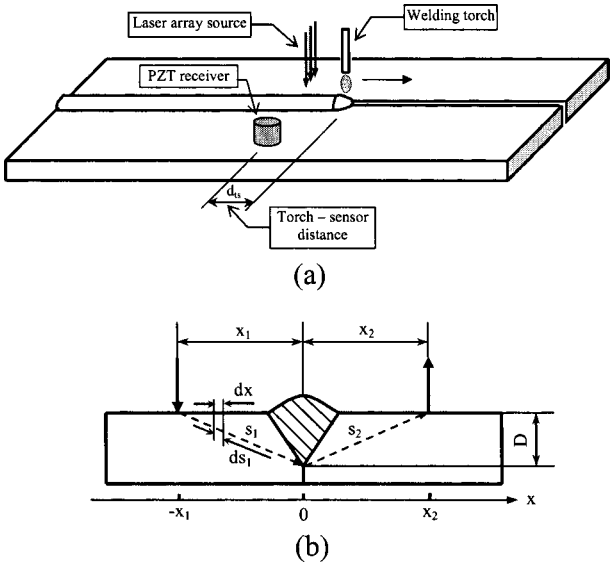


FIG. 1. Illustration of weld penetration depth monitoring with the TOF technique. (a) Laser ultrasonic system configuration. (b) Diagram of cross-section.

II. DERIVATION OF RAY EQUATIONS BASED ON FERMAT'S PRINCIPLE

A. Fermat's principle

Fermat's principle states that a ray path between any two points is a path of minimum propagation time.[15] The propagation time from one point P_0 to another P_1 is expressed by Eq. (2), where $n(x,y,z)$ is used to denote the slowness, namely the reciprocal of speed $\nu(x,y,z)$:

$$T=\int_{P_0}^{P_1} dt=\int_{P_0}^{P_1}\frac{ds}{\nu(x,y,z)}=\int_{P_0}^{P_1} n(x,y,z)ds. \tag{2}$$

When the ray path is expressed as parametric form with the independent variable t (note that the independent variable t could be time, but not necessarily), the differential ds of the path is given by Eq. (3):

$$ds=\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2}\,dt=\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}dt. \tag{3}$$

Then Eq. (2) can be expressed as Eq. (4):

$$T=\int_{P_0}^{P_1} n(x,y,z)\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}dt$$

$$=\int_{P_0}^{P_1} F(x,y,z,\dot{x},\dot{y},\dot{z})dt. \tag{4}$$

B. Calculus of variations

According to the calculus of variations, the minimum of the integral expressed by Eq. (4) must satisfy the Euler–Lagrange equation described by Eqs. (5a)–(5c).[15]

$$\frac{\partial F}{\partial x}-\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right)=0, \tag{5a}$$

$$\frac{\partial F}{\partial y}-\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{y}}\right)=0, \tag{5b}$$

$$\frac{\partial F}{\partial z}-\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{z}}\right)=0. \tag{5c}$$

Plugging the expression of $F(x,y,z,\dot{x},\dot{y},\dot{z})=n(x,y,z)\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}$ into Eqs. (5a)–(5c) gives Eqs. (6a)–(6c):

$$\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\frac{\partial n}{\partial x}-\frac{d}{dt}\left(\frac{n\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}}\right)=0, \tag{6a}$$

$$\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\frac{\partial n}{\partial y}-\frac{d}{dt}\left(\frac{n\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}}\right)=0, \tag{6b}$$

$$\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\frac{\partial n}{\partial z}-\frac{d}{dt}\left(\frac{n\dot{z}}{\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}}\right)=0. \tag{6c}$$

Since the velocity does not have a closed form expression, Eqs. (6a)–(6c) cannot be solved analytically, and a numerical technique has to be used. Equations (6a)–(6c) are second-order differential equations, and it is very difficult (if not impossible) to directly solve them numerically. In order to rewrite Eqs. (6a)–(6c) and express them as six first-order equations, the unit tangent vector $\mathbf{I}=(i_x,i_y,i_z)$ is introduced, and its three components are shown in Eqs. (7a)–(7c)

$$i_x=n(x,y,z)\dot{x}, \tag{7a}$$

$$i_y=n(x,y,z)\dot{y}, \tag{7b}$$

$$i_z=n(x,y,z)\dot{z}. \tag{7c}$$

By combining Eqs. (6a)–(6c) and Eqs. (7a)–(7c), Eqs. (8a)–(8c) are derived:

$$\frac{di_x}{dt}=i_x\frac{d}{dt}(\ln(\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}))+\frac{\partial n}{\partial x}(\dot{x}^2+\dot{y}^2+\dot{z}^2), \tag{8a}$$

$$\frac{di_y}{dt}=i_y\frac{d}{dt}(\ln(\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}))+\frac{\partial n}{\partial y}(\dot{x}^2+\dot{y}^2+\dot{z}^2), \tag{8b}$$

$$\frac{di_z}{dt}=i_z\frac{d}{dt}(\ln(\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}))+\frac{\partial n}{\partial z}(\dot{x}^2+\dot{y}^2+\dot{z}^2). \tag{8c}$$

C. Ray equations derived

When the independent variable t is considered as time, we have $\nu=\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}$ and $n=1/\nu$. Plugging the expressions of ν and n into Eqs. (8a)–(8c) and combining with Eqs. (7a)–(7c), the ray equations expressed in Eqs. (9a)–(9f) are obtained:

$$\frac{dx}{dt}=\nu i_x, \tag{9a}$$

$$\frac{dy}{dt}=\nu i_y, \tag{9b}$$

$$\frac{dz}{dt}=\nu i_z, \tag{9c}$$

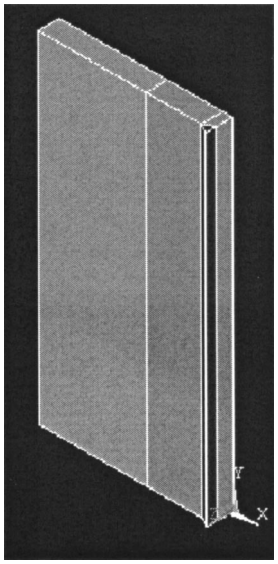


FIG. 2. Geometric modeling.

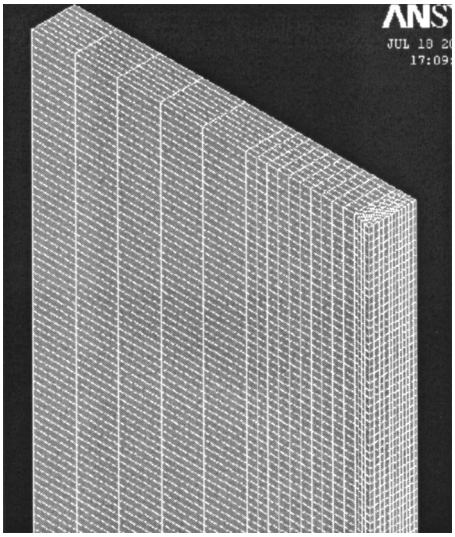


FIG. 3. Meshing of FE model.

$$\frac{di_x}{dt} = i_x \frac{d}{dt}(\ln \nu) - \frac{\partial \nu}{\partial x}, \tag{9d}$$

$$\frac{di_y}{dt} = i_y \frac{d}{dt}(\ln \nu) - \frac{\partial \nu}{\partial y}, \tag{9e}$$

$$\frac{di_z}{dt} = i_z \frac{d}{dt}(\ln \nu) - \frac{\partial \nu}{\partial z}. \tag{9f}$$

Given the initial values of source location **X** and propagation direction **I**, and the velocity (slowness) field over the region of interest, the entire ray path can be obtained by numerically integrating the ray equations. The numerical method for solving the ray equations is discussed in the next section.

III. NUMERICAL SOLUTION OF RAY EQUATIONS

A. Transient temperature distribution during gas metal arc welding (GMAW) by FEA

The temperature distribution in the welding specimen is required to solve the ray equations. The noncontact, real-time temperature measurement in the welding process is not readily available. Furthermore, such sensors would only be applicable to surface temperatures. The FEA has been proven to be a very economical and efficient approach to obtain temperature distributions in welding.[16–18]

Butt welding of two 1018 mild steel specimens has been performed with a robotic GMAW welding machine. The dimensions of the samples are 4 in.×8 in.×0.5 in. A FE model has been developed in Ansys 5.7 to obtain the transient temperature distribution. Only half of the weld is modeled due to symmetric conditions on geometry, boundary conditions, and heat inputs. Figure 2 shows the geometry of the half-plate. The model is meshed in Ansys with 3-D thermal brick elements. The vicinity of the weld path is meshed with finer elements because it has a greater temperature gradient than the rest of the model, as shown in Fig. 3.

An evenly distributed moving volume heat source is assumed. Based on earlier work,[19] the efficiency of heat input in the GMAW process is 0.65–0.90. A value of $\eta=0.7$ is used in this research. The net rate of thermal energy q supplied by the welding arc is expressed by Eq. (10):

$$q = \eta UI, \tag{10}$$

where η: heat transfer efficiency, U: welding voltage (V), I: welding current (A).

The heat generation corresponding to the volume in which the heat input is applied can be derived by dividing heat-input q by the volume. The continuous welding process is divided into 100 steps along the welding pathlength. During each step of 0.186 s (as a welding speed of 0.5 in./s), a stationary heat source is assumed. At the end of each step, the heat load is removed from the current volume unit and applied to the next unit. Temperature-dependent material properties shown in Table I are adopted for an accurate simulation result.

The initial condition is constant room temperature of 25 °C over the entire model. The boundary conditions include convection and radiation. With all the information described above, the simulation is carried out in Ansys 5.7.1. Figure 4 shows the temperature profiles at three time periods: $t=9$ s (when the welding torch reaches the middle of the weld), $t=18$ s (when the welding torch reaches the end of the weld), and $t=100$ s (after it cools down for about 80 s). The results look like a ship's wake in the water, which is ex-

TABLE I. Properties of mild steel (Refs. 16, 17).

Temperature (°C)	0	200	400	600	800	1000	1500
Coefficient of thermal conductivity k (W/m °C)	62	53	45	36	33	⋯	⋯
Density ρ (kg/m^3)	7860	7800	7750	7680	7600	7530	⋯
Enthalpy (J/m^3)×10^9	⋯	⋯	⋯	1.2	⋯	⋯	10
Specific heat capacity c (J/kg °C)				600			

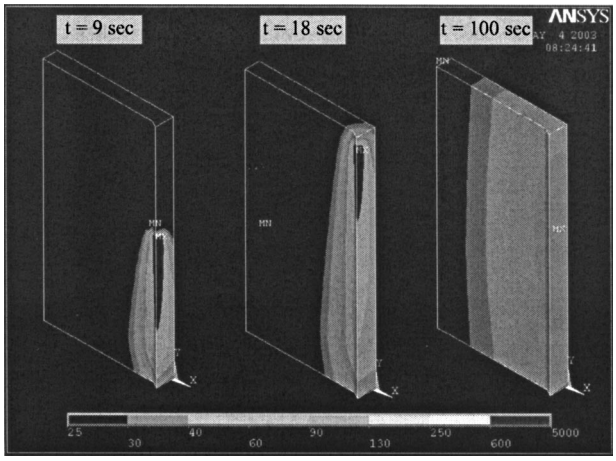


FIG. 4. Simulation results at three time periods.

pected. Since the welding arc moves across the plates faster than the heat propagates in the direction perpendicular to the welding direction, the wake is very narrow.

In order to verify the FEA results, temperatures are measured in selected locations on both the top and bottom of the weld samples, as shown in Fig. 5. The time history temperature profiles from both FEA simulation and experiments are compared in Fig. 6. The simulation results agree well with the experimental measurements at the specified locations. The difference between their peak values is less than 10%.

B. Temperature-dependent velocities of ultrasound

Once the temperature distribution is available, the ultrasound speed can be determined using empirical relationships between temperature and ultrasound speed. Scruby and Moss measured ultrasonic velocity in mild steel and published data for longitudinal and shear waves for 17 °C–1200 °C.[20] Figure 7 shows velocities of three major ultrasonic waves at different temperatures.

C. Shooting method to solve boundary value ray equations

Given initial values of source location and propagation direction and the velocity field obtained from last section, the entire ray path can be found by numerically integrating the ray equations given by Eqs. (9). In most real world applications, however, the ray tracing problem is a two-point boundary value problem, which means the source and destination locations are given.

For initial value problems, it is possible to start an acceptable solution at initial values and march it along by numerical integration to final values. For two-point boundary value problems, the boundary conditions at the starting point do not determine a unique solution to start with. A ''random'' choice among the solutions that satisfy these starting boundary conditions is almost certain not to satisfy the boundary conditions at the final point. The shooting method can be used to solve two-point boundary value problems. The strat-

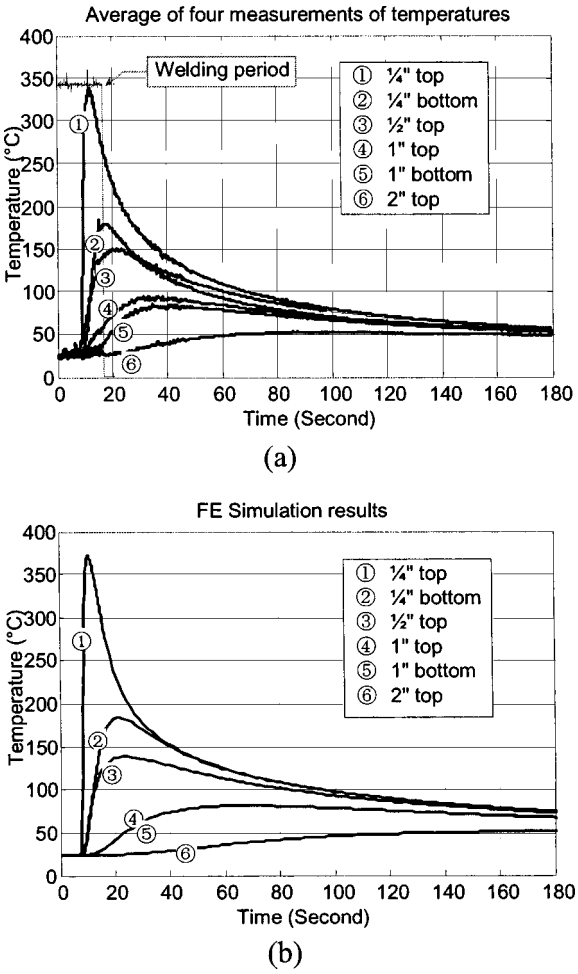


FIG. 6. Experimental verification of temperature distribution. (a) Temperature profile by experiments. (b) Temperature profile by FEA.

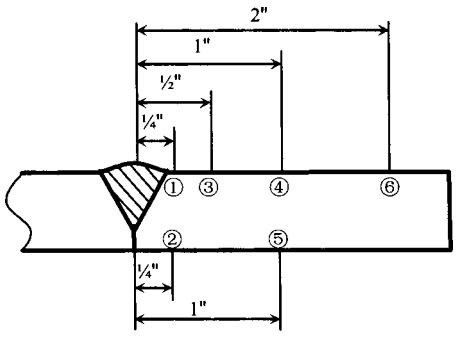


FIG. 5. Temperature measurement locations.

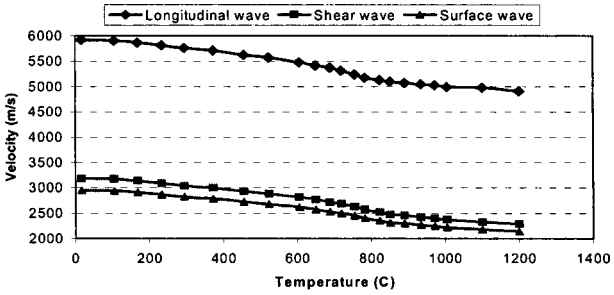


FIG. 7. Temperature-dependent velocities of ultrasound in mild steel (Ref. 18).

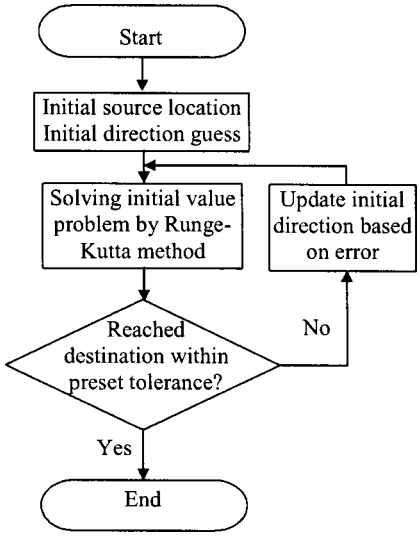


FIG. 8. Algorithm for shooting method.

egy is iteration on the initial value problem approach until the ray path arrives at the specified destination within a preset tolerance. The algorithm is depicted in Fig. 8.

D. Numerically traced ray for shear wave

The ray tracing algorithm is applied on half of the weld sample and the ray path for shear wave is traced. As shown in Fig. 9, the ultrasound propagation path obtained in real-time welding by ray tracing is no longer a straight line. The ray connecting the ultrasound source and the tip of weld is bent downward at the vicinity of the weld tip. This agrees with the Fermat's principle because the ray tries to find its way with higher traveling speed (lower temperature) so that it will take the least time to reach the destination.

IV. NUMERICAL TOF

Once the ray is obtained numerically, the TOF can be determined by integrating the time with Eq. (1) along the ray path [refer Fig. 1(b)]. Over a series of penetration depths, the ray tracing is performed and the TOF can be determined. Although the absolute TOF is obtained in this case, the relative TOF (relative to a reference value) is desired because it is more relevant to the applications. A relative TOF (relative to that with penetration depth of 5.5 mm) over penetration depth is plotted later in Fig. 14 for a comparison to experimental results.

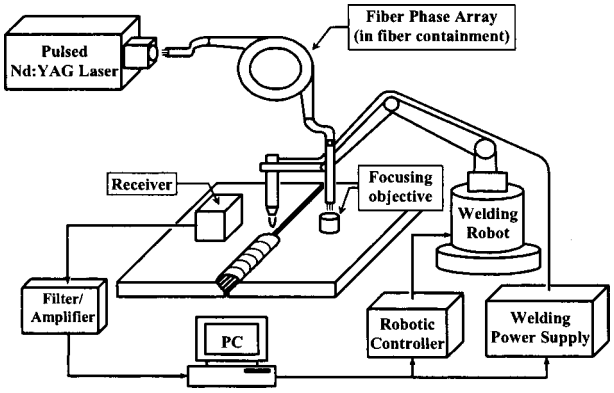


FIG. 10. Overall system setup.

V. EXPERIMENTAL VERIFICATION OF TOF

A. Experimental method and procedure

The experimental setup consists of the ultrasound generation and receiving subsystem, and welding subsystem. The overall system setup is shown in Fig. 10. The GMAW subsystem employs an Automatix AI 32v, General Electric Process Robot model P-50, and a Miller Pulstar 450 welding machine. The AI 32v running the RAIL v6.04 operating system controls the P-50 process robot. The welding torch is attached to the end effector of the P-50. An interface with the Pulstar 450 allows the AI 32v to control various welding parameters such as arc voltage, wire feed rate (welding current), and welding speed. The Pulstar 450 controls the flow of shielding gas, which is a mixture of 98% argon and 2% carbon dioxide.

A Nd:YAG Q-switched laser manufactured by Continuum is used to generate ultrasonic waves. This laser operates at a wavelength of 1064 nm with 5–7 ns pulses at up to 650 mJ per pulse, with a firing rate of 10 Hz. A fiber phased array contained in a copper ring is used to direct the laser onto the welding sample.[21] The fiber phased array is composed of three line sources. Each of the three sources consists of seven silica–silica fibers with a 400 μm core. The fibers in each of the three sources measure 2, 52, and 102 m,

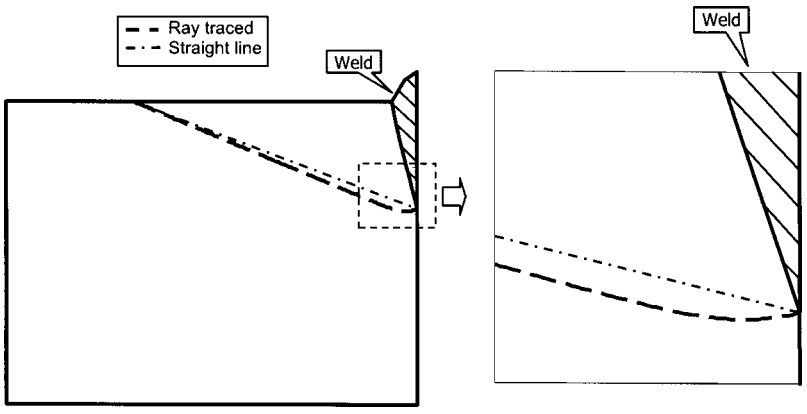


FIG. 9. Ray traced for shear wave.

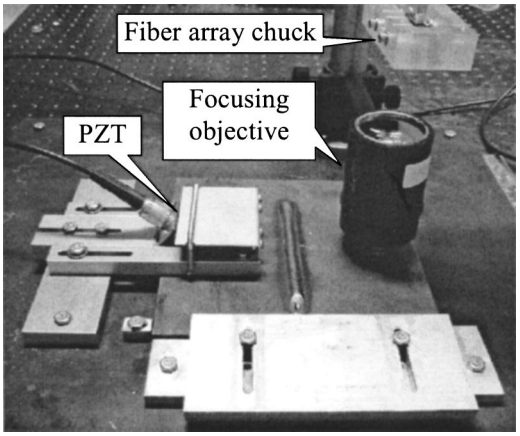


FIG. 11. Experimental setup.

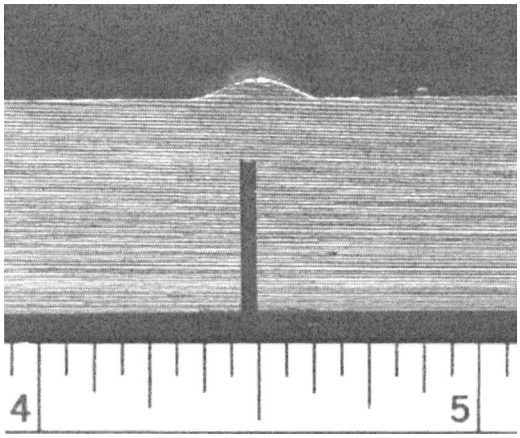


FIG. 13. Sample weld cross-section.

respectively. The difference in length introduces a time delay of 0.243 μs between adjacent line sources. As shown in Fig. 11, a fiber array chuck is designed to arrange the 21 fibers into three line sources with the source spacing adjustable. A custom designed focusing objective is used to focus the laser sources from the fiber array onto the sample. Two 70° angle beam PZTs consisting of 1 and 2.25 MHz of frequency are used to receive the ultrasonic signals. A fixture is designed to hold the angle beam PZT in place and apply constant pressure upon it so that the location of the PZT transducer is consistent between experiments and the signal strength is not influenced by the coupling. The typical signal received by a 1 MHz PZT is shown in Fig. 12. No longitudinal wave is received due to the configuration. The first received shear wave is marked on the plot and it is followed by surface wave and shear waves with multiple reflections.

B. TOF estimation with cross-correlation method

TOF refers to the time for the ultrasonic waves to travel from the source to the receiving point. The TOF can be measured by applying a threshold on the signal and locating the time when the signal crosses the threshold,[9] or by locating the signal's amplitude peak directly. When the signal-to-noise ratio is high and the amplitude is consistent, the methods using thresholding or locating the peak give reasonably accurate estimates of TOF. However, when the signal-to-noise ratio is low, the above methods become less accurate and reliable.

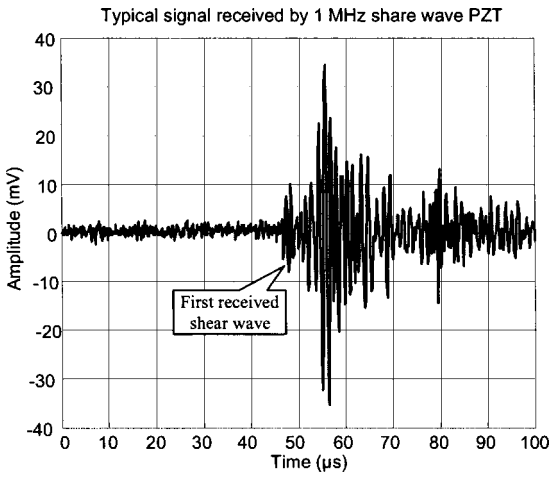


FIG. 12. Typical signal received by 1 MHz PZT.

In this paper, an overall matching method—the cross-correlation technique is applied to measure the TOF. Time delay measurements by the cross-correlation method have been used successfully with conventional pulsed ultrasonic contact techniques.[22,23] The cross-correlation method eliminates the need for the criterion that is usually affected by distortion or low signal-to-noise ratios, and hence gives the most reliable and reproducible results in difficult cases such as the welding environment. In this method, a reference signal is required. Each ultrasonic signal is correlated to the reference signal by the cross-correlation function. The relative TOF (difference between the real-time signal and the reference signal) is estimated by locating the peak of the cross-correlation function. The advantages of the cross-correlation method to estimate the TOF include reliability, insensitivity to random noise, and high resolution.

C. Comparison of experimental and numerical TOF results

The weld specimens are prepared with different groove sizes so that weld penetration depth varies in a desired range. After welding, all welds are cut perpendicular to weld path to measure the penetration depth. Figure 13 shows an image of a sample weld cross-section. The ruler in the image has the unit of inches.

The experimental and numerical results of TOF for a series of penetration depths are plotted in Fig. 14. Both 1.0 MHz PZT and 2.25 MHz PZT are used to measure ultrasonic signals. All experimental and numerical results of TOF are relative to the sample with a penetration depth of 5.5 mm, which is near the middle point of the investigated penetration depth range. The numerical results with ray tracing are obtained at a series of penetrations, and these discrete data points are connected with linear fitting. While the linearly fitted curve looks segmented and discontinuous, it is expected that the numerical results curve is continuous and smooth. Some disturbances are present in the numerical re-

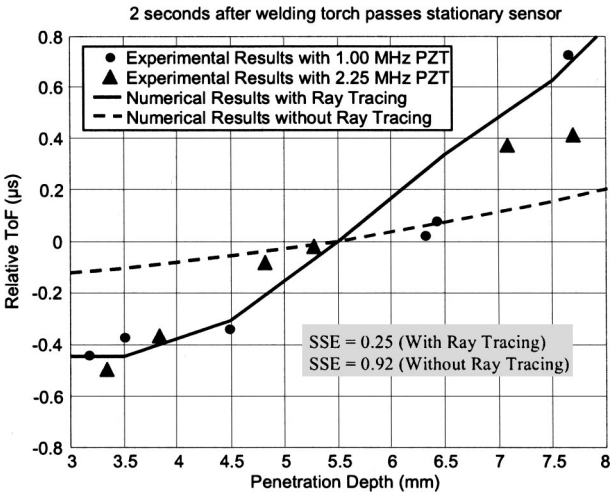


FIG. 14. A comparison of experimental and numerical TOF results.

sults. For example, the curve between the penetration depth of 6 and 7.5 mm is concave downward. These disturbances are not the real tread of the curve, but errors introduced by the numerical process, such as the FEA simulation of the temperature.

The sum of square error (SSE) is used to measure the deviation of experimental TOF from the numerical values. The smaller value of SSE means better agreement. The numerical results with ray tracing have smaller SSE than those without ray tracing.

VI. CONCLUSION

In this paper a 3-D ray tracing algorithm is developed directly from Fermat's principle. In solving the ray tracing equations, a FEA model is developed in Ansys to predict transient temperature distribution. The shooting method is applied to numerically trace the rays specified by two boundary locations. GMAW experiments have been performed and the experimental TOF is used to verify the numerical results. When SSE is used as a criterion, the experimental TOF results match the numerical TOF with ray tracing better than the numerical TOF when a straight path is assumed (without ray tracing).

The results show that the effects of temperature gradients on ultrasound propagation can be understood and calculated. The established relationship between the TOF and weld penetration depth will be used in the design of real-time weld penetration depth monitoring and control system. The replacement of the PZT receiver by an EMAT is being investigated so that both the generation and reception of ultrasound are noncontact, which is more practical in real-time welding situations.

ACKNOWLEDGMENT

This research was funded by The Manufacturing Machines and Equipment Program of the National Science Foundation under Grant No. DMI-9908082.

[1] J. Yang, N. DeRidder, C. Ume, and J. Jarzynski, "Non-contact optical fibre phased array generation of ultrasound for non-destructive evaluation of materials and processes," Ultrasonics **31**, 378–394 (1993).
[2] S. Dixon, C. Edwards, and S. B. Palmer, "A laser-EMAT system for ultrasonic weld inspection," Ultrasonics **37**, 273–281 (1998).
[3] S. W. Kercel, R. A. Kisner, M. B. Klein, G. D. Bacher, and B. Pouet, "In-process detection of weld defects using laser-based ultrasound," Proceedings of the SPIE **3852**, 81–92 (1999).
[4] P. Ridgway, R. Russo, E. Lafond, C. Habeger, Jr., and T. Jackson, "Laser ultrasonic in-process inspection of paper for elastic properties," J. Acoust. Soc. Am. **112**, 2350 (2002).
[5] M. A. Cornwell and Y. H. Berthelot, "Laser ultrasonics in copy paper: Bending stiffness dependence on temperature and moisture content," J. Acoust. Soc. Am. **112**, 2763–2770 (2002).
[6] S. Liu, D. Erdah, I. C. Ume, and A. Achari, "A novel approach for flip chip solder joint quality inspection: laser ultrasound and interferometric system," IEEE Transactions on Components and Packaging Technologies, **24**, 616–624 (2000).
[7] G. H. Glover and J. C. Sharp, "Reconstruction of ultrasound propagation speed distributions in soft tissue: time-of-flight tomography," IEEE Trans. Sonics Ultrason. **SU-24**, 229–234 (1977).
[8] F. A. Mauer, S. J. Norton, Y. Grinberg, D. Pitchure, and H. N. G. Wadley, "An ultrasonic method for reconstructing the two-dimensional liquid–solid interface in solidifying bodies," Metall. Trans. B **22**, 467–473 (1991).
[9] M. Miller, B. Mi, A. Kita, and I. C. Ume, "Development of automated real-time data acquisition system for robotic weld quality monitoring," Mechatronics **12**, 1259–1269 (2002).
[10] G. M. Graham and I. C. Ume, "Automated system for laser ultrasonic sensing of weld Penetration," Mechatronics **7**, 711–721 (1997).
[11] V. Y. Grechka and G. A. McMechan, "3-D two-point ray tracing for heterogeneous weakly transversely isotropic media," Geophysics **61**, 1883–1894 (1996).
[12] H. Sadeghi, S. Suzuki, and H. Takenaka, "A two-point, three-dimensional seismic ray tracing using genetic algorithms," Phys. Earth Planet. Inter. **113**, 355–365 (1999).
[13] D. T. Queheillalt, Y. Lu, and H. N. G. Wadley, "Laser ultrasonic studies of solid–liquid interfaces," J. Acoust. Soc. Am. **101**, 843–853 (1997).
[14] J. A. Johnson, N. M. Carlson, and L. A. Lott, "Ultrasonic wave propagation in temperature gradients," Journal of Nondestructive Evaluation **6**, 147–156 (1987).
[15] R. Weinstock, *Calculus of Variations: With Applications to Physics and Engineering* (Dover, New York, 1952), pp. 67–71.
[16] S. Prasad and T. K. S. Narayanan, "Finite element analysis of temperature distribution during arc welding using adaptive grid technique," Weld. J. (London) **75**, 123s–128s (1996).
[17] E. Kannatey-Asibu, N. Kikuchi, and A. Jallad, "Experimental finite element analysis of temperature distribution during arc welding," J. Eng. Mater. Technol. **111**, 8–18 (1989).
[18] J. Goldak and M. Bibby, "Computer modeling of heat flow in welds," Metall. Trans. B **17**, 587–600 (1986).
[19] D. Radaj, *Heat Effects of Welding* (Springer-Verlag, Berlin, 1992), pp. 20–30.
[20] C. B. Scruby and B. C. Moss, "Non-contact ultrasonic measurements on steel at elevated temperatures," NDT & E Int. **26**, 177–188 (1993).
[21] S. N. Hopko and I. C. Ume, "Laser generated ultrasound by material ablation using fiber optic delivery," Ultrasonics **37**, 1–7 (1999).
[22] J. D. Aussel and J. P. Monchalin, "Precision laser-ultrasonic velocity measurement and elastic constant determination," Ultrasonics **27**, 165–177 (1989).
[23] D. R. Hull, H. E. Kautz, and A. Vary, "Measurement of ultrasonic velocity using phase-slope and cross-correlation methods," Mater. Eval. **43**, 1455–1460 (1985).

Propagating and evanescent elastic waves in cylindrical waveguides of arbitrary cross section

Vesna Damljanović and Richard L. Weaver[a)]
Department of Theoretical & Applied Mechanics, University of Illinois at Urbana—Champaign, Urbana, Illinois, 61801-2983

(Received 12 August 2003; revised 18 January 2004; accepted 24 January 2004)

We study waves in elastic waveguides, with a view toward the nondestructive evaluation of slender structures by means of imposed vibrations. Envisioned applications demand an accurate understanding of both propagating and evanescent guided waves in waveguides of arbitrary cross section. Accordingly, we develop a theoretical framework in which energy principles and finite element discretization lead to a discrete set of solutions representing both wave types. We examine the solutions in great detail, with a particular emphasis on the accuracy of the finite element discretization. Results are compared with analytic solutions of the Pochhammer–Chree equations for the special case of a circular cross section, determining the combination of mesh parameters and frequency regimes for which the code yields accurate results. Convergence studies are conducted for the case of a more complex cross section, that of a typical railroad rail. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1687424]

PACS numbers: 43.40.Cw, 43.20.Mv, 43.35.Cg [ANN] Pages: 1572–1581

I. INTRODUCTION

Waveguides, as elastic slender bodies of a constant cross section with properties that are constant in the preferred direction(s), are common in engineering practice. Accurate predictive capability regarding their vibrations and acoustics, followed by comparison with measurements, can be used for a nondestructive evaluation (NDE) of such bodies. One such application is material characterization, e.g., a determination of the level of anisotropy, internal friction, or contained stress. Ultrasonic inspection and acoustic emission from flaws also require understanding of propagation in such bodies. The acoustics and vibrations of structures composed of plates, shells, and beams depend on understanding propagation in these components in more detail than is provided by simple beam and plate theories of the strength of materials.

The complexity of time-harmonic wave propagation in linear elastic waveguides arises from having two or more parallel boundaries that induce wave dispersion characterized by the dependence of frequency or phase velocity on wavelength. Multiple reflections of elastic waves from the lateral boundaries lead to a complicated elastic wave field that is not easily analyzed in terms of the familiar P and S waves. In an infinite waveguide, however, translational invariance implies that each natural mode of vibration has single wave number k and corresponding simple variation (e^{ikz}) along the waveguide, while variations in the other direction(s) remain complicated. In general, a multibranched dispersion relation is expected.

Pochhammer in 1876, and Rayleigh and Lamb in 1889, laid the foundations for the contemporary studies of waveguides. Pochhammer derived the general frequency equations for an infinite circular cylinder with a traction-free mantle and resolved it into various modes, while Rayleigh and Lamb derived the analogous equations for an infinite plate in plane strain and with traction-free faces.[1] The dispersion relation for flexural waves in a circular rod was studied by Pao and Mindlin,[2,3] who constructed approximate real and imaginary branches by analytically establishing a grid of boundary curves and asymptotic behavior. Onoe, McNiven, and Mindlin[1,4–6] found the complete frequency spectrum for longitudinal waves in a circular rod, including the complex branches. However, little attention has been given to wave propagation in rods of arbitrary cross section, and even less to the imaginary and complex branches of all the modes. Rosenfeld and Keller[7,8] found asymptotic expansions for both long and short waves propagating in elastic rods of arbitrary cross section, but did not calculate solutions for individual branches. Gavrić[9] developed a finite-element technique for computation of full dispersion curves for the purpose of application in the high-frequency regimes relevant for noise generation in rails. Although his formulation, in principle, enables the calculation of imaginary and complex branches as well as real, he did not obtain any complex or imaginary solutions. He focused only on real branches, i.e., propagating modes, and compared his results with a set of experimental results obtained for a rail of finite length, without a reference to existing analytic solutions and without confirming the accuracy of his solutions. Similar approaches without attention to evanescent modes were presented by Aalami[10] and recently by Widehammar *et al.*[11] and Hayashi *et al.*[12] Pfaffinger and Dual[13] recently presented a finite element analysis of the cutoff frequencies with special attention to the effects of damping. Venkatesan and co-workers[14] have applied a Fourier–Bessel expansion for the interior fields in lieu of finite elements, and then collocated boundary conditions on the surface to derive an alternative numerical method for such waves.

The accurate calculation of all the real and complex

a)Author to whom correspondence should be addressed; electronic mail: r-weaver@uiuc.edu

 0001-4966/2004/115(4)/1572/10/$20.00

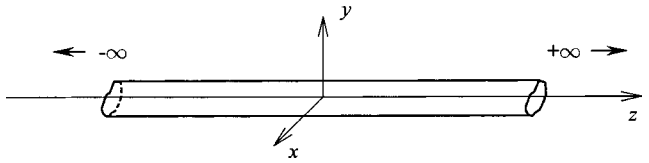


FIG. 1. Cylindrical waveguide of constant (arbitrary) cross section.

branches is necessary for a complete study of wave propagation in a beam of arbitrary cross section. According to Saint Venant's Principle, in a finite waveguide, the influence of evanescent waves is noticeable for shorter guides, or near the loads or ends. In an infinite waveguide, evanescent waves are important in the vicinity of loads and supports. While one expects evanescent waves to attenuate over distances of the order of the waveguide diameter, accurate NDE requires a more quantitative estimate of their influence, especially if measurements are made in the vicinity of such heterogeneities. Due attention was given to evanescent waves in some studies,[15,16] but the problem was solved for the special cases of composite or pretwisted beams. In order to generalize the assumptions employed in Refs. 9, 15, 16, Damljanović and Weaver[17] obtained the dispersion relation by first defining a Hamilton's principle for complex fields. A more rigorous approach is used here, employing the semianalytical discretization of Kantorovich–Krilov[18,19] and arriving at a dispersion relation identical to that of Refs. 9, 11, 12, 17.

In this paper we begin with the formulation of the Lagrangian for free vibration of a uniform cylindrical waveguide of arbitrary cross section. In Sec. III, the finite-element discretization of the cross section is employed and the Euler–Lagrange equations are invoked; a search for steady-state harmonic solutions then yields the dispersion relation. The resulting numerically obtained dispersion curves are discussed in detail in Sec. IV. Numerical convergence for the cases of a circular cross-section (for which there exists an analytic solution) and a complex cross section (e.g. rail, for which there is no analytic solution) is discussed in Sec. V.

II. ANALYTIC FORMULATION

Consider an isotropic elastic cylindrical waveguide of cross section that is constant along the z axis of a rectangular Cartesian coordinate system. The lateral boundary is traction-free (Fig. 1). In anticipation of partial FE discretization of the cross section, while keeping the z and t dependence analytic, the following representation for the admissible displacement field due to free vibration at constant frequency ω is adopted:

$$\{\mathcal{U}\}=\begin{Bmatrix} U(x,y;z,t) \\ V(x,y;z,t) \\ W(x,y;z,t) \end{Bmatrix}. \tag{1}$$

The Lagrangian for and infinitesimal volume of the waveguide is equal to the difference between the kinetic energy and total potential energy:[20]

$$\begin{aligned} \mathrm{d}L=\frac{1}{2}\Bigg\{ & \rho(\dot{U}^2+\dot{V}^2+\dot{W}^2)-\Bigg[\frac{E}{(1+\nu)(1-2\nu)}\Bigg(\left(\frac{\partial U}{\partial x}\right)^2 \\ & +\left(\frac{\partial V}{\partial y}\right)^2-\nu\left(\frac{\partial U}{\partial x}-\frac{\partial V}{\partial y}\right)^2\Bigg) \\ & +\frac{2\nu E}{(1+\nu)(1-2\nu)}\frac{\partial W}{\partial z}\left(\frac{\partial U}{\partial x}+\frac{\partial V}{\partial y}\right) \\ & +\frac{(1-\nu)E}{(1+\nu)(1-2\nu)}\left(\frac{\partial W}{\partial z}\right)^2 \\ & +\frac{E}{2(1+\nu)}\left(\left(\frac{\partial W}{\partial x}\right)^2+\left(\frac{\partial W}{\partial y}\right)^2+\left(\frac{\partial U}{\partial y}+\frac{\partial V}{\partial x}\right)^2\right) \\ & +\frac{E}{2(1+\nu)}\left(\left(\frac{\partial U}{\partial z}\right)^2+\left(\frac{\partial V}{\partial z}\right)^2\right) \\ & +\frac{E}{1+\nu}\left(\frac{\partial W}{\partial y}\frac{\partial V}{\partial z}+\frac{\partial W}{\partial x}\frac{\partial U}{\partial z}\right)\Bigg]\Bigg\}\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z, \end{aligned} \tag{2}$$

where overdots represent time derivatives. The same quantity, written in matrix form, is

$$\begin{aligned} \mathrm{d}L=\frac{1}{2}\{\mathcal{U}\}^T & [\overleftarrow{\partial}_t[m]\overrightarrow{\partial}_t-\overleftarrow{\partial}_z[\kappa_2]\overrightarrow{\partial}_z-[D]^T[\kappa_1]\overrightarrow{\partial}_z \\ & -[D]^T[\kappa_0][D]]\{\mathcal{U}\}\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z, \end{aligned} \tag{3}$$

where the operator $\overleftarrow{\partial}$ denotes the derivative (with respect to time or z) of the vector to the left of it and $\overrightarrow{\partial}$ denotes the derivative of the vector to the right of it. The material matrices are constant:

$$[\kappa_0]=\begin{bmatrix} \frac{1-\nu}{\nu}\lambda & 0 & 0 & \lambda & 0 & 0 \\ 0 & \mu & \mu & 0 & 0 & 0 \\ 0 & \mu & \mu & 0 & 0 & 0 \\ \lambda & 0 & 0 & \frac{1-\nu}{\nu}\lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{bmatrix},$$

$$[\kappa_1]=\begin{bmatrix} 0 & 0 & 2\lambda \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2\lambda \\ 2\mu & 0 & 0 \\ 0 & 2\mu & 0 \end{bmatrix}, \tag{4}$$

$$[m]=\begin{bmatrix} \rho & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho \end{bmatrix}, \quad [\kappa_2]=\begin{bmatrix} \mu & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \frac{1-\nu}{\nu}\lambda \end{bmatrix}.$$

Here, $[D]$ is the operator matrix defined in the Appendix, and λ and μ are Lamé constants:[21]

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)} \quad \text{and} \quad \mu = \frac{E}{2(1+\nu)}. \tag{5}$$

III. STEADY-STATE FINITE ELEMENT SOLUTION

Leaving the axial (z) and time (t) dependences of the displacement vector as arbitrary, the cross section is discretized using three-noded triangles (with the commercially available FE preprocessor MSC. Patran) and linear interpolation functions—details of the derivation are in the Appendix. These linear constant-strain elements satisfy the continuity requirement C^0 imposed by the highest-order derivative with respect to x and y in the Lagrangian. Their shape functions interpolate the displacement field linearly between the nodes, resulting in continuity of displacement between the elements, and discontinuity of its derivatives of order 1 and higher. Furthermore, the same shape functions are used for interpolation of both the nodal coordinates and the nodal displacements (isoparametric elements[21]). The result of this partial discretization is the Lagrangian for a single, generic cross-section element of infinitesimal length $\mathrm{d}z$:

$$\begin{aligned} L^{\mathrm{e}} = \frac{1}{2}\{d^{\mathrm{e}}(z,t)\}^T \Bigg[\int_0^1 \int_0^{1-\eta} & [\overleftarrow{\partial}_t [N]^T [m] [N] \vec{\partial}_t \\ & - \overleftarrow{\partial}_z [N]^T [\kappa_2] [N] \vec{\partial}_z - [B]^T [\kappa_1] [N] \vec{\partial}_z \\ & - [B]^T [\kappa_0] [B]] \det[J] \mathrm{d}\xi \, \mathrm{d}\eta \Bigg] \{d^{\mathrm{e}}(z,t)\} \mathrm{d}z, \end{aligned} \tag{6}$$

where $\{d^{\mathrm{e}}(z,t)\}$ is the array of nine nodal displacements, that is, degrees of freedom (d.o.f.) of the element, as defined in Refs. 21–23. Integration over the element using a three-point Gauss quadrature and assembling into global form yield the Lagrangian for the entire waveguide:

$$\begin{aligned} L = \frac{1}{2} \int \{d(z,t)\}^T & [\overleftarrow{\partial}_t [M] \vec{\partial}_t - \overleftarrow{\partial}_z [K_2] \vec{\partial}_z - [K_1] \vec{\partial}_z \\ & - [K_0]] \{d(z,t)\} \mathrm{d}z = \int \mathcal{L} \mathrm{d}z, \end{aligned} \tag{7}$$

where $\mathcal{L}$ is Lagrangian lineal density, $\{d(z,t)\}$ is the array of $3N$ global d.o.f. and $[M]$, $[K_2]$, $[K_1]$ and $[K_0]$ are global ("mass" and "stiffness") matrices of order $3N \times 3N$. Introducing indicial notation, with $d_j = d_j(z,t)$ being the jth component of $\{d(z,t)\}$, the Euler–Lagrange equations are applied, as defined in Refs. 20 and 22:

$$\frac{\partial \mathcal{L}}{\partial d_i} - \frac{\mathrm{d}}{\mathrm{d}z} \frac{\partial \mathcal{L}}{\partial d_i'} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{d}_i} = 0, \tag{8}$$

and the governing equation (in indicial form) is

$$\begin{aligned} & (\mathrm{sym}[K_2])_{ij} d_j'' + (\mathrm{skw}[K_1])_{ij} d_j' - (\mathrm{sym}[K_0])_{ij} d_j \\ & - (\mathrm{sym}[M])_{ij} \ddot{d}_j = 0. \end{aligned} \tag{9}$$

The prime and double prime denote first and second derivatives with respect to z, respectively, and the double overdot denotes the second temporal derivative. This equation represents the loci of stationary points of the action over the restricted space spanned by all possible vectors $\{d(z,t)\}$.

In this and all subsequent FE derivations, the governing equations are normalized so that Young's modulus and density are $E=1$ and $\overline{\rho}=1$, respectively. Positions x, y, z and displacements d_j are taken in inches (pursuant to applications envisioned with US rail dimensions). Therefore, the normalized bar wave speed is $\overline{c_b} = \sqrt{E/\overline{\rho}} = 1$ and the normalized frequency is $\overline{\omega} = \omega / c_b$ (rad/in.). This normalization does not affect wave numbers; that is, $k = k$ in units of in^{-1}. It is understood hereon that all the variables are in a normalized form, but the bars are dropped for convenience, except where needed for clarity.

IV. SOLUTION TO EIGENVALUE PROBLEM—DISPERSION CURVES

Guided wave solutions are of the form

$$d_j(z,t) = h_j e^{i\omega t - ikz}, \tag{10}$$

where h_j are constants. Substituting (10) into the governing equation (9) and changing back to matrix notation yield the following quadratic eigenvalue problem (EVP):

$$(k^2 \, \mathrm{sym}[K_2] + \mathbf{i} k \, \mathrm{skw}[K_1] + \mathrm{sym}[K_0] - \omega^2 \, \mathrm{sym}[M])\{h\} = 0. \tag{11}$$

In order to lose the imaginary unit, a $3N \times 3N$ unitary transformation matrix $[T]$ is introduced, with each diagonal element equal to 1, except every third (corresponding to the displacement in the z direction) being equal to $\mathbf{i}$. The matrix $[T]$ has the property that $[T]^*[T] = [T][T]^* = [I]$, where an asterisk denotes the complex conjugate. Equation (11) is multiplied from the left by $[T]$ and $\{h\}$ is replaced by $[T]^*[T]\{h\} = \{h\}$. The symmetric matrices $[K_2]$, $[K_0]$, and $[M]$ do not mix the w component of displacement with u and v. Therefore,

$$\begin{aligned} & [T][K_2][T]^* = [K_2], \\ & [T][K_0][T]^* = [K_0], \\ & [T][M][T]^* = [M]. \end{aligned} \tag{12}$$

On the other hand, the skew matrix $[K_1]$ mixes u and v with w but does not mix u and v displacement components with each other, so it follows that

$$[T][K_1][T]^* = -\mathbf{i}[\hat{K}_1], \tag{13}$$

where $[\hat{K}_1]$ is a real and symmetric matrix. Defining a new vector $\{\hbar\} = [T]\{h\}$, one obtains the final form of the eigenvalue problem:

$$(k^2[K_2] + k[\hat{K}_1] + [K_0] - \omega^2[M])\{\hbar\} = 0. \tag{14}$$

The matrices in Eq. (14) are real and symmetric and of order equal to the total number of d.o.f. used in the FE approximation, i.e., $3N$, where N is number of nodes. This equation may be compared to those of Gavrić[9] and others.[11,12,17]

At fixed real k (i.e., for propagating modes) Eq. (14) is a $3N \times 3N$ real symmetric positive-definite generalized eigenvalue problem with real positive roots ω^2. If interest is restricted to propagating modes (as it is in most previous studies[9–12]) then it may be most efficient to specify k and

then solve for ω. Our interest is, however, in highly accurate complex k's at specified real ω, where the simpler approach is not useful. Our objective is to find the wave numbers k that satisfy Eq. (14). A simple examination of this equation reveals that, considering the expression in parentheses is quadratic in k and the matrices are of order $3N$, the determinant is a polynomial of order $2\times 3N$. Therefore, the vanishing determinant should result in $6N$ solutions for k, which are generally distinct (providing there is no degeneracy) and complex. The complex solutions should appear in pairs of complex conjugates.[24] Furthermore, due to reflection symmetry in z, it is clear that if some k is a solution, then $-k$ should also be a solution.[25]

For $k=0$, Eq. (14) reduces to a generalized EVP with real, symmetric, positive-definite matrices:

$$([K_0]-[M]\omega^2)\{\hbar\}=\{0\}. \tag{15}$$

Solutions to this equation, the eigenvalues $\pm\omega_{\text{c.o.}}$, are real and are known in elastodynamics as cutoffs—the frequencies at which some of the evanescent waves (imaginary branches) become propagative (real branches) and all the points of the waveguide vibrate in phase. Much is known about the convergence of solutions of generalized EVPs. Most notably, according to the Eigenvalue Separation Theorem,[26] for the complete approximation space, the eigenvalues of the $N+1$ d.o.f. model bracket the eigenvalues of the N d.o.f. model and so, as N increases, the approximate eigenvalues approach the actual ones asymptotically and from above (the actual eigenvalues being the lower bounds for approximate ones). However, there is no similar rule for the evanescent waves, which appear together with propagating waves in the case of the full-fledged problem, $k\neq 0$ in Eq. (14). In order to benefit somewhat from the above-mentioned theorem, a parallel is drawn between the current problem and the generalized EVP.

In a manner that bears similarity to the state-space reformulation of dynamics, Eq. (14) is rewritten in the form of a generalized EVP. A new vector is introduced:

$$\{\xi\}=k\{\hbar\}, \tag{16}$$

and multiplied with the matrix $[K_2]$:

$$[K_2]\{\xi\}=k[K_2]\{\hbar\}. \tag{17}$$

Due to $[K_2]$ being nonsingular, the above multiplication incurs no loss of information. Combining Eqs. (14) and (17), one obtains the following generalized EVP:

$$[A]\{\psi\}=k[B]\{\psi\}, \tag{18}$$

with

$$[A]=\begin{bmatrix}[K_2] & [0]\\ [0] & \omega^2[M]-[K_0]\end{bmatrix},$$

$$[B]=\begin{bmatrix}[0] & [K_2]\\ [K_2] & [\hat{K}_1]\end{bmatrix},\quad \{\psi\}=\begin{Bmatrix}k\{\hbar\}\\ \{\hbar\}\end{Bmatrix}. \tag{19}$$

Both $6N\times 6N$ matrices are real and symmetric—symmetry of the matrix on the right-hand side is ensured by the use of $[K_2]$ in (17)—although not positive-definite. Just like Eq. (14), this equation has $6N$ distinct, generally complex solution pairs $(k_r,\{\psi_r\})$, with $r=1,...,6N$.

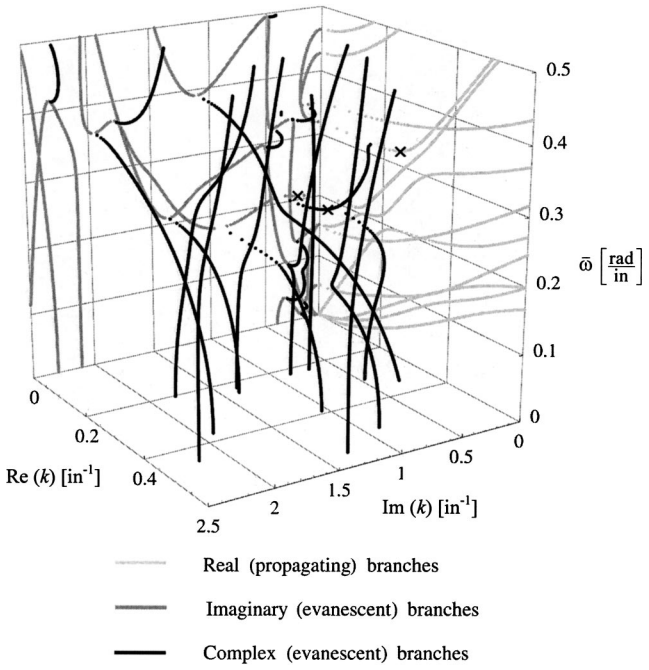


FIG. 2. Dispersion curves for the rail 136 lb AREMA (mesh #2), $\nu=0.3$.

Full dispersion curves (complex k as a function of real ω) were obtained for the case of 136 lb AREMA rail[27] of Poisson ratio $\nu=0.3$ using a moderately refined mesh. A computer code was written to obtain eigenvalue wave numbers k_r, for frequencies ω from just above 0 through 0.5 (the 7th cutoff), in densely spaced steps of $\Delta\omega=0.001$ (in normalized units). Equation (18) was solved using NAG (Numerical Algorithms Group) Fortran library routines that compute all the eigenvalues and eigenvectors of a real, symmetric but not necessarily positive-definite generalized EVP (for branches) or real symmetric-definite generalized EVP (for cutoffs).

The first quadrant of the resulting dispersion curves for the rail, with all the cutoffs and branches, is shown in Fig. 2. To the authors' knowledge, this is the first time that a complete frequency spectrum for the cylindrical waveguide of a noncircular cross section has been presented in the literature. In order to keep the computational burden within reasonable limits, coarse mesh #2 (defined below) was used to compute the data for this plot. Because the EVP was solved for one ω at a time, the frequency spectrum in Fig. 2 was constructed of consecutive horizontal "slices" of Re(k)–Im(k)-ω space. Consequently, in regions where the branches become close to horizontal (i.e., at minima and maxima), the points that comprise them become sparse. The wave numbers obtained by the code came, as expected, mostly in quadruples, except for a few pairs. The purely real and purely imaginary solutions appeared in pairs of opposite signs. The fully complex ones appeared in quadruples of complex conjugates and opposite signs.

One way to look at the frequency spectrum of a waveguide is to study the behavior of the branches while increasing the frequency. Up to the first cutoff, there are exactly four purely real, propagating branches. These are the familiar bending, torsional, and extensional modes. In addition, there are many evanescent ones (mostly complex with few purely

imaginary). Above the first cutoff, the purely imaginary branch that reaches the ω axis passes through a new cutoff and continues as a new, purely real branch—the evanescent mode becomes propagating. New purely real branches appear, as the frequency is increased, only from the new cutoffs. On the other hand, most of the purely imaginary branches start and end on the Im(k) axis; some of them start at the Im(k) axis and end at the cutoffs on the ω axis, while a few of them arc between two neighboring cutoffs. Many complex branches start at the $\omega=0$ plane and collapse onto the minima of imaginary branches. Other complex branches interconnect the gaps between neighboring imaginary branches (from the minimum of one to the maximum of another).

Interestingly, while local inflection points occasionally exist on higher-order real branches, there are no local maxima and very few local minima. The lowest local minimum appears on the ninth real branch (that emanates from the fourth cutoff) and has a complex branch collapsing on it from its origin at the local maximum of the imaginary branch that ends at the third cutoff—the three involved branches are marked with an $\times$ in Fig. 2. This example shows effectively the role of local minima and maxima of imaginary branches as potential wells that trap (if close enough) or deflect (if farther away) the complex branches as they ''travel'' through the Re(k)–Im(k)-ω space.

The nature of local extrema in general is not well understood, but it was shown by Onoe and Mindlin (as reviewed by Graff[5]) that the position of local extrema is sensitive to Poisson's ratio and the geometry of the cross section. Therefore, the peculiar complex branch mentioned above would have different behavior or might not even exist for a waveguide with a different Poisson's ratio or cross section. For example, Onoe *et al.*[4] showed that, for a circle with $\nu=0.31$, the complex branch connecting to the local minimum of the first real branch (longitudinal mode), starts at the $\omega=0$ plane instead of at the local maximum of an imaginary branch. The very existence of local minima (and negative group velocity) on a real branch is still as striking as it was 40 years ago.

V. ACCURACY AND CONVERGENCE

A sequence of four meshes was used to assess the convergence of the FE solutions for the dispersion relation $k(\omega)$ in two cases—the case of a circular waveguide, for which there exists an analytic solution (Pochhammer–Chree equations) for comparison; and the case of a waveguide with an irregular cross section (rail 136 lb AREMA) for which there is no analytic solution. Structured h refinement[22] was used in both cases by dividing each triangle of the preceding mesh into four triangles, with one new node being created at the middle of each of the old triangle's sides. Therefore, the set of nodes of each preceding mesh is a subset of the set of nodes of the following mesh and each following mesh has four times more elements than the preceding one. Mesh characteristics are given in Table I and meshes for both cross sections are shown in Figs. 3 and 4.

Because the three-noded triangles used in this study have first-order polynomial interpolation functions, the appropriate mesh fineness parameter is $1/h$, where h is a characteristic length of the worst element (the element with the greatest surface area and/or sharpest angle). From the nested property of the mesh, it follows that $h \propto 1/\sqrt{N_{\mathrm{el}}}$, where N_{el} is the number of elements, and then it holds that

$$|\Omega^{\mathrm{c.o.}} - \Omega^{\mathrm{c.o.}}_{\mathrm{exact}}| \propto (\sqrt{N_{\mathrm{el}}})^{s} \quad \text{and, similarly,}$$
$$|\overline{k} - \overline{k}_{\mathrm{exact}}| \propto (\sqrt{N_{\mathrm{el}}})^{s}, \tag{20}$$

where $\Omega^{\mathrm{c.o.}}$ is the cutoff frequency of interest and the subscript ''exact'' should be taken with reserve, as the available analytic solution is approximate within the precision of the transcendental equation solver used to obtain it.

TABLE I. Number of nodes and elements for all meshes used in the study.

	Circle of diameter 1		Rail 136 lb AREMA	
No.	# of nodes	# of elements	# of nodes	# of elements
1	17	20	42	50
2	53	80	133	200
3	185	320	465	800
4	689	1280	1729	3200

It was considered appropriate to evaluate the convergence by plotting the dependence of the relative error of the frequency (for cutoffs) and wave numbers (for branches) on $\sqrt{N_{\mathrm{el}}}$, both in terms of natural logarithms. In such a coordinate system, the curve is expected to be linear for this type of mesh refinement in the limit as $h \to 0$. The rate of convergence is then equal to the slope s of the asymptotic convergence curve, and it represents the rate at which the accuracy of the solution increases with the mesh refinement. Finally, the vertical position of the convergence curve in the chosen coordinate system is the measure of accuracy of that particular solution—the lower the convergence curve is in the plot, the more accurate is the corresponding mode. The root finding error was judged to be less than the discretization error. This judgment is based on the machine precision, and on the accuracy with which the (exact) torsional roots were found for the circular cylinder.

It is important to notice that, at the curved edges, during the division of an old triangle into four new triangles, the new node in the middle of the triangle side that is following the curve has to be ''nudged'' onto the curve by nearest-point projection, so the four new triangles create an element whose surface does not coincide with the surface of the old triangle. This geometric discretization error (here referred to as ''node shifting'') affects the convergence curves of different modes to a varied degree, depending on the sensitivity of the particular mode shape to the location of the point on the cross section.

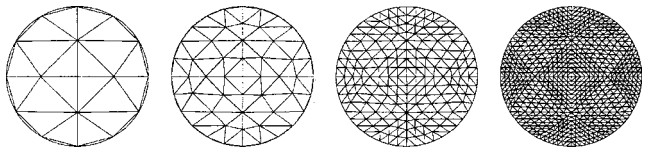


FIG. 3. Nested meshes for circle of diameter 1 in.

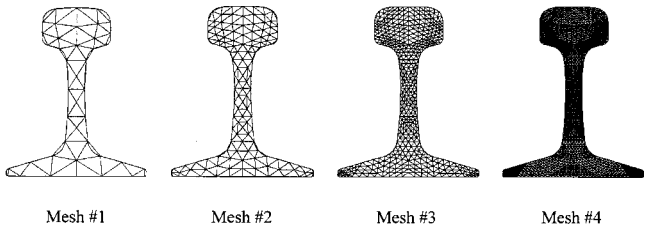


FIG. 4. Nested meshes for the rail 136 lb AREMA.

A. Circular cross section—comparison with exact solution

The Pochhammer–Chree equations for guided elastic waves in a circular cylinder are well studied.[1,2,3,5] We have solved these transcendental equations for the case of a cylinder of diameter $2a=1$ in., and Poisson ratio $\nu=0.3$, in terms of the nondimensional frequency $\Omega=\overline{\omega}ac_b/c_{\text{shear}}=\omega a/c_{\text{shear}}$ and nondimensional wave number $k=ka$. Our Pochhammer–Chree solutions are denoted with subscript "Poch."

Although the code solves for all the cutoffs, it was considered sufficient to study the convergence of only the lowest two cutoffs (lowest two flexural cutoffs of the order $n=1$ in the Pochhammer–Chree equations), mainly because our frequencies of interest are well below the first cutoff, but also because the accuracy is known to decrease considerably above the first cutoff for this type of problem. As seen in Fig. 5, the accuracy is lower but the convergence rate is only slightly higher for the higher cutoff frequency. There is no apparent ordering of levels of accuracy according to the cutoff order or within the families of branches with the same azimuthal order—it appears that only for some *n*-families the accuracy is lower for the higher branches, as might be suggested by intuition.

The lower real, imaginary, and complex solutions for a single frequency of $\Omega=1.0$ (about halfway below the first cutoff) were considered in order to present the convergence of the branches. The quantities "*s*" appearing here are the slopes of linear fits, and not necessarily equal to the familiar asymptotic convergence rates that should be achieved in the limit of a very fine mesh. In most cases, however, the asymptotic behavior does appear to be achieved, at $s\approx 2.0$.

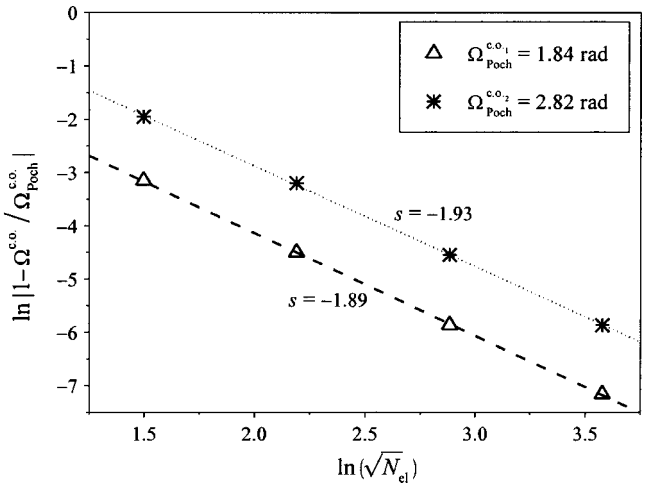


FIG. 5. Numerical convergence and convergence rates of selected cutoffs (for a circular rod).

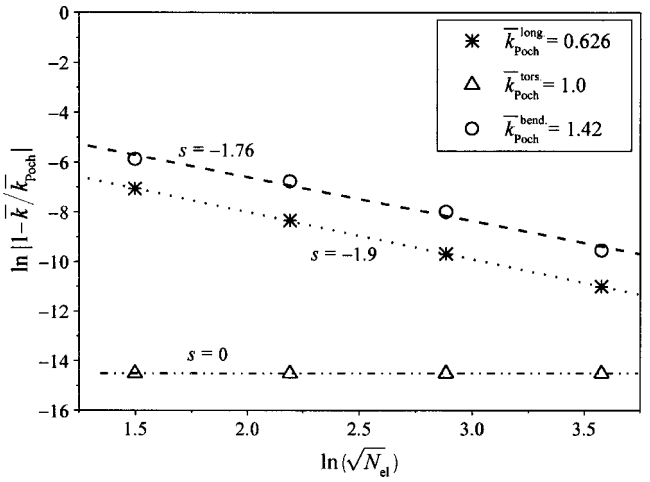


FIG. 6. Numerical convergence and convergence rates of selected real branches (for a circular rod).

The convergence of solutions for real branches was studied only in the case of three branches that exist below the first cutoff—longitudinal, torsional, and flexural modes (Fig. 6). Their analytically obtained values correspond to $n=0$ (longitudinal and torsional) and $n=1$ (bending) azimuthal order. The torsional mode (the lowest dataset) is almost exact and with practically a zero convergence rate, since it does not improve with mesh refinement. This mode is interesting in that its value is the same regardless of whether it was obtained by the strength of materials or elasticity, and in that the FE result is essentially exact. In hindsight, this is not surprising; torsion in a circular cylinder has constant strain, and is thus fully compatible with the constant strain elements assumed here. The longitudinal mode is more accurate than the bending mode. Its data points lie more closely to the straight line, as might be expected given that this mode has the wave number that is almost independent of the cross-section shape, so the node-shifting issues are less important.

The bending mode here consists of two degenerate branches—lateral bending and vertical bending—that coincide due to rotational symmetry of the cross section. For the coarsest mesh, those two modes have slightly different values at a given ω, but the difference between them diminishes with mesh refinement. The lower of the two is used in the convergence study. The bending mode is converging slightly more slowly than the longitudinal mode, with $s=-1.76$, as opposed to $s=-1.9$ for the longitudinal mode.

As shown in Fig. 7, the rates of convergence of the lowest two imaginary branches (corresponding to the purely imaginary solution of the $n=1$ case in the Pochhammer–Chree equations) are close to those of the real branches. Their accuracy is somewhat lower (their curves lying higher along the vertical axis) than that of the real branches and the accuracy of higher imaginary branch is fairly low, which is expected for evanescent modes.

Note that the analytic branches are grouped according to the azimuthal order *n*, while the FE code sorts the solutions in the ascending order of absolute value (in the case of

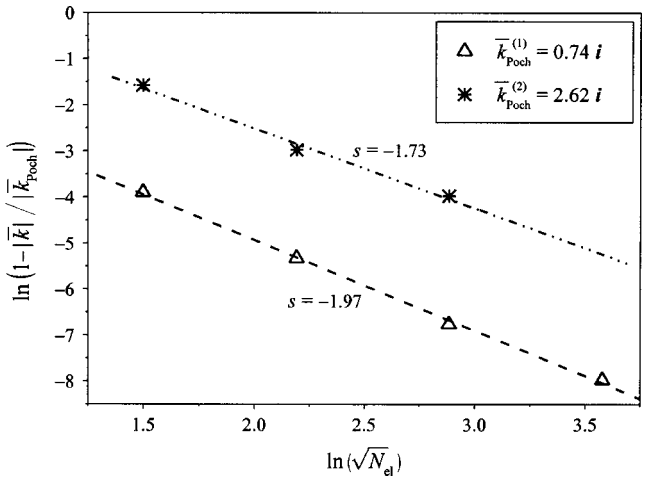


FIG. 7. Numerical convergence and convergence rates of selected imaginary branches (for a circular rod).

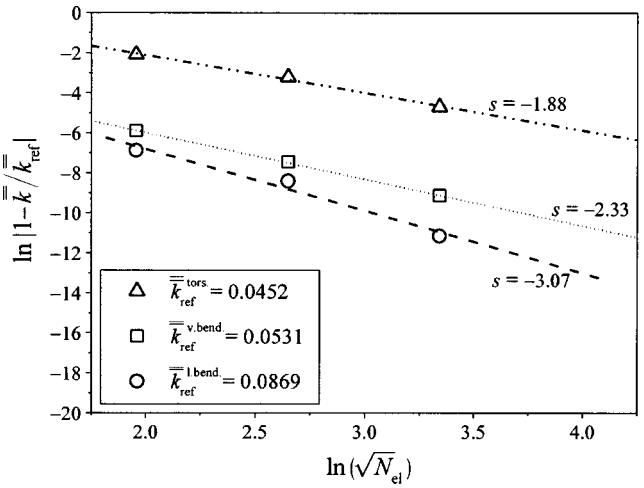


FIG. 9. Convergence and rates of change of selected real branches with respect to mesh refinement (for rail).

purely real or imaginary solutions) or modulus (in the case of complex solutions) with no reference to n. There is no consistent method of correlating the points on the numerically and analytically obtained branches, and for very coarse meshes or higher-order branches, it becomes increasingly difficult to recognize the corresponding roots—thus the missing fourth point on the convergence curve of the higher mode. However, this is not of much concern for the current study, because the frequencies of interest are very low, and the solutions for branches are consequently very accurate.

In order to locate the complex branches easily, the lowest three solutions from the code were used as starting guesses for the Newton–Raphson method of root searching of the Pochhammer–Chree equations, trying different values of azimuthal order n. The two lowest complex solutions were found to correspond to $n=2$ and $n=0$, respectively. The accuracy of those solutions is shown in Fig. 8 and is good, but lower than for the real branches and the lowest-order imaginary branch. The higher complex mode is of slightly higher accuracy than the higher imaginary mode—it is not clear whether this difference can be explained by the magnitude of the imaginary part, which is smaller for the complex branch.

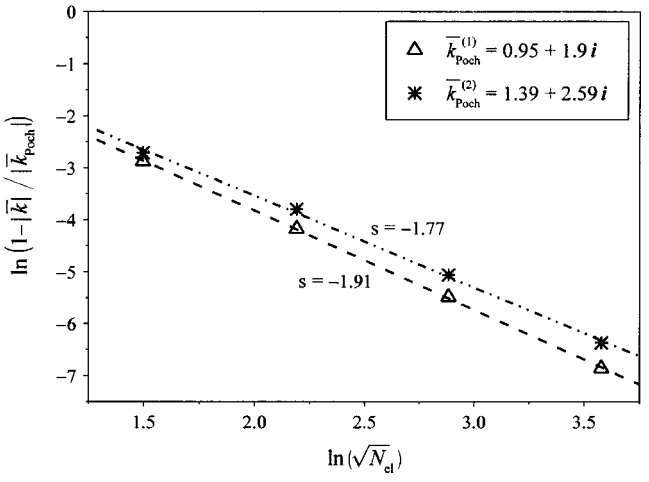


FIG. 8. Numerical convergence and convergence rates of selected complex branches (for a circular rod).

In summary, the FE solutions examined here are in good agreement with the analytic solutions of the Pochhammer–Chree equation. High accuracy was generally achieved for the lowest branches below the first cutoff; good accuracies were obtained for mesh #3. Most notably, for the bending mode that is of our particular interest, the accuracy (as judged against Pochhammer–Chree solution) for mesh #3 is excellent, 0.03%; convergence rates are generally close to asymptotic values, with $s=(-1.73)$ to (-1.97), and $s=-1.76$ for the flexural mode.

B. Complex cross section

The standard rail 136 lb AREMA is an example of a waveguide with irregular, complex cross section with one (vertical) axis of symmetry confining the bending modes to two perpendicular preferred planes. The accuracy of the FE solution for the lowest propagating and evanescent modes in such a waveguide was studied for the frequency $\overline{\omega}=0.007\,141$ rad/in. ($f=233$ Hz). The resulting convergence curves are shown in Figs. 9–11. The radius of gyration $r_g=0.979$ in. was chosen as the characteristic length analogous to the radius a of the Pochhammer–Chree theory. The corresponding normalized wave number that is used throughout this section, $\overline{\overline{k}}=kr_g$, is nondimensional.

The convergence curves are plotted analogously to those for the circular waveguide, but lacking the analytic solution for comparison, the values corresponding to the finest, fourth mesh, are used as a reference (denoted by subscript ''ref''). Therefore, the trends seen in those graphs do not represent convergence in the strict sense. The slope s of such a convergence curve is not equal to the convergence rate, but rather to the rate of change of the relative error with refinement of a particular mesh to the finest mesh #4. The magnitude of error for a particular mesh is analogous to accuracy in the case of a circular waveguide and, therefore, the vertical position of a convergence curve in such a plot can be viewed as an indicator of accuracy. In the following paragraphs, the words ''accuracy'' and ''convergence'' are used

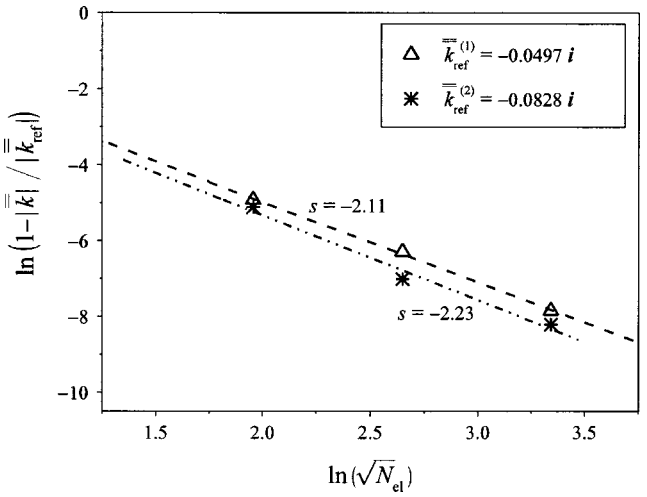


FIG. 10. Convergence and rates of change of selected imaginary branches with respect to mesh refinement (for rail).

in a colloquial manner, keeping in mind the distinctions explained above. That all the data points for all the studied modes lie near straight lines, is the indicator that the choice of mesh #4 as a reference is feasible.

The torsional (real, propagating) mode has the slowest rate of change of error, with the smallest s among four real modes, and with the lowest accuracy (Fig. 9). The accuracy of the axial (extensional) mode (not shown) is high, with a relative error of 0.2%, which did not change noticeably with refinement.

In the case of our most important mode—the lowest lateral bending—even the refinement from mesh #2 to mesh #3 changed the result by only 0.02%—a precision more than sufficient for the purpose of this project. The relatively large value (3.07) for s in this data, and the irregular fit to a straight line, are indications that the asymptotic behavior has not yet been reached. Nevertheless, one concludes with good (if not full) confidence that the small difference between the results for meshes #3 and #4 implies that $k_{\text{lat. bend.}}$ is accurately estimated. The fact that the Euler–Bernouilli (E–B) beam theory prediction for the lateral bending wave number ($\bar{\bar{k}}_{E-B}=0.0836$) is in error by about 3.5% at this frequency,

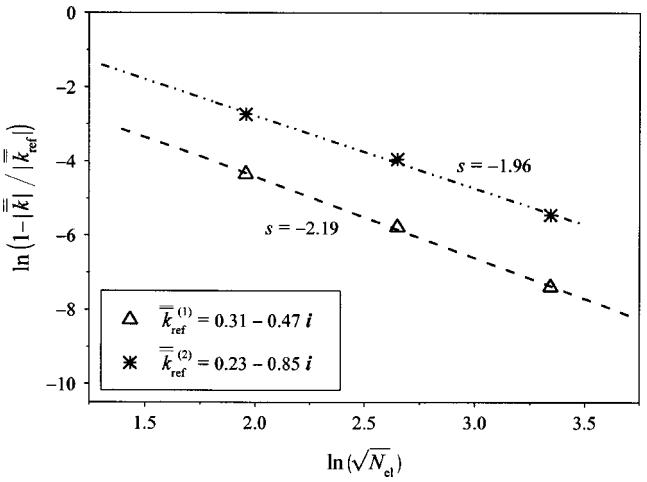


FIG. 11. Convergence and rates of change of selected complex branches with respect to mesh refinement (for rail).

confirms the necessity for the use of FEM, but also confirms the utility of E–B theory for rough estimates.

In summary, the accuracy and convergence of the FE code presented here are satisfactory, as proven by comparison with the analytic solutions for a circular guide. The generally good convergence in the case of a rail, especially for the propagating lateral bending mode, renders mesh #3 highly accurate, with perhaps some caveats in regard to the torsional mode. Further mesh refinement is increasingly impractical, because of the considerable computing time that the currently used EVP solver requires for a finer mesh.

VI. CONCLUSIONS

We have developed a code for the calculation of wave propagation in an elastic waveguide of arbitrary cross section. It is capable of accurately determining the real, imaginary, and complex branches at moderate values of dimensionless frequency and wave number, as demonstrated by a comparison with known solutions in the circular waveguide, and as demonstrated by convergence studies in a waveguide of a more complicated cross section. The emphasis on evanescent waves, and the emphasis on accuracy, make this work relevant for applications in the nondestructive evaluation of material or structural properties such as contained load and anisotropy. We anticipate applications of this method to waveguides with various common cross sections, such as railroad rails and rectangular waveguides, in particular to the nondestructive assessment of contained stress and anisotropy in rail. For a more cost-effective performance needed in most NDE applications, there is much room for improvement. Suggested modifications include the use of higher-order (quadratic) elements for a lower computational burden, the use of isoparametric elements with curved boundaries for lower error of the geometric approximation, and a more efficient algorithm for eigenvalue extraction.

ACKNOWLEDGMENTS

This work was supported by the Technical Scanning Committee of the Association of American Railroads and the Transportation Research Board's High Speed Rail IDEA program. The National Computational Science Alliance provided resources on its Silicon Graphics Origin2000.

APPENDIX: FINITE ELEMENT FORMULATION

The isoparametric mapping for a three-noded triangle, with the same shape functions both for coordinates and the displacements,[22] is

$$x(\xi,\eta)=\sum_{i=1}^{3} x_i N_i(\xi,\eta); \quad y(\xi,\eta)=\sum_{i=1}^{3} y_i N_i(\xi,\eta), \tag{A1}$$

where x_i and y_i are the nodal coordinates of the element. The above expressions specify x and y coordinates of any point in the triangle in terms of the parameters ξ and η, such that $0\leqslant\eta\leqslant1$ and $0\leqslant\xi\leqslant1-\eta$. Similarly, the displacements at any point (ξ,η) are

$$u(\xi,\eta)=\sum_{i=1}^{3} u_i N_i(\xi,\eta);$$

$$v(\xi,\eta)=\sum_{i=1}^{3} v_i N_i(\xi,\eta); \quad w(\xi,\eta)=\sum_{i=1}^{3} w_i N_i(\xi,\eta), \tag{A2}$$

where u_i, v_i, and w_i are the nodal displacements. The shape functions are

$$N_1=1-\xi-\eta; \quad N_2=\xi; \quad N_3=\eta. \tag{A3}$$

Then the approximate displacement field within the element is

$$\{\tilde{u}^{\mathrm{e}}(\xi,\eta)\}=[N(\xi,\eta)]\{d^{\mathrm{e}}\}, \tag{A4}$$

where

$$\{d^{\mathrm{e}}\}=\{u_1 \;\; v_1 \;\; w_1 \;\; u_2 \;\; v_2 \;\; w_2 \;\; u_3 \;\; v_3 \;\; w_3\}^T, \tag{A5}$$

$$[N(\xi,\eta)]=\begin{bmatrix} N_1 & 0 & 0 & N_2 & 0 & 0 & N_3 & 0 & 0 \\ 0 & N_1 & 0 & 0 & N_2 & 0 & 0 & N_3 & 0 \\ 0 & 0 & N_1 & 0 & 0 & N_2 & 0 & 0 & N_3 \end{bmatrix}. \tag{A6}$$

In addition, the displacement gradients are approximated with

$$\partial\{\tilde{u}^{\mathrm{e}}\}=[D]\{\tilde{u}^{\mathrm{e}}\}=[D][N]\{d\}, \tag{A7}$$

where $[D]$ is merely the operator matrix:

$$[D]=\begin{bmatrix} \partial/\partial x & 0 & 0 \\ \partial/\partial y & 0 & 0 \\ 0 & \partial/\partial x & 0 \\ 0 & \partial/\partial y & 0 \\ 0 & 0 & \partial/\partial x \\ 0 & 0 & \partial/\partial y \end{bmatrix}. \tag{A8}$$

It operates on $[N]$, resulting in

$$[B]=[D][N]=\begin{bmatrix} [J]^{-1} & & \\ & [J]^{-1} & \\ & & [J]^{-1} \end{bmatrix}\begin{bmatrix} N_{1,\xi} & 0 & 0 & N_{2,\xi} & 0 & 0 & N_{3,\xi} & 0 & 0 \\ N_{1,\eta} & 0 & 0 & N_{2,\eta} & 0 & 0 & N_{3,\eta} & 0 & 0 \\ 0 & N_{1,\xi} & 0 & 0 & N_{2,\xi} & 0 & 0 & N_{3,\xi} & 0 \\ 0 & N_{1,\eta} & 0 & 0 & N_{2,\eta} & 0 & 0 & N_{3,\eta} & 0 \\ 0 & 0 & N_{1,\xi} & 0 & 0 & N_{2,\xi} & 0 & 0 & N_{3,\xi} \\ 0 & 0 & N_{1,\eta} & 0 & 0 & N_{2,\eta} & 0 & 0 & N_{3,\eta} \end{bmatrix}, \tag{A9}$$

where $[J]^{-1}$ is the inverse of the Jacobian: $[J]=\partial(x,y)/\partial(\xi,\eta)$. Noting that $\mathrm{d}\Omega^{\mathrm{e}}=\mathrm{d}x\,\mathrm{d}y=\det[J]\mathrm{d}\xi\,\mathrm{d}\eta$, we finally arrive at the discretized form of Eq. (3)—the Lagrangian for a single element:

$$L^{\mathrm{e}}=\frac{1}{2}\{d^{\mathrm{e}}(z,t)\}^T\left[\int_0^1\int_0^{1-\eta}[\vec{\partial}_t[N]^T[\mathfrak{M}][N]\vec{\partial}_t-\vec{\partial}_z[N]^T[\kappa_2][N]\vec{\partial}_z-[B]^T[\kappa_1][N]\vec{\partial}_z-[B]^T[\kappa_0][B]]\det[J]\mathrm{d}\xi\,\mathrm{d}\eta\right]$$
$$\times\{d^{\mathrm{e}}(z,t)\}\mathrm{d}z. \tag{A10}$$

[1] J. Miklowitz, *Elastic Waves and Waveguides*, Series in Applied Mathematics and Mechanics (North-Holland, New York, 1978), Vol. 22.
[2] Y. H. Pao and R. D. Mindlin, "The dispersion of flexural waves in an elastic circular cylinder," J. Appl. Mech. **27**, 513–520 (1960).
[3] Y. H. Pao, "The dispersion of flexural waves in an elastic circular cylinder, Part II," J. Appl. Mech. **29**, 61–64 (1962).
[4] M. A. Onoe, H. D. McNiven, and R. D. Mindlin, "Dispersion of axially symmetric waves in elastic rods," J. Appl. Mech. **29**, 729–734 (1962).
[5] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).
[6] T. R. Meeker and A. H. Meitzler, "Guided wave propagation in elongated cylinders and plates," *Physical Acoustics* (Academic, New York, 1964), Vol. 1, Sec. A.2.
[7] G. Rosenfeld and J. B. Keller, "Wave propagation in elastic rods of arbitrary cross section," J. Acoust. Soc. Am. **55**, 555–561 (1974).
[8] G. Rosenfeld and J. B. Keller, "Wave propagation in non-uniform elastic rods," J. Acoust. Soc. Am. **57**, 1094–1096 (1975).
[9] L. Gavrić, "Computation of propagative waves in free rail using a finite element technique," J. Sound Vib. **185**, 531–543 (1995).
[10] B. Aalami, "Waves in prismatic guides of arbitrary cross section," J. Appl. Mech. **40**, 1067–1072 (1973).
[11] S. Widdehammar, P. A. Gradin, and B. Lundberg, "Approximate determination of dispersion relations and displacement fields associated with elastic waves in bars," J. Sound Vib. **246**, 853–876 (2001).
[12] T. Hayashi, W.-J. Song, and J. L. Rose, "Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example," Ultrasonics **41**, 175–183 (2003).
[13] M. R. Pfaffinger and J. Dual, "Higher vibration modes in railway tracks at their cut-off frequencies," J. Sound Vib. **235**, 1015–1037 (2002).
[14] H. S. Paul and M. Venkatesan, "Propagation in a piezoelectric solid cylinder of arbitrary cross section," J. Acoust. Soc. Am. **82**, 2013–2020 (1987).

[15] K. H. Huang and S. B. Dong, ''Propagating waves and edge vibrations in anisotropic composite cylinders,'' J. Sound Vib. **96**, 363–379 (1984).
[16] V. V. Volovoi, D. H. Hodges, V. L. Berdichevsky, and V. G. Sutyrin, ''Dynamic dispersion curves for non-homogeneous, anisotropic beams with cross-section of arbitrary geometry,'' J. Sound Vib. **215**, 1101–1120 (1998).
[17] V. Damljanović and R. L. Weaver, ''Elastic waves in cylindrical waveguides of arbitrary cross section,'' TAM Report, Department of Theoretical & Applied Mechanics, University of Illinois at Urbana—Champaign, No. 969, 2001.
[18] O. Onipede and S. B. Dong, ''Propagating waves and end modes in pretwisted beams,'' J. Sound Vib. **195**, 313–330 (1996).
[19] H. Taweel, S. B. Dong, and M. Kazic, ''Wave reflection from the free end of the cylinder with an arbitrary cross-section,'' Int. J. Solids Struct. **37**, 1701–1726 (2000).
[20] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[21] I. S. Sokolnikoff, *Mathematical Theory of Elasticity*, 2nd ed. (McGraw-Hill, New York, 1956).
[22] R. D. Cook, D. S. Malkus, and M. E. Plesha, *Concepts and Applications of Finite Element Analysis*, 3rd ed. (Wiley, New York, 1989).
[23] C. L. Dym and I. H. Shames, *Solid Mechanics: A Variational Approach*, Advanced Engineering Series (McGraw-Hill, New York, 1973).
[24] The complex roots of any polynomial with real coefficients always appear in pairs of complex conjugates.
[25] Left- and right-going waves are related; the solution set $(-k;u,v,-w)$ is the mirror image of the solution set $(k;u,v,w)$.
[26] L. Meirovitch, *Fundamentals of Vibrations* (McGraw-Hill, Boston, 2001).
[27] AREMA 136 lbs/yard rail standard has a 6 in. base and a 7 5/16 in. height; its detailed specification can be found at L. B. Foster Company's website: www.lbfoster.com

Forced response of a cylindrical waveguide with simulation of the wavenumber extraction problem

Vesna Damljanović and Richard L. Weaver[a)]
Department of Theoretical & Applied Mechanics, University of Illinois at Urbana—Champaign, Urbana, Illinois, 61801-2983

(Received 25 August 2003; revised 14 January 2004; accepted 25 January 2004)

The steady state response of a cylindrical elastic waveguide of arbitrary cross section to a harmonic load is considered. The inverse problem of wavenumber extraction is simulated using finite element discretization of the cross section. The case of a concentrated lateral point load on a railroad rail is used for illustration, at a frequency of 202 Hz, corresponding to a frequency well below the first cutoff, but above the regime where simple strength of materials concepts are accurate. The work is conceived with a view towards applications in the nondestructive characterization of such beams by means of scanned laser vibrometry for measurement of lateral bending wavenumbers dependent on a contained static axial load. Simulations of such measurements are generated, and subjected to noise and to a variety of potential systematic errors. Linear and nonlinear least squares minimization of the residual between simulated measurements and fits show that the lateral bending wavenumber can be recovered accurately, with remarkable robustness, even in the presence of noise and systematic errors. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1675818]

PACS numbers: 43.40.Cw, 43.20.Mv, 43.40.Le [ANN] Pages: 1582–1591

I. INTRODUCTION

Cylindrical elastic waveguides of arbitrary cross section have been receiving increased attention in recent years. This work has been driven in part by significant potential applications to guided vibrations in a railroad rail. Gavric,[1] Hayashi *et al.*,[2] Damljanović and Weaver,[3] and Damljanović[4] have discussed Finite Element methods for approximate solutions for the propagating guided waves. Pfaffinger and Dual[5] discussed the cutoff frequencies. Relevant other work includes that of Aalami[6] and Widdehammer *et al.*,[7] Huang and Dong[8] and Volovoi *et al.*[9] Damljanović and Weaver addressed evanescent waves as well as propagating waves and focused attention on high accuracy. In all this work there has been little or no attention directed towards the problem of responses of such structures to specified forcing; with the exception of Hayashi *et al.*[2] the published work has emphasized only the free vibration of propagating waves. In this paper we formulate the solution for steady state responses to specified forcing on cylindrical elastic waveguides of an arbitrary cross section, including both propagating and evanescent waves, and apply it to simulate an inverse problem in which measured steady state responses are used to infer guided wave wavenumbers, thus characterizing the waveguide.

In this work we are motivated by an application to a railroad rail, in which temperature extremes, and constrained expansions, lead to large contained loads that in turn lead to bucklings and rail breaks. The nondestructive determination of such a contained load in rails has long been of concern to the industry. We note that the bending rigidity of a rail is relatively sensitive to a contained load (indeed others have investigated using such sensitivity to determine load from resonant frequencies[10]) and we have proposed measuring bending wavenumbers k at a specified frequency in order to infer such a load.[4] This sensitivity is greatest at long wavelengths, which leads, given a finite span over which to make measurements, to the interesting problem of determining the wavenumber from measurements over a span of only about one wavelength. Spatial Fourier transforms—specifically their peaks—can be used to identify wavenumbers, but only if the spatial data range covers many wavelengths; clearly such a technique is not going to be effective here. The problem is further complicated by the presence, in a steady-state vibration of a waveguide, of other modes of vibration, not just the bending waves. In the next section we extend FEM formulations for propagating free vibrations in such waveguides to the related problem of steady state responses to forcing on such waveguides, and derive an expression for the response to any such forcing. This expression is then applied to the special case of lateral responses to a concentrated lateral force. It is found to consist of a sum of many waves, both propagating and evanescent, most of which are unimportant at points not close to the source, thus allowing the sum to be truncated except at the source.

In Secs. III and IV we attempt to extract the lateral bending wavenumber from measured data. A short span of the data, corresponding to about one wavelength, is treated as measured data and fit to a linear superposition of waves with specified wavenumbers. By adjusting the wavenumber of the lateral bending wave in this fit, we determine a best estimate for that wavenumber. Of course if the data is perfect, we recover the actual wavenumber used in the original data construction. However, by adding simulations of noise, or other laboratory imperfections, we investigate the robust-

a)Author to whom the correspondence should be addressed; electronic mail: r-weaver@uiuc.edu

 0001-4966/2004/115(4)/1582/10/$20.00

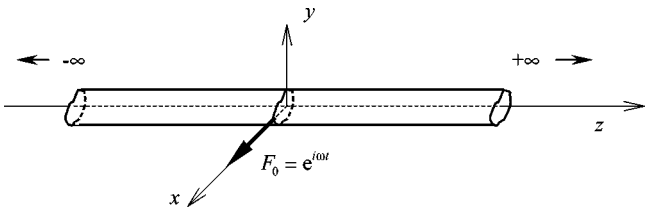


FIG. 1. Cylindrical waveguide with constant (arbitrary) cross-section and harmonic lateral point-force at $zi=i0$.

ness of the proposed recovery of the lateral bending wavenumber.

Finally, in Sec. V we summarize the conclusions.

II. SOLUTION TO FORCED VIBRATION PROBLEM

Consider an isotropic elastic cylindrical waveguide with a traction-free lateral boundary and constant cross-section along the z axis of a rectangular Cartesian coordinate system. The waveguide is excited laterally by a point-force of unit magnitude along the x axis (Fig. 1). The semi-analytical discretization of Katorovich-Krilov[11,12] is applied to the admissible displacement field, taking it to be of a finite-element-like form given in terms of Cartesian nodal displacements $d_i(z,t)$, $(i=1,2,...,3N$, where N is the number of nodes) by using linear constant strain isoparametric triangular elements.[13] The details are given elsewhere.[4] The procedure is similar to that described in other studies.[1,2,8,9,11] Other types of elements could also be chosen; the form of the equations to be discussed here would remain the same.

The lineal kinetic energy density and strain energy density consequent to the chosen form for the admissible displacement field then determine a Rayleigh–Ritz Lagrangean lineal density $\mathcal{L}$ and a corresponding Euler Lagrange equation of motion:[3,14]

$$\frac{\partial \mathcal{L}}{\partial d_i} - \frac{\mathrm{d}}{\mathrm{d}z}\frac{\partial \mathcal{L}}{\partial d_i'} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{d}_i} = Q_i(z,t), \tag{1}$$

where $Q_i(z,t)$ is the generalized force associated with the degree of freedom $d_i(z,t)$ and may be obtained from the actual force distribution by using standard considerations of virtual work.[14–16] One deduces

$$(\mathrm{sym}[K_2])_{ij} d_j'' + (\mathrm{skw}[K_1])_{ij} d_j' - (\mathrm{sym}[K_0])_{ij} d_j - (\mathrm{sym}[M])_{ij} \ddot{d}_j = -q_i(z) e^{i\omega t}, \tag{2}$$

where $[M]$, $[K_2]$, $[K_1]$ and $[K_0]$ are global (''mass'' and ''stiffness'') matrices of order $3N\times 3N$. The prime and double prime denote the first and second derivatives with respect to z, respectively, and the double dot denotes the second temporal derivative. If the forcing is confined to the nodes, then $q_i(z)$ is a Cartesian component of the amplitude of the harmonic forcing at a node. Otherwise, $q_i(z)$ is a linear combination of the continuously distributed forcing on the neighboring elements.

The steady-state response of the waveguide is taken in the following form:

$$d_j(z,t) = h_j(z) e^{i\omega t}, \quad j=1,...,3N \ \text{ or } \ \{d(z,t)\} = \{h(z)\} e^{\mathbf{i}\omega t}. \tag{3}$$

The nonhomogeneous problem statement is obtained by substituting Eq. (3) into the governing equation (2) and changing back to matrix notation:

$$[[K_2]\partial_z^2 + [K_1]\partial_z - [K_0] + \omega^2[M]]\{h(z)\} = -\{q(z)\}. \tag{4}$$

A $3N\times 3N$ unitary transformation matrix $[T]$ is introduced next, with each diagonal element equal to 1, except every third (corresponding to displacement in the z direction) being equal to $\mathbf{i}$. The matrix $[T]$ has the property $[T]^*[T] = [T][T]^* = [I]$, where * denotes the complex conjugate. Equation (4) is multiplied from the left by $[T]$ and $\{h\}$ is replaced by $[T]^*[T]\{h\} = \{h\}$. The symmetric matrices $[K_2]$, $[K_0]$ and $[M]$ do not mix the w component of displacement with u and v, so $[T][K_2][T]^* = [K_2]$, $[T][K_0][T]^* = [K_0]$ and $[T][M][T]^* = [M]$. On the other hand, the skew matrix $[K_1]$ mixes u and v with w but does not mix u and v displacement components with each other, so $[T][K_1][T]^* = -\mathbf{i}[\hat{K}_1]$, where $[\hat{K}_1]$ is real and symmetric. Equation (4) then reads as

$$[[K_2]\partial_z^2 - \mathbf{i}[\hat{K}_1]\partial_z - [K_0] + \omega^2[M]]\{\hbar(z)\} = -[T]\{q(z)\}, \tag{5}$$

where $\{\hbar\} = [T]\{h\}$.

In a manner that bears similarity to the state-space reformulation of dynamics, this equation is then rewritten in the form consisting of a generalized EVP on the left-hand side and the generalized forcing term on the right-hand side. A new vector is introduced:

$$\{\xi\} = -\mathbf{i}\partial_z\{\hbar\} \tag{6}$$

and multiplied by the matrix $[K_2]$:

$$[K_2]\{\xi\} = -\mathbf{i}\partial_z[K_2]\{\hbar\}. \tag{7}$$

Due to $[K_2]$ being nonsingular, the above multiplication incurs no loss of information. Combining Eqs. (5) and (7) one obtains the following governing equation:

$$[A]\{p(z)\} - \mathbf{i}[B]\partial_z\{p(z)\} = \{\mathcal{F}(z)\}, \tag{8}$$

with

$$[A] = \begin{bmatrix} [K_2] & [0] \\ [0] & \omega^2[M] - [K_0] \end{bmatrix}, \quad [B] = \begin{bmatrix} [0] & [K_2] \\ [K_2] & [\hat{K}_1] \end{bmatrix},$$
$$\{p(z)\} = \begin{Bmatrix} \mathbf{i}[T]\partial_z\{h(z)\} \\ [T]\{h(z)\} \end{Bmatrix}, \quad \{\mathcal{F}(z)\} = \begin{Bmatrix} \{0\} \\ -[T]\{q(z)\} \end{Bmatrix}. \tag{9}$$

Both $6N\times 6N$ matrices are real and symmetric, although not positive-definite. We take the forcing to be concentrated at one node, P_f, at $z=0$, and to be in one direction. The vector $\{q(z)\}$ is, in indicial notation,

$$q_j(z) = \begin{cases} F_0\delta(z), & \text{if } j = j_{P_f} \\ 0, & \text{if } j \neq j_{P_f} \end{cases}, \tag{10}$$

where F_0 is the amplitude of forcing and j_{P_f} is the index related to the appropriate d.o.f. at the point P_f where the forcing is applied. If $i_{\text{d.o.f.}}$ is the index of a particular d.o.f. (1 for u, 2 for v and 3 for w), and the forcing is at a point that corresponds to node n_f, then the correct forcing index is

$j_{P_f}=3(n_f-1)+i_{\text{d.o.f.}}$. In case of force being applied over multiple points, $q_j(z)$ would be nonzero for all j's corresponding to the nodes where the force is applied.

We solve Eq. (8) by using the spatial Fourier transform (FT) defined[17] as

$$\{\overline{a}(k)\}=\int_{-\infty}^{+\infty}\{a(z)\}e^{ikz}\,dz, \tag{11}$$

where $\{a(z)\}$ is a vector function of position z and k is the wavenumber. The spatial FT of Eq. (8) yields the governing equation in the k domain:

$$([A]-k[B])\{\overline{p}(k)\}=[\overline{\mathcal{F}}(k)]=\{F_0\}, \tag{12}$$

where

$$\{F_0\}_j=\begin{cases}F_0, & \text{if } j=3N+j_{P_f}\\ 0, & \text{if } j\neq 3N+j_{P_r}\end{cases}. \tag{13}$$

The homogeneous part of Eq. (12) is exactly the eigenvalue problem discussed previously:[1–3]

$$[A]\{\psi\}=k[B]\{\psi\}, \tag{14}$$

whose solutions are $6N$ eigenvalue–eigenvector pairs k_r and $\{\psi_r\}$, with $r=1,...,6N$. The symmetry of $[A]$ and $[B]$ allows the conclusion that the eigenvectors have the desired orthogonality: $\{\psi_r\}^T[B]\{\psi_s\}=0$, unless $k_r=k_s$. The accuracy and convergence of our FE solution were proven satisfactory, but their detailed study[4] is outside the scope of this paper.

The solution $\{\overline{p}(k)\}$ of Eq. (12) can be written as an expansion in terms of eigenvectors:[18]

$$\{\overline{p}(k)\}=\sum_{r=1}^{6N}\frac{1}{k_r-k}\frac{\{\psi_r\}^T\{F_0\}}{\{\psi_r\}^T[B]\{\psi_r\}}\{\psi_r\}. \tag{15}$$

The solution in the z domain is obtained by taking the inverse FT:

$$\{p(z)\}=\frac{1}{2\pi}\sum_{r=1}^{6N}\int_{-\infty}^{+\infty}\frac{1}{k_r-k}\frac{(\{\psi_r\}^T\{F_0\})\{\psi_r\}}{\{\psi_r\}^T[B]\{\psi_r\}}e^{-ikz}\,dk. \tag{16}$$

The above integral has simple poles k_r; it was solved with aid of the Cauchy residue theorem.[19] Closing the integration contour in the lower half-plane (valid for $z>0$) and taking residues from all the positive real[20] poles, negative imaginary poles and all complex poles with negative imaginary parts (a total of $3N$ poles), one obtains the solution as the sum of the residues of these poles. The resulting state-space vector is

$$\{p(z)\}=-\mathbf{i}\sum_{r=1}^{3N}\frac{(\{\psi_r\}^T\{F_0\})\{\psi_r\}}{\{\psi_r\}^T[B]\{\psi_r\}}e^{-ik_rz}, \tag{17}$$

where the sum is only over the above-mentioned poles. The lower half of that vector is the spatial function of the steady-state response:

$$([T]\{h(z)\})_j=\left(-\mathbf{i}\sum_{r=1}^{3N}\frac{(\{\psi_r\}^T\{F_0\})\{\psi_r\}}{\{\psi_r\}^T[B]\{\psi_r\}}e^{-ik_rz}\right)_{3N+j}. \tag{18}$$

The elements of $[T]$ associated with the lateral displacements are unity, so if interest is confined to u or v, then Eq. (18) is equal to $h_j(z)$ itself. Finally, the lateral response itself is obtained by substituting the above function into Eq. (3):

$$d_j(z,t)=\left(-\mathbf{i}\sum_{r=1}^{3N}\frac{(\{\psi_r\}^T\{F_0\})\{\psi_r\}}{\{\psi_r\}^T[B]\{\psi_r\}}e^{\mathbf{i}\omega t-ik_rz}\right)_{3N+j},\quad j=1,...,3N. \tag{19}$$

These solutions are complex numbers and their real parts are the corresponding true displacements. Again, the index r runs over only the $3N$ eigenvalues k_r that are positive real, negative imaginary, or complex with negative imaginary parts.

For the case of constant lateral forcing with amplitude $F_0=1$ at P_f, the complex amplitude of the steady-state harmonic response at a point P is obtained by dropping the time dependence:

$$d_{j_P}(z)=-\mathbf{i}\sum_{r=1}^{3N}\frac{\{\psi_r\}_{j_{P_f}}\{\psi_r\}_{j_P}}{\{\psi_r\}^T[B]\{\psi_r\}}e^{-\mathbf{i}k_rz}, \tag{20}$$

where $\{\psi_r\}_{j_{P_f}}$ and $\{\psi_r\}_{j_P}$ are the $j_{P_f}^{\text{th}}$ and j_Pth components of eigenvector $\{\psi_r\}$.

This form for the response to a concentrated forcing indicates that the response is a superposition of waves propagating, or evanescing, away from the point of forcing. If the forcing were distributed over a finite region, the conclusion is essentially identical; at all points z outside this region the response consists of a sum of waves propagating, or evanescing, away from the region of forcing. Thus the form of the response is unaltered, only the coefficients of the factors e^{-ik_rz} are different. In the case of a rail with many distributed forces and constraints and supports, the conclusion is much the same. By representing the effect of such constraints and supports as additional distributed forcings $q(z)$, one finds that the response at any point z *not within any region of such forcing or support or constraint* must be a superposition of propagating and evanescing waves. There is one difference: If such forces lie both to the left and the right of the point of interest, then the superposition must include waves which propagate and evanesce in both directions, not just away from the origin.

For the purpose of illustration but also for convenience in simulations presented in the following section, the real (in-phase) and imaginary (out-of-phase) parts of expression (20) are evaluated for the waveguide with the rail cross-section and a coarse mesh (42 nodes, 50 elements; see Fig. 2), for 202 Hz (or $\omega=1271$ rad/s), which corresponds to a frequency well below the first cutoff, and a bending wavelength of 77 inches. The mesh fineness is of no importance for the study presented in this paper. The displacement is calculated at a point P (1.46 in.; 6.21 in.) that coincides with one of the nodes on the vertical edge of the rail head. The forcing is taken to be in the lateral direction at P_f (0.37 in.; 2.75 in.) and $z=0$. The real (in-phase) and imaginary (out-of-phase) parts of the response over 0–100 inches of the infinite waveguide are shown in Fig. 3.

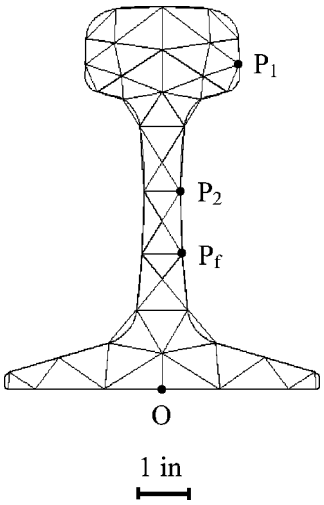


FIG. 2. Rail cross-section with the coarsest mesh (42 nodes and 50 elements). The origin O, the points P_1 and P_2 used in the simulated measurement and the point P_f used for the simulated forcing, are indicated. The y-axis points up from O, the x-axis points towards the right.

The apparent wavelength of the response is of almost the same magnitude as the wavelength of the lateral bending mode, showing that the dynamic lateral displacement is dominated by the contribution of the lateral bending mode, except near the forcing at $z=0$ where the deviations from perfect sine and cosine curves indicate small contributions from the torsional, evanescent lateral bending and other evanescent waves. It is interesting to note that the maximum displacement of an infinite steel rail to a unit forcing is only of the order of 10^{-6} inches per pound force.

III. INVERSE PROBLEM

Critical to the development of the rail stress measurement technique is a robust procedure for the extraction of $k_{\text{lat. bend.}}$[21] from the measurements of displacements along the z-axis of a railroad rail that is vibrated laterally. The extraction is not a trivial process if only a short range—less than one wavelength—of rail length is expected to be available for the analysis.[22] Furthermore, there are other modes present in the rail—vertical bending, torsional, extensional and many evanescent modes—that distort the distribution of vibration amplitude. The major steps in the procedure are described below.

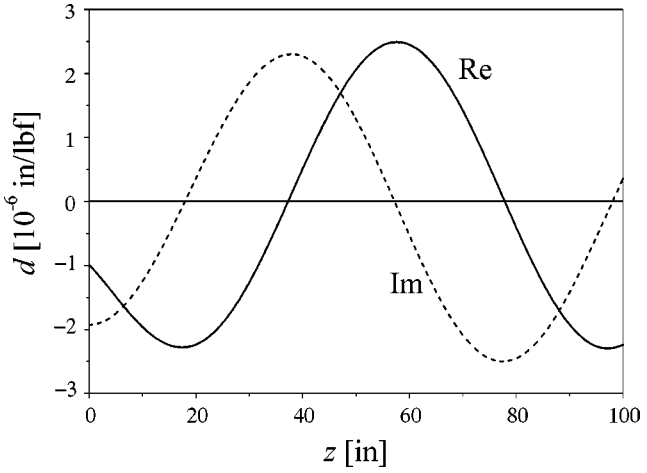


FIG. 3. Response of the uniform infinite rail to the unit harmonic point-force at $z=0$.

Here we imagine that measurements $d(z)$ of steady-state vibration amplitude, have been made, at a fixed point (x,y) on the cross-section, over a short span in z that is free of supports, constraints or forcing. As discussed following Eq. (20), $d(z)$ is necessarily of the form

$$d(z)=\sum_{n=1}^{\infty} A_n U_n(x,y;\omega)\mathrm{e}^{-\mathrm{i}k_n z}, \tag{21}$$

consisting of an infinite superposition of rightward and leftward propagating and evanescent waves. This is valid, with constant A's, over any span z without loads or constraints. Thus fasteners and rail ties are irrelevant. Here $U_n(x,y;\omega)$ is the nth guided mode shape at frequency ω and k_n is the nth wavenumber at frequency ω. The constant coefficients A_n depend on the forcing and supports and are in general difficult to predict. The wavenumbers k_n (except for the lateral bending wavenumber) and mode shapes U_n are taken to be well known, as determined by a high precision FEM code.[3,4] The lateral bending wavenumber depends to a significant degree on the unknown contained load, and is therefore uncertain.[21]

The inverse problem to be solved consists of the determination of $k_{\text{lat. bend.}}$ (and the coefficients A_n) that best fit Eq. (21) to the measured $d(z)$. The difference between the best-fit $k_{\text{lat. bend.}}$ and that expected at zero load is the quantitative indicator of contained load. In practice it suffices to truncate the infinite sum, and to use approximate, albeit highly accurate, FEM values for the U_n and k_n. Thus we truncate (20) and write

$$d(z)\approx d(z)=\sum_{m=1}^{N_r} \alpha_m \upsilon_m(x,y;\omega)\mathrm{e}^{-\mathrm{i}k_m z}, \tag{22}$$

where α_m, $\upsilon_m(x,y;\omega)$ and k_m are analogous to A_n, $U_n(x,y;\omega)$ and k_n (although obtained using a finite mesh) and index m sweeps over the indices of the N_r more influential guided modes. At sufficiently low frequency there are but 4 propagating waves (real k) in each direction. At a z position sufficiently far removed from the loads or supports only a small number of the evanescent (imaginary or complex k) terms needs to be included. Because the excitation is in the lateral direction, one expects the lateral bending mode to dominate the sum.

The fit procedure involves the minimization of the mean-square residual (χ^2) between the measurement and the sum in Eq. (22) with respect to variations in the wavenumber $k_{\text{lat. bend.}}$ (with all other wavenumbers fixed) and the coefficients α_m. The best-fit $k_{\text{lat. bend.}}$ is thus obtained. The coefficients α_m obtained in the process are unimportant for the projected application.

Before this procedure is used in practice, a careful scrutiny of the $k_{\text{lat. bend.}}$ extraction procedure must be employed in order to estimate the conditions that would have to be satisfied in practice to provide a desired precision for the best-fit $k_{\text{lat. bend.}}$. The envisioned application has asked for $k_{\text{lat. bend.}}$ to

be determined within about 0.25% at 202 Hz, using data over a range in z less than one wavelength. The following section describes some of the more important issues of the $k_{\text{lat. bend.}}$ extraction procedure.

IV. SIMULATIONS OF EXTRACTION OF LATERAL BENDING WAVENUMBER

The simulations were inspired by several questions, formulated as follows. What scan length and how many terms in the truncated series (22) would be sufficient for approximating $d(z)$, and over what range in z, as it pertains to the railroad rail? These questions are addressed in Sec. IV A, which deals with general quality of the fit. How well would the proposed method be able to discern the small perturbations in $k_{\text{lat. bend.}}$ due to longitudinal stress in the rail? In Sec. IV B we show sensitivity estimates obtained by simulations of distortion of the lateral bending wavenumber. What level of measurement error (due to noise and drift of electrical instrumentation, or vertical slant of scan path) would allow for the prescribed precision in the lateral bending wavenumber? This is addressed in Sec. IV C.

The simulations were completed using the FE solution for the displacement of point P (1.46 in.; 6.21 in.) on the rail surface, as shown in Sec. II. From the $6N$ eigenvalues obtained by solving the EVP at $\omega=1271$ rad/s, only the eigenvalues with an imaginary part in the range 0 to -2.0 in.$^{-1}$ (a total of N_r) were used to construct the curves in Fig. 3. This cutoff was judged satisfactory due to $e^{-\mathbf{i}k_m z}$ for the excluded waves decaying by a factor of at least e^2 per inch along the z axis, thus implying that such waves are negligible at distances more than couple of inches from the load. For the four meshes used in the FE study,[3,4] the number N_r ranged between 35 for the coarsest mesh and 39 for the finest. Note that N_r is just a fraction of $6N$ d.o.f. The region in immediate proximity to the source (at $z=0$) was ignored, being heavily influenced by the neglected evanescent waves. This data set of complex displacement amplitudes, evaluated at 981 points z_i $(i=21,1001)$ along the z axis in the scan range of $z \in [2\text{ in.}, 100\text{ in.}]$ in steps of 0.1 inch, was adopted as the "measurement" $d(z)$.

A. General fit quality

How well does the "measurement" described above fit the following heavily truncated sum of exponentials in k_m, the wavenumbers obtained by FE:

$$\hat{d}(z_i)=\sum_{m=1}^{\mathrm{M}=11}\alpha_m e^{-\mathbf{i}k_m z_i}, \tag{23}$$

with only 11 out of $6N$ terms? The 11 retained terms correspond to 4 purely real eigenvalues (the only real solutions for this frequency considerably below the first cutoff) and 7 complex and imaginary eigenvalues that have the smallest (by absolute value) imaginary part. The 11 wavenumbers are listed in Table I. Their mode types were identified by comparing with strength-of-materials solutions and by examination of their mode shapes.

TABLE I. Wavenumbers of the 11 most influential modes for the rail cross-section [for 202 Hz (or $\omega=1271$ rad/s)].

k (in.$^{-1}$)	Mode type
0.0814369−0.0**i**	Lateral bending (propagating)
0.0498041−0.0**i**	Vertical bending (propagating)
0.0348009−0.0**i**	Torsion
0.0−0.047203**i**	Vertical bending (evanescent)
0.00620008−0.0**i**	Longitudinal
0.0−0.078377**i**	Lateral bending (evanescent)
0.324088−0.46016**i**	Less known evanescent
−0.324088−0.46016**i**	"
0.0−0.441982**i**	"
0.0−0.217283**i**	"
0.0−0.184638**i**	"

A linear least-squares fitting procedure was used to fit the "data" into expression (23) by minimizing the regression function:[22]

$$\chi^2(\alpha_1,\ldots,\alpha_{11})=\sum_{i=21}^{1001}\left|d(z_i)-\hat{d}(z_i,\omega;\alpha_1,\ldots,\alpha_{11})\right|^2, \tag{24}$$

where $d(z)$ was taken from the truncated series (20), (22). Because the quantity in (24) is complex, the coefficients α_m are expected to be complex. The modulus squared is equal to the product of the expression in brackets and its complex conjugate. Substituting (23), results in

$$\chi^2=\sum_{i=21}^{1001}\left[d_i d_i^* - d_i\sum_{m=1}^{11}\alpha_m^* e^{\mathbf{i}k_m^* z_i} - d_i^*\sum_{m=1}^{11}\alpha_m e^{-\mathbf{i}k_m z_i} + \left(\sum_{m=1}^{11}\alpha_m e^{-\mathbf{i}k_m z_i}\right)\left(\sum_{m=1}^{11}\alpha_m^* e^{\mathbf{i}k_m^* z_i}\right)\right], \tag{25}$$

with complex conjugates denoted by * and $d(z_i)$ by d_i. Minimizing with respect to either α_m or α_m^* results in the same equations, so conjugates are chosen for convenience:

$$\frac{\partial\chi^2}{\partial\alpha_p^*}=0, \quad p=1,\ldots,11. \tag{26}$$

The resulting system of 11 simultaneous linear equations:

$$\sum_{m=1}^{11}\alpha_m\sum_{i=21}^{1001}e^{\mathbf{i}(k_p^*-k_m)z_i}=\sum_{i=21}^{1001}d(z_i)e^{\mathbf{i}k_p^* z_i}, \quad \forall p=1,\ldots,11, \tag{27}$$

is solved for the coefficients α_m and the fit (23) is compared with the "measurement." The difference is termed the residual.

As shown in the plots (Fig. 4), the real and imaginary parts of the residual are negligible at a sufficient distance from the origin (the location of the forcing). Similarly, the residual decays with distance from the $z=100$ inch end of the selected region. Both ends of the selected region are associated with a larger residual than in the interior, but for different reasons. At $z=0$ the reason is physical, with forcing introducing numerous evanescent waves that were not included in the truncated sum. At $z=100$ in., the high residual is an artifact of the fit that has to match an infinite sum to a

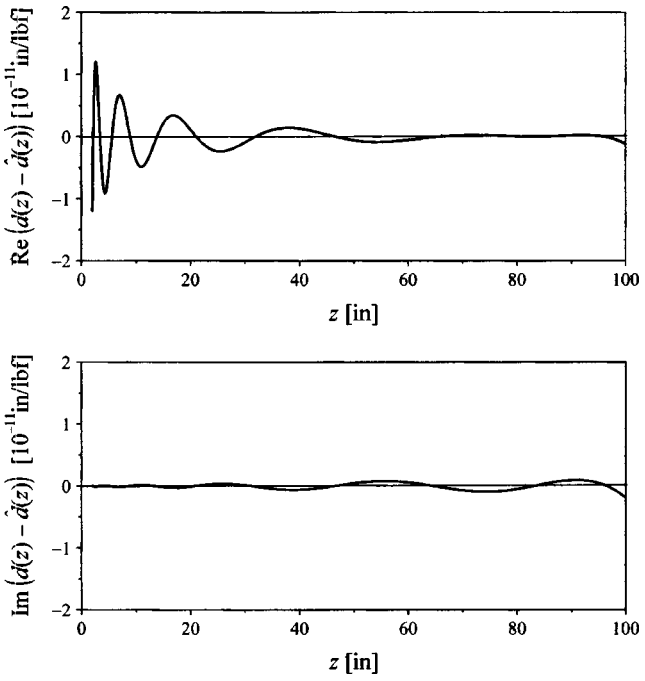


FIG. 4. The real (top) and imaginary (bottom) part of the residual between the displacement ''measurement'' and the fit.

finite one, so it is attempting to compensate for the high residual at the $z_i=0$ end by sacrificing some fit quality at the other end. It may be noted that the residual is of the order of or less than 10^{-11} in./lbf, and is negligible compared to the characteristic amplitude of the ''measurement'' (Fig. 3), which is of the order 10^{-6} in./lbf. There is apparently no significant error associated with the truncation (23), at least if measurements are confined to distances at least 3 inches from loads or supports.

B. Sensitivity to distortions of lateral bending wavenumber

Having proved that the general fit quality is good, one asks the next question: to how small a change in lateral bending wavenumber is this method sensitive? To address this, the wavenumber k_m for the lowest bending mode, $k_{\text{lat. bend.}}$—the most dominant among the 11 terms—was declared unknown and a new value of $k_{\text{lat. bend.}}$ was sought that would give the best fit to the ''measurement.'' In the following studies, the region $z<3$ in. was excluded from the fit.

1. One-point scan

''Guesses'' for 'unknown' $k_{\text{lat. bend.}}$ were obtained by distorting both propagating and evanescent $k_{\text{lat. bend.}}$ by a varying amount δ, which never exceeded more than a few percent. The wavenumbers used in the fit are herein denoted as k_m:

$$k_m=\begin{cases} k_m, & \text{if not lateral bending,} \\ k_m(1+\delta), & \text{if prop. lateral bending,} \\ k_m(1-\delta), & \text{if even. lateral bending.} \end{cases} \tag{28}$$

The different signs of δ are consistent with expectations for the effect of the contained load. The same fitting procedure is employed as in the previous section and the residual and χ^2 values are studied for different values of distortion δ. For small distortions, the dependence of χ^2 on δ is quadratic. The sharpness of the parabola indicates the fit quality, while the position of the minimum is the best-fit solution for δ. In this particular study, the minimum is expected to be at exactly $\delta=0$, and achieved sharply.

Surprisingly, distorting $k_{\text{lat. bend.}}$ by as much as 1% left χ^2 almost unchanged (not shown). While the curve did achieve its minimum at the correct value ($\delta=0$), it lacked the needed sharp minimum of $\chi^2(\delta)$. That means that, in a real measurement, this procedure would not be able to discern small changes in $k_{\text{lat. bend}}$, thus raising a need for a more accurate procedure.

2. Two-point scan

A closer scrutiny of the behavior of the fit revealed that the least-squares algorithm allowed the fit quality to remain high, even with badly distorted $k_{\text{lat. bend.}}$, by adjusting the amplitudes α_m of the other terms (most notably the longitudinal, torsional and vertical bending waves) to absurdly large values. It appears that these long-wavelength terms can adjust their coefficients α_m so as to mimic the shorter-wavelength lateral bending wave, at least over a short range in z. This ability of the long-wavelength terms was curtailed in part by incorporating an additional scan, at a different point on the rail surface, P_2 (0.34 in.; 3.9 in.), coinciding with a mesh node on the upper portion of the rail web. The expression ''two scan points'' refers herein to scans along the z axis at two different positions on the surface of the rail.

The ''fitted'' displacements for the two scan points, $\hat{d}_1$ and $\hat{d}_2$, are calculated using Eq. (23), by reformulating the coefficients α_m as weighted eigenvectors:

$$\alpha_m^{P_1}=\beta_m\psi_m^{P_1}, \quad \alpha_m^{P_2}=\beta_m\psi_m^{P_2}, \tag{29}$$

where $\psi_m^{P_1}$ and $\psi_m^{P_2}$ are lateral components of eigenvector $\{\psi_r\}$ at points P_1 and P_2, respectively [a shorthand nomenclature for $\{\psi_r\}_{j_p}$ in Eq. (20), for points P_1 and P_2].

In addition, a closer scrutiny of coefficients α_m used to construct the ''measurement:''

$$\alpha_m^{P_1}=\frac{\psi_m^{P_1}\psi_m^{P_f}}{\psi_m^T[B]\psi_m}=\frac{\psi_m^{P_1}}{\sqrt{\psi_m^T[B]\psi_m}}\cdot\frac{\psi_m^{P_f}}{\sqrt{\psi_m^T[B]\psi_m}}, \tag{30}$$

revealed the relative contribution (marked by the curly bracket) of each of the 11 modes. It was concluded that 6 of the modes—propagating longitudinal, propagating and evanescent vertical bending and the 3 most evanescent modes—could be eliminated due to their negligible contribution to the displacement at that point. The remaining 5 modes are the propagating and evanescent lateral bending, the torsional and the two lowest less well known evanescent modes (emphasized by bold borders in Table I). It is perhaps not surprising that a lateral displacement response to a lateral forcing has little contribution from the vertical bending modes or longitudinal modes.

The suitably reformulated regression function (24) is

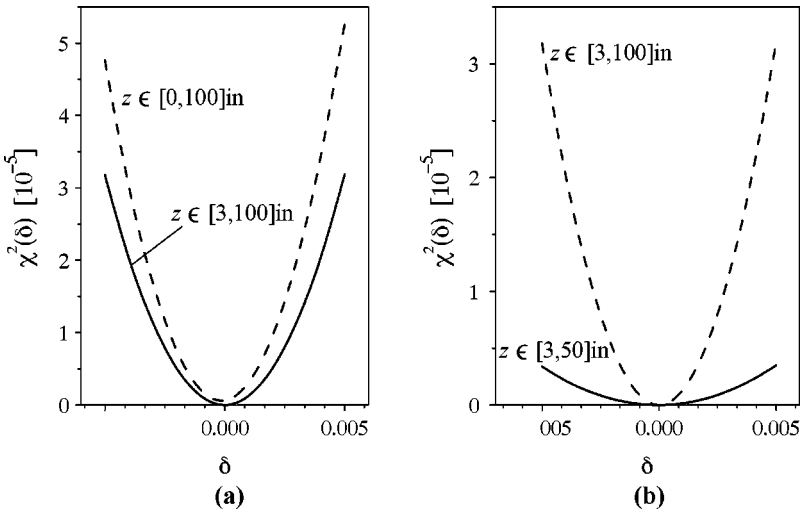


FIG. 5. Inferring the correct $k_{\text{lat. bend.}}$ ($\delta=0$) from the minimum of normalized $\chi^2(\delta)$ for a two-point scan: (a) Effect of proximity to point of loading; (b) effect of span length.

$$\chi^2(\beta_1,\ldots,\beta_5)=\sum_{i=31}^{1001}|d_1(z_i)-\hat{d}_1(z_i,\omega;\beta_1,\ldots,\beta_5)|^2 + \sum_{i=31}^{1001}|d_2(z_i)-\hat{d}_2(z_i,\omega;\beta_1,\ldots,\beta_5)|^2. \tag{31}$$

The data $d(z)$ is again taken from Eqs. (20), (22). $\hat{d}$ is taken from Eqs. (23), (29), now truncated to 5 terms. The same minimization procedure is applied, resulting in the following system of 5 linear equations with respect to 5 unknown coefficients β_m:

$$\sum_{m=1}^{5}\beta_m\sum_{i=31}^{1001}(\psi_m^{P_1}\psi_p^{P_1^*}+\psi_m^{P_2}\psi_p^{P_2^*})\mathrm{e}^{\mathrm{i}(k_p^*-k_m)z_i} = \sum_{i=31}^{1001}[d_1(z_i)\psi_p^{P_1^*}+d_2(z_i)\psi_p^{P_2^*}]\mathrm{e}^{\mathrm{i}k_p^*z_i}, \quad \forall p=1,\ldots,5. \tag{32}$$

This constrained-fit procedure is found to be satisfactory, with a sharp minimum of $\chi^2(\delta)$, a minimum located at the correct value ($\delta=0$), as shown in Figs. 5. The position of the minimum of $\chi^2(\delta)$ represents this method's guess for the most likely value of $k_{\text{lat. bend.}}$. The figures illustrate that, as expected, fit quality deteriorates (χ^2 increases) when the $z>3$ in. fit is used to predict the displacements below 3 in. [Fig. 5(a)], and that a sharper minimum is obtained by using a longer scan range [Fig. 5(b)]. The $\chi^2(\delta)$ in both plots was normalized by the mean square of combined "measurements" d_1 and d_2. Thus it represents the number of data points times the mean square fractional residual. The small values for $\chi^2(\delta)$ indicate very good fits.

C. Effects of measurement errors on fit quality

We studied several anticipated sources of measurement error in order to determine the acceptable precision with which the displacement should be measured to provide the desired accuracy in δ. All cases were considered with a scanning laser vibrometer in mind, as a displacement measurement device.

1. Random noise

To simulate the effect of measurement noise, a random number in a certain amplitude range was added to the old "measured" displacement:

$$\overline{d}(z_i)=d(z_i)+\vartheta(z_i), \tag{33}$$

where the clean data $d(z)$ is taken from the FEM solution, Eqs. (20) and (22), and the noise ϑ is uncorrelated. Noise is quantified as a fraction of rms data:

$$\vartheta(z_i)=\mathrm{rms}(|d|)\cdot(\%\text{rel. noise})\cdot(\text{random complex } \#\in[-1,+1]^2), \tag{34}$$

with only the last factor depending on z (a different complex random number for each data point). The second factor represents the dimensionless measure of the noise strength. The same 2-scan 5-term fitting procedure as in the previous subsection was employed, and the reduced χ^2 was plotted versus the distortion in $k_{\text{lat. bend.}}$ for 3 noise levels, with any value close to 1 representing a reasonable fit. The reduced χ^2 is in this case[23]

$$\chi^2_{\text{red}}=\frac{\sum_{i=31}^{1001}\dfrac{|\overline{d}_1(z_i)-\hat{d}_1(z_i)|^2+|\overline{d}_2(z_i)-\hat{d}_2(z_i)|^2}{\sigma^2}}{2\cdot 971-5}, \tag{35}$$

with $\hat{d}_1(z_i)$ and $\hat{d}_2(z_i)$ taken from a 5 term version of the series (23). The denominator is simply the number of points used in the fit (two points P and 971 values of z) minus the number of independent fit parameters (the five wave amplitudes β). The noise variance is $\sigma^{2a}=\mathrm{a\,meani}(i|\vartheta(z_i)|^2{}_i)$. A linear least-squares fit was employed, as in previous subsections, to obtain the coefficients β_m. The minima of parabolas for measurements with noise are expected to be shifted randomly to nonzero values of distortion δ, erroneously indicating the presence of longitudinal stress or other regular source of distortion of $k_{\text{lat. bend.}}$. This study is aimed at estimating what noise strength can the proposed technique tolerate while still achieving the desired accuracy of $k_{\text{lat. bend.}}$.

The error in guessing the $k_{\text{lat. bend.}}$ (i.e., the distortion δ) is greater for higher levels of noise. As shown in Fig. 6(a), the desirable sharpest curve corresponds to the lowest level

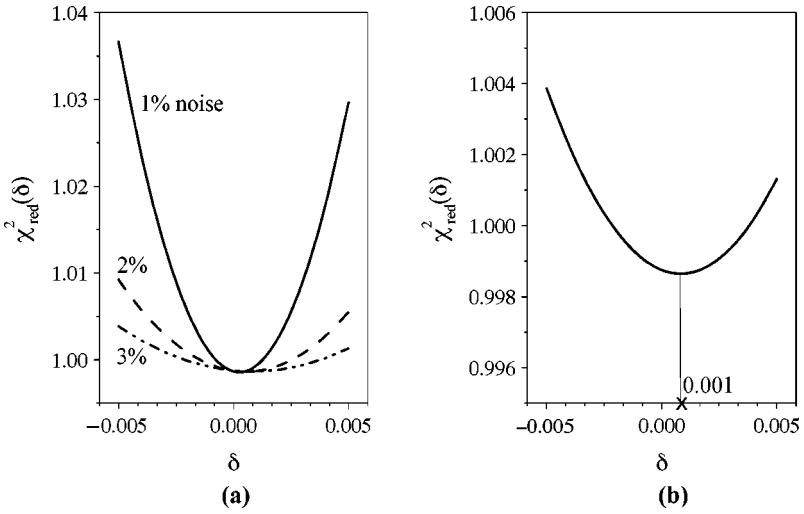


FIG. 6. Effect of noise in the "measurement" on $\chi^2_{\text{red}}(\delta)$: (a) Different levels of noise; (b) 3% noise.

of noise. Using a shorter scan ($z \in [3,50]$ in., corresponding to about 2/3 of the bending wavelength) at a noise level of 3%—the worst-case scenario—the curve peaks at around 0.001, i.e., 0.1% distortion [Fig. 6(b)]. That level of distortion and, therefore, the imprecision of retrieval of $\delta=0$ is tolerable, as it is less than the limiting 0.25%. Different sets of data with independent sources of noise (different seeds for the random number generator) yield similar values for the δ that minimizes χ^2, with random signs. These values fluctuated by $\pm 0.15\%$ as the precise realization of noise varied (different seed, same $[-1,+1]^2$ random number range). For a longer scan, $z \in [3,100]$ in, and the same level of noise, 3%, the fluctuations were only $\pm 0.06\%$. For lower levels of noise, the fluctuations in the best fit for $k_{\text{lat. bend.}}$ were correspondingly weaker.

The conclusion is that, for practical levels of noise in a displacement measurement, $k_{\text{lat. bend.}}$ can be ascertained with sufficient precision using the above fitting procedure. Considering that the imprecision of the $k_{\text{lat. bend.}}$ roughly scales with $\vartheta_{\%}/\sqrt{N_z}$ and calculating the proportionality coefficient from the above results for noise of 3%, one can conclude that this procedure gives $k_{\text{lat. bend.}}$ to within 1.5 $\vartheta_{\%}/\sqrt{N_z}$ for N_z scan points and a $\vartheta_{\%}$ level of noise at a fixed scan length, where $\vartheta_{\%}$ is given by (34) in percent. It is, of course, assumed here that there are no systematic measurement errors that might distort the fit. In the following subsection we deal with two characteristic systematic measurement errors that were identified during preliminary scans completed in the lab.

2. Vertical slant in scan trajectory

During *in situ* measurements of dynamic displacement, it is often more difficult to align the measurement apparatus than to control the noise levels of electronic instruments used in signal processing. Considering that the scan length required by the proposed technique (40 in.) is large in comparison with the characteristic length of the rail cross-section (7.3 in.), a small slant of the scanning apparatus with respect to the horizontal plane containing the waveguide axis results in a scan trajectory that does not follow the same point (x,y) on the cross-section along the rail, as shown in schematics in Fig. 7. The total rise at the end of the scan region is denoted by s. In order to assess the effect of such a slant in scan trajectory on the best estimate of $k_{\text{lat. bend.}}$, a certain slant was added to the "measurement" from the previous sections, under several assumptions. The lateral displacement due to the lateral bending mode changes very little along the vertical axis of the rail, y, and the change of lateral displacement due to the torsional mode is almost linear in y. Because the lateral component of the response has major contributions from these two modes, the lateral bending displacement was assumed to vary linearly in the y direction. The slope of this linear dependence was calculated using the lateral displacement values for points P_1 and P_2 from the previous subsection. It was also assumed that the same magnitude and character of misalignment are associated with both scans.

Consider a scan that systematically slants upwards. At every new z position, the laser is reading a point that is slightly above the point read at the previous z position. The slant-affected lateral displacement d^s of point P_1 at any z_i is taken to be a linear combination of displacements of P_1 and P_2 and it is obtained from the schematics in Fig. 8, by evoking Thales' theorem for the slant-affected displacements:

$$d_1^s(z_i) = d_1(z_i) + \frac{s}{\Delta y_{12}} \frac{d_1(z_{\text{end}}) - d_2(z_{\text{end}})}{z_{\text{end}} d_2(z_{\text{end}})} z_i d_2(z_i), \quad (36)$$

where $z_{\text{end}} = 100$ in. and $\Delta y_{12} = y_{P_1} - y_{P_2}$. An expression analogous to Eq. (36) holds for $d_2^s(z_i)$. The slant-affected displacements were then considered as the "measurement" that was then used in the 2-point 5-term fit procedure, with a varied magnitude of total rise (0.04 in., 0.08 in. and 0.24 in.). The result was a set of $\chi^2(\delta)$ curves that were almost coin-

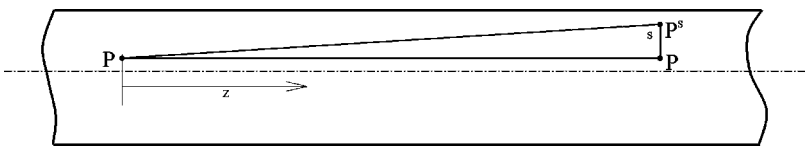


FIG. 7. Schematics of slanted scan (lateral view).

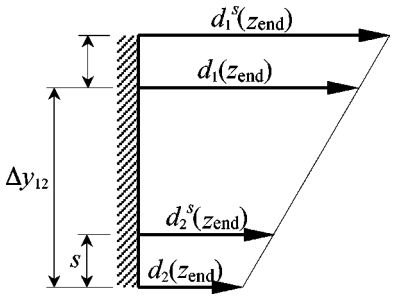


FIG. 8. Schematics of lateral displacement amplitudes at the end of the slant-affected scan region.

cidental, with only slightly different values for their minima—higher minimum for a greater total slant. However, the best fit value of δ remained practically zero (to within the roundoff error) in all cases. It is somewhat surprising that a vertical slant in the scan, of as much as 0.24 in, does not affect $k_{\text{lat. bend.}}$ significantly, although it slightly increases the minimum mean square residual. However, as a recommendation for the real measurements, the implication that a scan slant does not affect the accuracy of the stress estimate bodes well for implementation of the proposed method.

3. Sensitivity calibration drift

Preliminary measurements indicated that instrument calibration tended to drift slowly with time, in an uncontrolled but practically linear fashion (e.g., due to changes in temperature). In order to estimate how much such a drift would affect the best fit value for $k_{\text{lat. bend.}}$, the "measurement" was distorted by a linearly increasing gain at each z position. The drift-affected measurement d^d is taken to be

$$d_1^d(z_i) = d_1(z_i)\left(1 + \frac{z_i}{z_{\max}}\frac{\Delta_{\%}}{100}\right), \tag{37}$$

where $\Delta_{\%}$ is percent total drift. The usual fit procedure was employed with 0.25, 0.5 and 1% drift. The fit residuals for all cases are shown in Fig. 9.

The lower fit quality and higher residual are associated with greater drift, but, unlike the previous case, there is some change in the best fit value for $k_{\text{lat. bend.}}$, more noticeable for greater drift. Although $\delta = -0.03\%$ for a 1% drift is still

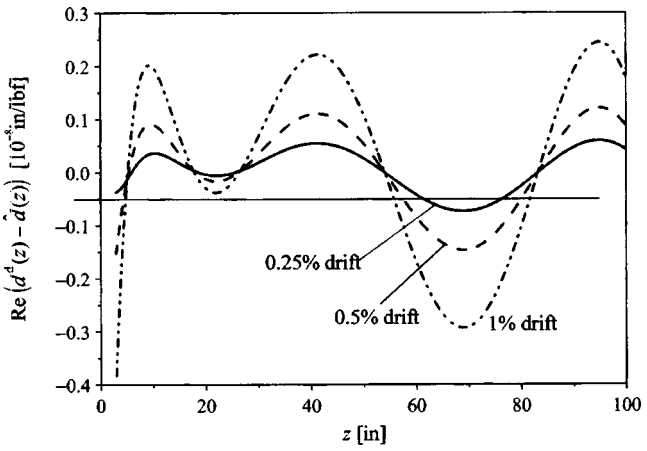


FIG. 9. Residuals between the drift-affected "measurements" and their fit.

within the limits of required accuracy, it indicates that the distortion due to drift should not be neglected.

In summary, the simulations show that, although all the sources of measurement error do affect the measured displacement and the fit quality and residuals, only the noise affects the $k_{\text{lat. bend.}}$ and, consequently, the accuracy of the stress estimate, while the slant in scan alignment does not. A drift in sensitivity of the measurement instruments results in only a slight effect. The plots of the fit residuals emerged as a "forensic" tool for recognizing which type of measurement error occurred, in addition to being useful indicators of the imprecision of the fit. Although these simulations do not pretend to model the *in situ* conditions exactly, they serve as guidelines for implementation of the proposed stress measurement technique.

V. CONCLUSIONS

A formal description for the steady-state harmonic response of a forced cylindrical waveguide has been developed in terms of the eigenvectors and eigen-wavenumbers of a FEM formulation for the free vibration of a uniform infinite beam of an arbitrary cross section. It has been illustrated for the particular case of a lateral point load on a railroad rail at low frequency. Simulated measurements, from data generated by these methods, were used to assess the potential for practical measurements to lend themselves to the inference of a lateral wavenumber by fitting such measurements to the specified general form for harmonic responses. Laboratory measurements in support of this are reported in Ref. 4.

It is noteworthy that the recovery of a lateral bending wavenumber was as robust as it appears to be, at least for the two-scan measurement scheme envisioned here, and for the types of systematic errors considered here. This is especially remarkable in light of the obvious challenges present in recovering the wavenumber from a data record consisting of only about one wavelength, and contaminated by other waves as well. That robustness bodes well for the eventual practical application of these ideas to the assessment of static contained axial load. Nevertheless, one can imagine a variety of other systematic errors that could be encountered in practice for which the tools developed here should be useful, e.g., an error in the assumed value for one of the other wavenumbers such as the torsional, or one of the mode shapes $\{\psi_r\}$, or a non-negligible contribution from vertical bending waves or extensional waves. A laboratory implementation is discussed elsewhere.[4]

ACKNOWLEDGMENTS

This work was supported by the Technical Scanning Committee of the Association of American Railroads and the Transportation Research Board's High Speed Rail IDEA program. The National Computational Science Alliance provided resources on its Silicon Graphics Origin2000.

[1] L. Gavrić, "Computation of propagative waves in free rail using a finite element technique," J. Sound Vib. **185**, 531–543 (1995).
[2] T. Hayashi, W.-J. Song, and J. L. Rose, "Guided wave dispersion curves for a bar with an arbitrary cross-section, a Rod and rail example," Ultrasonics **41**, 175–183 (2003).

[3] V. Damljanović and R. L. Weaver, ''Propagating and evanescent elastic waves in cylindrical waveguides of arbitrary cross section,'' J. Acoust. Soc. Am. **115**, 1572–1581 (2004).
[4] V. Damljanović, ''Guided wave technique for measurement of contained stress in uniform slender elastic bodies,'' Ph.D. dissertation, University of Illinois at Urbana—Champaign, 2003.
[5] M. R. Pfaffinger and J. Dual, ''Higher vibration modes in railway tracks at their cut-off frequencies,'' J. Sound Vib. **235**, 1015–1037 (2002).
[6] B. Aalami, ''Waves in prismatic guides of arbitrary cross section,'' J. Appl. Mech. **40**, 1067–1072 (1973).
[7] S. Widdehammar, P. A. Gradin, and B. Lundberg, ''Approximate determination of dispersion relations and displacement fields associated with elastic waves in bars,'' J. Sound Vib. **246**, 853–876 (2001).
[8] K. H. Huang and S. B. Dong, ''Propagating waves and edge vibrations in anisotropic composite cylinders,'' J. Sound Vib. **96**, 363–379 (1984).
[9] V. V. Volovoi, D. H. Hodges, V. L. Berdichevsky, and V. G. Sutyrin, ''Dynamic dispersion curves for non-homogeneous, anisotropic beams with cross-section of arbitrary geometry,'' J. Sound Vib. **215**, 1101–1120 (1998).
[10] T. Livingston, J. G. Béliveau, and D. R. Huston, ''Estimation of axial load in prismatic members using flexural vibrations,'' J. Sound Vib. **179**, 899–908 (1995).
[11] O. Onipede and S. B. Dong, ''Propagating waves and end modes in pretwisted beams,'' J. Sound Vib. **195**, 313–330 (1996).
[12] H. Taweel, S. B. Dong, and M. Kazic, ''Wave reflection from the free end of the cylinder with an arbitrary cross-section,'' Int. J. Solids Struct. **37**, 1701–1726 (2000).
[13] R. D. Cook, D. S. Malkus, and M. E. Plesha, *Concepts and Applications of Finite Element Analysis*, 3rd ed. (Wiley, New York, 1989).
[14] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[15] C. L. Dym and I. H. Shames, *Solid Mechanics: A Variational Approach*, Advanced Engineering Series (McGraw-Hill, New York, 1973).
[16] L. Meirovitch, *Fundamentals of Vibrations* (McGraw-Hill, Boston, 2001).
[17] J. Miklowitz, *Elastic Waves and Waveguides*, Series in Applied Mathematics and Mechanics (North-Holland, New York, 1978), Vol. 22.
[18] I. S. Sokolnikoff, *Mathematical Theory of Elasticity*, 2nd ed. (McGraw-Hill, New York, 1956).
[19] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[20] Of those that are real, only those with positive group velocity $(\mathrm{d}k/\mathrm{d}\omega)^{-1}$ are selected; for our case and for frequencies below the first cutoff, this corresponds to all real positive k's.
[21] For the problem that motivates this work, it is only the lateral bending wave number that is uncertain—due to uncertain static axial load. Other wave numbers are largely independent of this uncertainty, and can, in principle, be determined *a priori* from moduli and geometry. Should that data be unreliable, there would be a need for extraction of other wave numbers as well, and a corresponding need for more dynamic measurements.
[22] Practical constraints on removing rail fasteners, together with the enhanced sensitivity of $k_{\text{lat. bend}}$ to static load at long wavelengths lead us to the challenging regime in which $k_{\text{lat. bend}}$ must be extracted from a limited range of data $d(z)$.
[23] S. G. Rabinovich, *Measurement Errors and Uncertainties—Theory and Practice*, 2nd ed. (Springer-Verlag, New York, 2000).

Extension of measurement surface in near-field acoustic holography

Angie Sarkissian
Naval Research Laboratory, Washington D.C. 20375-5350

(Received 25 August 2003; revised 9 December 2003; accepted 12 December 2003)

In near-field acoustic holography, when measurements are made over a limited surface, there has been recent interest in numerically enlarging the measurement surface tangentially. Current algorithms use iterative methods to extrapolate the field tangentially outward. An algorithm based on the method of superposition is applied here which may be used to either extrapolate the field to enlarge the measurement surface or to interpolate in case there is a ''hole'' in the measurement surface where data are not available. [DOI: 10.1121/1.1645609]

PACS numbers: 43.40.Rj, 43.20.Tb [EGW] Pages: 1593–1596

I. INTRODUCTION

In near-field acoustic holography[1] or ''patch'' holography,[2–4] where measurements are made over a limited surface, there has been recent interest in numerically extending the measurement surface tangentially outward.[4–7] Current algorithms[4–7] use iterative methods to extrapolate the field. Williams *et al.* use this iterative procedure when applying patch holography.[4]

An algorithm based on the method of superposition[8–10] is applied here which approximates the field on and near the measurement surface by the superposition of fields produced by a number of sources placed on a surface near the measurement surface. The source strengths are determined by applying boundary conditions on the measurement surface. The algorithm may be used to either extrapolate the field to enlarge the measurement surface or to interpolate in case there is a ''hole'' in the measurement surface where data are not available.

Wu *et al.* have developed the Helmholtz equation least-squares method for acoustic holography where they expand the field in spherical functions.[11–13] The algorithm used here uses simple monopole sources placed on a surface and is thus very easy to implement.

II. METHOD OF SUPERPOSITION

Figure 1 contains the geometry, showing a structure radiating a harmonic field and a measurement surface containing N measurement locations. Measurement location j has coordinates $\mathbf{r}_j^m$. We place a surface σ at a constant distance δ from the measurement surface on the side where the radiating structure is located. We next state that the acoustic field on and near the measurement surface may be approximated by the field produced by N' sources placed on surface σ. Source j has coordinates $\mathbf{r}_j^s$ and source strength q_j. The field on the measurement surface may be written as

$$p(\mathbf{r}_i^m)=\sum_{j=1}^{N'} G_{ij}q_j, \tag{1}$$

where **G** is the Green's function matrix

$$G_{ij}=\frac{e^{ik|\mathbf{r}_i^m-\mathbf{r}_j^s|}}{|\mathbf{r}_i^m-\mathbf{r}_j^s|}; \tag{2}$$

$k=\omega/c$ is the wave number, c is the speed of sound in the fluid outside the structure. We may rewrite Eq. (1) in matrix form

$$\mathbf{p^m}=\mathbf{Gq}. \tag{3}$$

The source strengths are determined by applying boundary condition on the measurement surface. We invert matrix **G** and rewrite Eq. (3) as

$$\mathbf{q}=\mathbf{G}^{-1}\mathbf{p^m}. \tag{4}$$

If the number of measurement points is not equal to the number of sources, $N\neq N'$, matrix **G** will not be square and if the number of sources N' is large the matrix **G** will be ill-conditioned. We thus use the singular value decomposition method to invert it. We write **G** as

$$\mathbf{G}=\mathbf{USV}^{\dagger}, \tag{5}$$

where **U** and **V** are unitary matrices and **S** is a real, diagonal matrix containing positive singular values σ_j in order of decreasing value. The inverse of **G** may be approximated by the sum

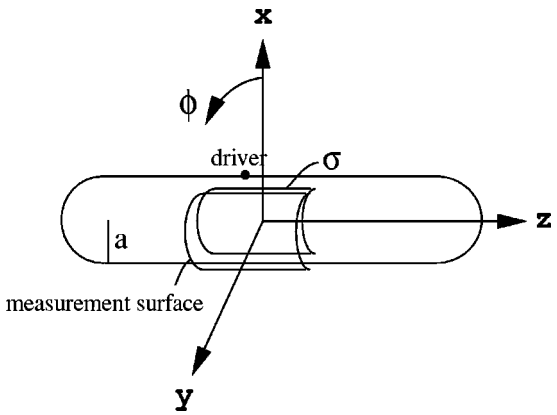


FIG. 1. Geometry.

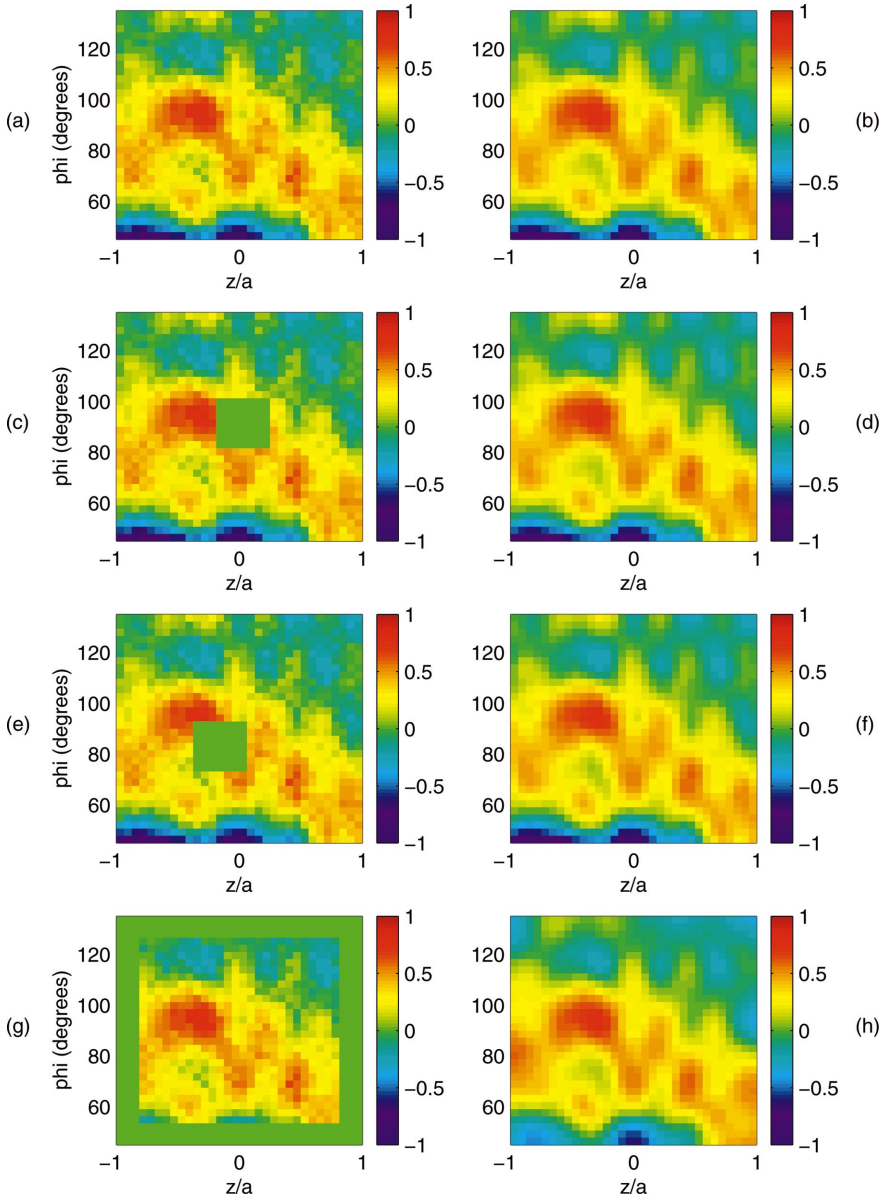


FIG. 2. (a) The real part of the simulated radiated field on the measurement surface (including noise) for $ka=4$; (c), (e), and (g) show the fields on the measurement surfaces where in green are shown regions where no data are available; (b), (d), (f), and (h) show the fields reconstructed from measurements shown in (a), (c), (e), and (g), respectively.

$$G_{ij}^{-1}=\sum_{1}^{N_t} V_{ik}\sigma_k^{-1}U_{jk}^{*}, \tag{6}$$

where the sum is truncated to keep only $N_t<N$ terms in order to remove the very small singular values that produce large errors in the reconstructed field.

Once the source strengths q_j are determined using Eq. (4) with the approximation of $\mathbf{G}^{-1}$ above, the field near the measurement surface may be approximated by

$$p(\mathbf{r})=\sum_{j=1}^{N'} \frac{e^{ik|\mathbf{r}-\mathbf{r}_j^s|}}{|\mathbf{r}-\mathbf{r}_j^s|}q_j\,. \tag{7}$$

This approach does not use iteration and no zero padding of the data is necessary since the Green's function contains elements G_{ij} where all points $\mathbf{r}_i^m$ lie only on the measurement surface where data are available.

III. IMPLEMENTATION

The algorithm is applied to numerically generated data. The structure is a hemispherically end-capped cylindrical shell having radius a. Its total length, including end caps, is $7a$ and its thickness is $0.0074a$. The structure contains 85 identical straight frames along its cylindrical region with equal spacing between them. The frames have length $0.078a$ and thickness $0.0065a$. We use parameters of nickel for the shell, where the Young's modulus is 2.1×10^{11} Pascal, the Poisson's ratio is 0.3, the density is 8800 kg/m^3, and the loss factor is 0.001. The sound speed in the fluid medium outside the structure is $c=1510$ m/s. We use SARA-2D (Ref. 14) to compute the radiated field, which uses finite and infinite elements to model the structure and the fluid medium.

We place the coordinate origin at the center of the cylinder, as shown in Fig. 1. Using cylindrical coordinates, the

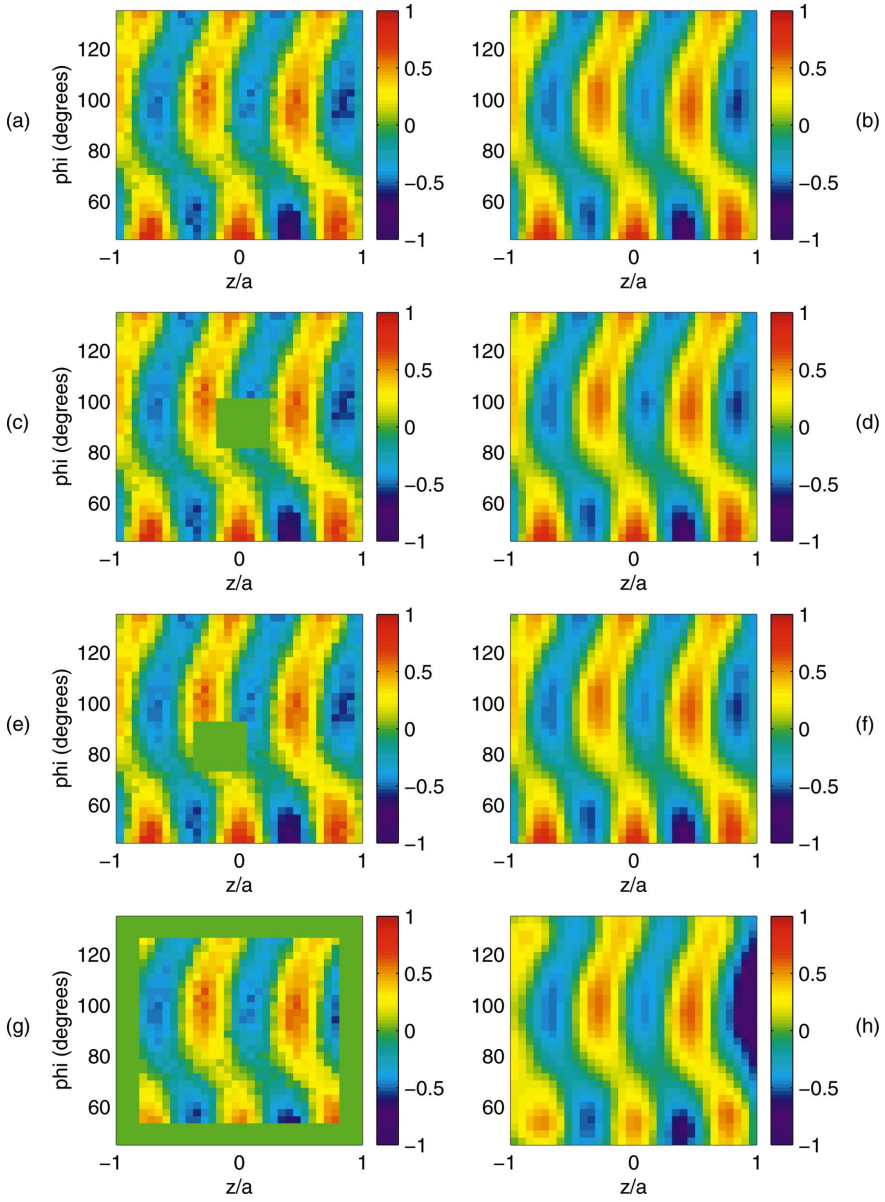


FIG. 3. (a) The real part of the simulated radiated field on the measurement surface (including noise) for $ka=8$; (c), (e), and (g) show the fields on the measurement surfaces where in green are shown regions where no data are available; (b), (d), (f), and (h) show the fields reconstructed from measurements shown in (a), (c), (e), and (g), respectively.

driver is located at $(a,0,-a/3)$, as shown. First, the field is computed for $ka=4$ at a 33×33 grid of points at a constant distance of $a/3$ away from the surface of the structure, on the curved surface $\rho=4a/3$, $\pi/4\leqslant\phi\leqslant3\pi/4$, $-a\leqslant z\leqslant a$. To this numerically computed field, we add random error that has 10% of the magnitude of the field so that the ratio of the variance of the error to the variance of the field values is 0.01. The normalized real part of the resultant field is shown in Fig. 2(a). We next take the four measurement surfaces shown in Figs. 2(a), (c), (e), and (g), where in green are shown regions where there are no data present. We apply the method of superposition to reconstruct the field everywhere on the measurement surface including the missing locations and show them in Figs. 2(b), (d), (f), and (h), respectively. We note that Fig. 2(b) is a reconstruction from a complete measurement surface, shown in Fig. 2(a).

When doing the reconstruction, we choose surface σ to be at a constant distance of $\delta=a/3$ from the measurement surface, which is approximately one-fifth of the acoustic wavelength. Although the value of δ is arbitrary, good judgment must be used to choose it. We found a value equal to $\frac{1}{3}$ to $\frac{1}{5}$ the acoustic wavelength to produce reasonable results. The grid of points on surface σ had identical z and ϕ coordinates as the measurement points in each case. Thus, we chose the number of points on surface σ, N', to equal the number of measurement points, N, in each case, although this is not a requirement. Also, our surface σ coincided with the structure surface, which is again not a requirement. We found reducing the offset distance δ by 20%, thus placing σ outside the structure surface, produced similar results. If δ becomes too large the Green's function matrix $\mathbf{G}$ will be more poorly conditioned.

In Eq. (6), the sum is truncated to keep only N_t singular values. We found for this case that choosing N_t that made the

ratio of the smallest to the largest singular value, $[\sigma(1)]/[\sigma(N_t)]$ of order 10^{-5} produced reasonable results. For the reconstructions shown in Figs. 2(b), (d), and (f), the sum was truncated keeping only $N_t=300$ singular values. For the reconstruction shown in Fig. 2(h), where the measurement surface was smaller, $N_t=150$ was used. One approach to choosing N_t would be to vary it until the field reconstructed over the measurement area agrees best with measurement values.

The reconstructed fields look visually similar to the original simulation shown on top. We compute the error in the region where data were missing [shown by green in Figs. 2(c), (e), and (g)] by comparing the reconstructed value to the field plus noise value shown in Fig. 2(a). For the regions shown in green in Fig. 2(c), (e), and (g), the average errors are 0.17, 0.21, and 0.62, respectively. The numerically simulated field was normalized to have a maximum magnitude of unity.

We next apply the same procedure to the case where the frequency is doubled to $ka=8$. In this higher-frequency case a smaller offset distance $\delta=a/4$ was used in the computations, which is approximately one-third of the acoustic wavelength.

Figure 3(a) shows radiated field plus noise. The noise was 10% in amplitude again. Figures 3(c), (e), and (g) show in green regions where data are missing. To the right of each is shown the reconstructed field on the entire measurement surface. In Fig. 3(b) we show the reconstruction from a complete surface. The average reconstruction errors for the regions shown in green in Figs. 3(c), (e), and (g) are 0.34, 0.39, and 0.67, respectively. The numerically simulated field values for this case also were normalized to have a maximum magnitude of unity. The errors are higher for the higher-frequency case as expected.

For this higher-frequency case, the sum in Eq. (6) was truncated, keeping $N_t=250$ singular values for the reconstructions in Fig. 3(b), (d), and (f). $Nt=125$ was used for the reconstruction in Fig. 3(h).

The reconstruction is reasonable as long as it is applied to regions near the measurement surface. It degrades in accuracy if applied at points far from the measurement surface because of the ill-posed nature of the problem. This ill-posed nature is not necessarily due to the presence of evanescent waves but because of the finite size of the measurement surface. When using the Helmholtz integral equation, we write the field at a given point in terms of the field values over a closed surface. In our case, we have a ''patch'' for a measurement surface instead of a closed surface which produces this ill-posed nature. For the lower frequency case, $ka=4$, reconstructions were made at distances of up to one-eighth of the acoustic wavelength away from the measurement surface. In the higher-frequency case, $ka=8$, fields were reconstructed at distances up to one-fourth of an acoustic wavelength from the surface, thus producing higher errors.

IV. SUMMARY

The method of superposition may be applied to reconstruct the acoustic field in regions tangential to the measurement surface located near the measurement surface. The algorithm is very easy to implement either to flat or curved surfaces. It may be used to extrapolate the field outward to enlarge the measurement surface or to interpolate if there is a ''hole'' in the measurement surface where no data are available.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1] J. D. Maynard, E. G. Williams, and Y. Lee, ''Near-field acoustic holography. I. Theory of generalized holography and the development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[2] A. Sarkissian, C. F. Gaumond, E. G. Williams, and B. H. Houston, ''Reconstruction of the acoustic field over a limited surface area on a vibrating cylinder,'' J. Acoust. Soc. Am. **93**, 48–54 (1993).
[3] J. Hald, ''Patch near-field acoustical holography using new statistically optimal method,'' Proceedings of Inter-noise 2003, p. 2203.
[4] E. G. Williams and B. H. Houston, ''Fast Fourier transform and singular value decomposition formulation for patch near-field acoustical holography,'' J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[5] K. Saijyou and S. Yoshikawa, ''Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography,'' J. Acoust. Soc. Am. **110**, 2007–2023 (2001).
[6] E. Williams, ''Continuation of acoustic nearfields,'' J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[7] K. Saijyou and H. Uchida, ''New data extrapolation method for boundary element method based near-field acoustical holography,'' J. Acoust. Soc. Am. **115**, 785–796 (2004).
[8] G. H. Koopmann, L. Song, and J. B. Fahnline, ''A method for computing acoustic fields based on the principle of wave superposition,'' J. Acoust. Soc. Am. **86**, 2433–2438 (1989).
[9] R. D. Miller, E. T. Moyer, Jr., H. Huang, and H. Uberall, ''A comparison between the boundary element method and the wave superposition approach for the analysis of the scattered fields from rigid bodies and elastic shells,'' J. Acoust. Soc. Am. **89**, 2185–2196 (1991).
[10] A. Sarkissian, ''Method of superposition applied to scattering from a target in shallow water,'' J. Acoust. Soc. Am. **95**, 2340–2345 (1994).
[11] Z. Wang and S. F. Wu, ''Helmholtz equation least-squares method for reconstructing the acoustic pressure field,'' J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[12] S. F. Wu and J. Yu, ''Reconstructing interior acoustic pressure fields via Helmholtz equation least-squares method,'' J. Acoust. Soc. Am. **104**, 2054–2060 (1998).
[13] S. F. Wu, ''On reconstruction of acoustic pressure fields using the Helmholtz equation least-squares method,'' J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[14] H. Allik, R. Dees, S. Moore, and D. Pan, SARA-2D *User's Manual* (BBN Systems and Technologies, New London, CT, 1995).

Annoyance caused by the sounds of a magnetic levitation train

Joos Vos[a)]
TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands

(Received 25 April 2003; accepted for publication 5 January 2004)

In a laboratory study, the annoyance caused by the passby sounds from a magnetic levitation (maglev) train was investigated. The listeners were presented with various sound fragments. The task of the listeners was to respond after each presentation to the question: ''How annoying would you find the sound in the preceding period if you were exposed to it at home on a regular basis?'' The independent variables were (a) the driving speed of the maglev train (varying from 100 to 400 km/h), (b) the outdoor A-weighted sound exposure level (ASEL) of the passbys (varying from 65 to 90 dB), and (c) the simulated outdoor-to-indoor reduction in sound level (windows open or windows closed). As references to the passby sounds from the maglev train (type Transrapid 08), sounds from road traffic (passenger cars and trucks) and more conventional railway (intercity trains) were included for rating also. Four important results were obtained. Provided that the outdoor ASELs were the same, (1) the annoyance was independent of the driving speed of the maglev train, (2) the annoyance caused by the maglev train was considerably higher than that caused by the intercity train, (3) the annoyance caused by the maglev train was hardly different from that caused by road traffic, and (4) the results (1)–(3) held true both for open or closed windows. On the basis of the present results, it might be expected that the sounds are equally annoying if the ASELs of the maglev-train passbys are at least 5 dB lower than those of the intercity train passbys. Consequently, the results of the present experiment do not support application of a railway bonus to the maglev-train sounds. [DOI: 10.1121/1.1650330]

PACS numbers: 43.50.Ba, 43.50.Lj, 43.50.Qp [DKW] Pages: 1597–1608

I. INTRODUCTION

Magnetic levitation (maglev) trains utilize an advanced technology in which magnetic forces lift, propel, and guide the vehicle over a guideway. The technology permits maximum speeds of up to about 500 km/h, which is almost twice as high as that of conventional high-speed trains. Because of its high speed, the maglev train is able to compete with auto and aviation modes for travel distances between about 75 and 1000 km, and is therefore an interesting travel option for the 21st century.

Test tracks of the maglev train have been built in Japan, Germany, and China. In the USA there are currently no maglev systems in operation, although there are several corridors that have been studied in detail and that are awaiting funding decision by the Federal Railroad Administration.[1] For some of these projects environmental impact statements are being prepared.

Similarly, in the interest of improving the infrastructure of the Northern part of The Netherlands, an intelligent choice among various alternative measures required detailed knowledge about the *annoyance* caused by the passby sounds from a maglev train. Since at least in Europe, there are at present no tracks of the maglev train located in or close to residential areas, a field survey could not be carried out. Consequently, the research was performed in the laboratory.

Some data on overall loudness (rather than annoyance) of passby sounds of a maglev train (type Transrapid 07) and more conventional trains (types EC, IC, ICE, and a freight train) have been reported by Fastl and Gottschling (1996) and by Gottschling and Fastl (1997). In these two related laboratory studies, the overall (or global) loudness ratings for the maglev and more conventional train sounds presented at comparable A-weighted equivalent sound levels were not significantly different.

In a laboratory study reported by Neugebauer and Ortscheid (1997), the overall loudness and other subjective evaluations were determined for passages of a maglev train and a conventional short-range train. For three relevant factors (evaluation, activity, and potency) summarizing the responses obtained with the method of the semantic differential, the sounds of the maglev train yielded significantly higher values than those of the short-range train, indicating that, overall, the subjects were more negative about the maglev train. Moreover, especially at the higher A-weighted equivalent sound levels, the sounds from the maglev train were considerably louder than those of the short-range train.

Results on a semantic study of acoustic and nonacoustic aspects in the evaluation of maglev and short-range train passby sounds have also been reported by Quehl (1999). The limited number of experimental conditions and imperfections in a portion of the passby sounds, however, prevented her from drawing firm conclusions.

In sum, the available data on the subjective evaluation of maglev-train sounds are limited, and the results are inconsistent: The results described in Fastl and Gottschling (1996) and in Gottschling and Fastl (1997) suggest that the railway bonus for conventional trains might also be applied to the maglev train, whereas the results from Neugebauer and

[a)]Electronic mail: vos@tm.tno.nl

Ortscheid (1997) might indicate that application of the bonus to the maglev train is not permitted.

In the present laboratory study, listeners had to rate the annoyance of various sound fragments. With respect to the validity of the experimental results obtained in laboratory studies, our experiences are positive. For example, differences in annoyance between road-traffic sounds and shooting sounds produced by small firearms, as found in field surveys, are obtained in laboratory studies also (Vos, 1995).

One of the independent variables in the present study was the driving speed of the maglev train (varying from 100 to 400 km/h). Since most environmental noise ordinances are based on sound levels measured outside residences, the second independent variable was the outdoor A-weighted sound exposure level (ASEL) of the passbys (varying from 65 to 90 dB). The annoyance inside the dwelling furthermore depends on the attenuation of the facade structure. As a result, the third independent variable was the simulated outdoor-to-indoor reduction in sound level (windows open or windows closed). As references to the sounds from the maglev train (type Transrapid 08), sounds from road traffic (passenger cars and trucks) and more conventional railway (intercity trains) were included for rating also.

II. METHODS

A. Sound fragments

The stimuli were sound passages of a maglev train, various intercity trains, a high-speed train, and various passenger cars and trucks. For all passages, free-field digital recordings were made.

The sounds of the maglev train (type Transrapid 08) were recorded in Lathen, Germany (de Graaff *et al.*, 2001). We selected passages at four driving speeds (100, 200, 325, and 400 km/h), each passage being simultaneously recorded at three distances of 25, 50, and 100 m. From these 12 recordings, 16 different sound fragments were prepared. Each fragment consisted of one passage with a duration of 15–20 s. Since the total duration of all fragments included in the present study was fixed at 45 s, the maglev-train passages were preceded and followed by silent periods of about 12–15 s. A realistic presentation of the fragments requires that, given a specific distance between the source and the receiver, the sound level at which a passage is reproduced in the laboratory does not significantly deviate from the level found in the field. A satisfactory representativeness was obtained by presenting the passages with driving speeds of 100 and 200 km/h at outdoor ASELs of 65–80 dB, and those with driving speeds of 325 and 400 km/h at outdoor ASELs of 75–90 dB.

For the passages with driving speeds of 100 and 200 km/h that were presented at the ASELs of 65 and 70 dB, as well as for the passages with driving speeds of 325 and 400 km/h that were presented at the ASELs of 75 and 80 dB, the sounds recorded at a distance of 100 m were used. For the passages with driving speeds of 100 and 200 km/h presented at an ASEL of 75 dB, and for those with driving speeds of 325 and 400 km/h presented at an ASEL of 85 dB, the recordings at a distance of 50 m were used. For the remaining passages presented at an ASEL of 80 dB (driving speeds of 100 and 200 km/h) or for those presented at an ASEL of 90 dB (driving speeds of 325 and 400 km/h), the sounds recorded at a distance of 25 m were used.

The fragments of the intercity trains were based on the sounds from passenger trains (types ICR/ICM and IRM/DD) recorded at distances of 35 and 100 m. For the outdoor ASELs of 75, 80, and 85 dB, the passages of train types ICR/ICM and IRM/DD recorded at a distance of 100 m were used. For the sound fragment presented at an outdoor ASEL of 90 dB, a passage of train type IRM/DD recorded at a distance of 35 m was used. The driving speed of the various trains was estimated to range between 120 and 140 km/h. Again, each fragment consisted of one passage. The duration of a passage was equal to 25–30 s.

The fragments of the high-speed train (type TGV-Atlantic) were based on passages recorded at distances of 50 and 200 m. The passage recorded at a distance of 200 m was presented at ASELs of 75 and 80 dB. The passage recorded at a distance of 50 m was presented at ASELs of 85 and 90 dB. The driving speed of the train was equal to about 300 km/h.

The fragments of road traffic were based on the sounds from passenger cars and trucks recorded at distances of 12.5–60 m from a provincial road. The driving speed was equal to about 80 km/h. Each fragment consisted of partly overlapping passages of 10–12 different passenger cars and one truck, with a total duration of 45 s. The maximum A-weighted levels of the truck passbys were about 10 dB higher than those of the passenger car passbys (Versfeld and Vos, 1997, 2002). The passages recorded at a distance of 60 m were presented at ASELs of 65 and 70 dB. The passages recorded at a distance of 25 m were presented at ASELs of 75 and 80 dB, and, for the sound fragments presented at an ASEL of 85 dB, the recordings at a distance of 12.5 m were used.

The original sound recordings were further processed. To isolate the sound produced by a specific source from the background noise, the amplitude of the background noise was shaped over short time intervals of about 1 s just prior to the beginning (fade-in) and directly after the end (fade-out) of the audible source-specific sound. All traffic sounds were subjected to this way of processing. In some recordings of the maglev train, highly prominent bird singing was removed by filtering as much as possible. The passage of the high-speed train recorded at a distance of 200 m contained non-specific low-frequency sounds. These sounds were removed with the help of a high-pass filter.

For each facade attenuation type, the level reduction is shown in Fig. 1. For the condition which simulated wide-open windows, an attenuation of 5 dB was assumed for frequencies between 12.5 and 1000 Hz. For higher frequencies the attenuation was 8 dB at most. With the windows closed, the facade attenuation increased from 12 dB for the 16- and 31.5-Hz octave bands up to 35 dB for the 8-kHz octave band, and represented the average of noise attenuations that are frequently found for Dutch dwellings with the windows closed (Vos, 2001).

The sounds were reproduced in a relatively small listening room ($w \times l \times h = 3.5 \times 5.9 \times 3.3$ m^3). Specific resonance

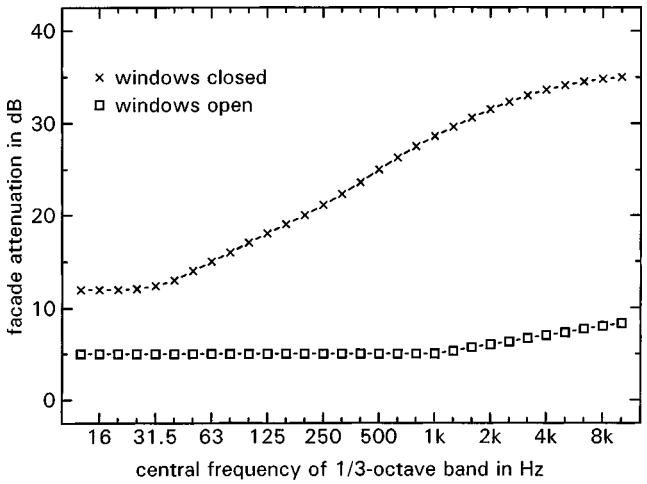


FIG. 1. Frequency-dependent outdoor-to-indoor sound reduction for two conditions.

frequencies (such as those around 30 and 60 Hz) of the room, resulting in changes in the sound spectrum, as well as nonflat frequency characteristics of the audio equipment, were compensated as much as possible. Finally, for frequencies between 25 and 8000 Hz, a flat frequency characteristic was obtained within about 4 dB (standard deviation of the deviations in the various $\frac{1}{3}$-octave bands equal to 2.6 dB).

Spectra of the sound fragments were determined with the help of a Larson-Davis spectrum analyzer (Model 3200) and a Brüel & Kjaer sound level meter (type 2236), with the microphone positioned at the ears of the subjects. The sound spectra are expressed as the linear sound exposure level in the various $\frac{1}{3}$-octave bands. The reference sound pressure is always equal to 20 μPa. In the present report, spectra are shown in the condition which simulated wide-open windows only.

Figure 2 shows the linear sound exposure level in the various $\frac{1}{3}$-octave bands for the passages of the maglev train. For each driving speed four spectra are shown, corresponding to outdoor ASELs of 65, 70, 75, or 80 dB for driving speeds of 100 and 200 km/h, and to outdoor ASELs of 75, 80, 85, or 90 dB for driving speeds of 325 and 400 km/h. Figure 2(a) shows a characteristic spectral peak around 315 Hz for the driving speed of 100 km/h. For a speed of 200 km/h [Fig. 2(b)], such a peak is found around 630 Hz. The spectral peak around 315 Hz for the driving speed of 100 km/h and that around 630 Hz for the driving speed of 200 km/h result from the groove passage frequency of stator grooves spaced at 0.083 m (de Graaff *et al.*, 2001). Due to aerodynamic noise at the speeds of 325 [Fig. 2(c)] and 400 km/h [Fig. 2(d)], the groove passage related frequency components do not longer determine the levels in the $\frac{1}{3}$-octave bands: Relevant spectral energy is found over a wide range between 100 and 2000 Hz.

Figure 3 shows the four sound spectra for the intercity trains. In addition to the smaller spectral peak around 31.5 and 63 Hz, a highly significant peak around 1600 Hz is obtained.

Figure 4 shows that the sounds of the high-speed train passbys contain very much energy up to frequencies of about 3–4 kHz. For frequencies between 25 and about 160 Hz,

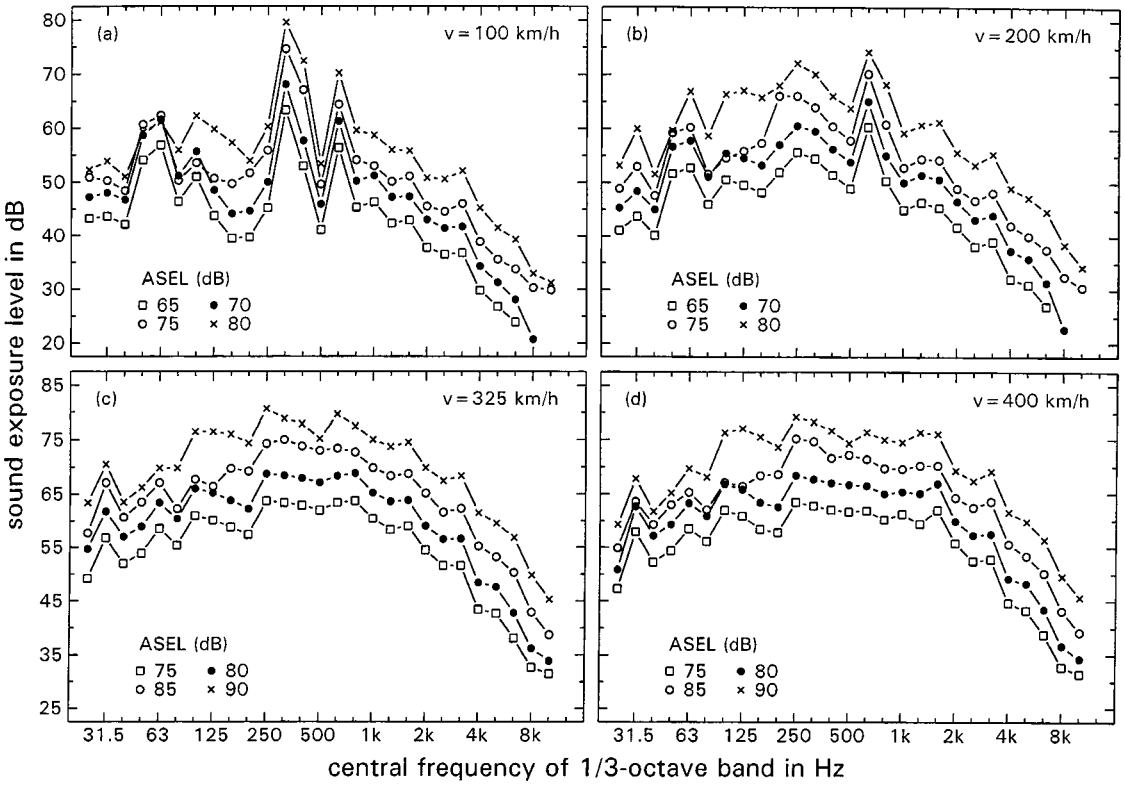


FIG. 2. Sound exposure levels in the various $\frac{1}{3}$-octave bands for four speeds of the maglev-train passbys, as determined at the ears of the subjects in the conditions which simulated open windows. For each driving speed, four spectra are shown with overall outdoor ASELs as indicated. Driving speed in km/h: (a) 100, (b) 200, (c) 325, and (d) 400.

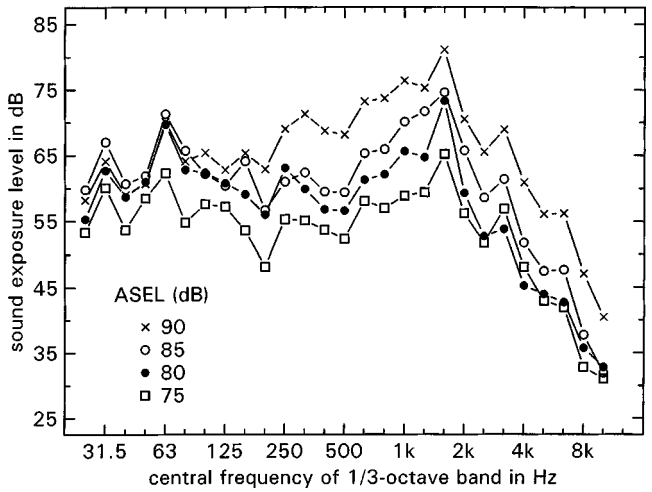


FIG. 3. Sound exposure levels in the various $\frac{1}{3}$-octave bands for intercity-train passbys, as determined at the ears of the subjects in the conditions which simulated open windows. Four spectra are shown with overall outdoor ASELs as indicated.

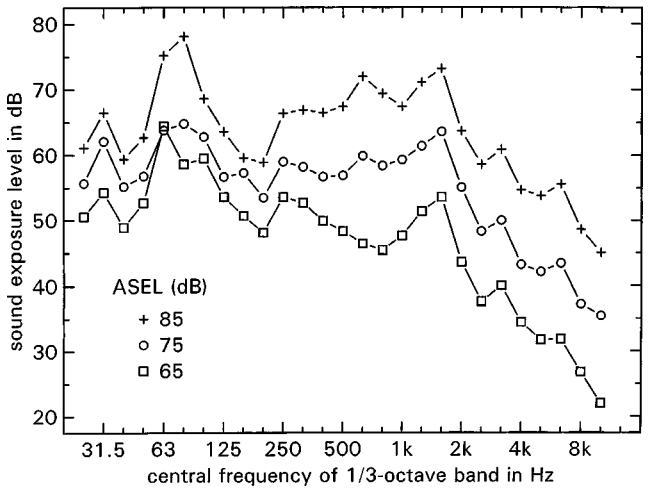


FIG. 5. Sound exposure levels in the various $\frac{1}{3}$-octave bands for road-traffic sounds, as determined at the ears of the subjects in the conditions which simulated open windows. Three of the five spectra are shown with overall outdoor ASELs as indicated.

large differences in sound level were obtained between the passby that was recorded at a distance of 50 m (ASELs of 85 and 90 dB) and the passby that was recorded at a distance of 200 m (ASELs of 75 and 80 dB). This difference must be the result of the high-pass filter that was used for ''improving'' the quality of the passby sound recorded at the larger distance of 200 m. Apparently, the filtering had also affected the source-specific spectral content. As a result, the representativeness of the passby sounds presented at ASELs of 75 and 80 dB must be questioned.

Figure 5 shows three of the five spectra for the fragments with road-traffic sounds. Again, these sounds contain relatively much energy in a wide frequency range. A significant decrease in sound exposure level is found for frequencies higher than about 1600 Hz.

In order to make the acoustic environment more realistic, a soft, spectrally shaped noise was continuously present throughout the experiment. In the condition which simulated open windows, the sound resembled traffic sounds as heard at a relatively great distance from a roadway. This background noise was presented at an A-weighted equivalent sound level of 35 dB, measured at the ears of the subjects. In the condition which simulated closed windows, it was felt that the background noise should resemble the sounds from the central heating system or the ventilation system. To achieve this, the background noise was further subjected to an overall reduction in sound level and an additional attenuation of the low-frequency components. In the listening room, this background noise was presented at an A-weighted equivalent sound level as low as 29 dB.

The spectra of the two background noise types are shown in Fig. 6. The binaural hearing threshold of otologically selected young listeners is inserted in the figure to emphasize that, in general, the sound components with frequencies lower than 63–100 Hz are no longer audible.

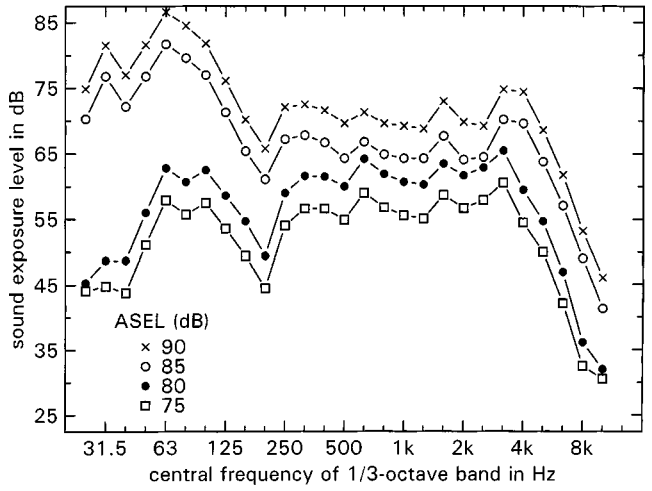


FIG. 4. Sound exposure levels in the various $\frac{1}{3}$-octave bands for high-speed train passbys, as determined at the ears of the subjects in the conditions which simulated open windows. Four spectra are shown with overall outdoor ASELs as indicated.

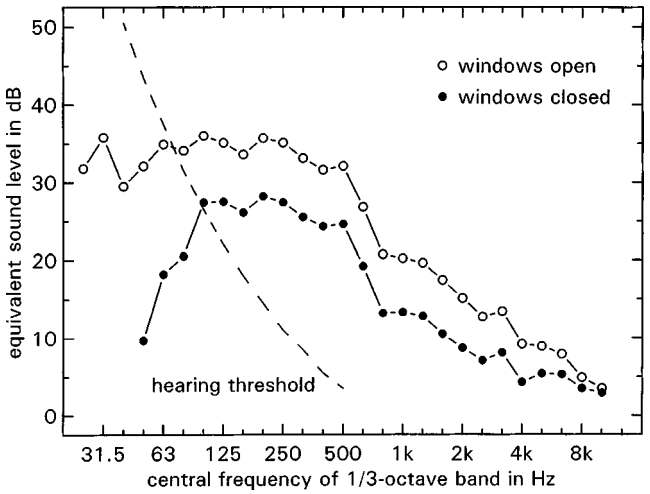


FIG. 6. Linear weighted equivalent sound level in the various $\frac{1}{3}$-octave bands of the background noise in two conditions, measured at the ears of the subjects. Inserted is the hearing threshold of young listeners.

B. Apparatus

The experiment was entirely computer controlled. The sounds were reproduced in the listening room by means of a loudspeaker (JBL-4425) hidden behind a curtain. The subjects were sitting behind a table furnished with a monitor and a keyboard. The distance between the listeners and the loudspeaker was about 3 m. For frequencies above 100 Hz, the reverberation time of the sound-insulated room was shorter than 0.5 s. Hearing thresholds were determined with the help of a Madsen memory threshold audiometer (MTA 86) with the function switch in the auto-threshold mode with pulsating tones.

C. Subjects

Twelve normally hearing subjects (six males and six females) between 23 and 34 years of age participated in the experiment. The mean age was equal to 27.2 years; the standard deviation equaled 4.1 years. Before the experimental sessions, the hearing thresholds of the subjects were determined between 250 and 8000 Hz for the left- and right-hand ears separately. Ten subjects had hearing levels $\leqslant$10 dB, and two subjects had hearing levels $\leqslant$15 dB in any part of the audiogram (best ears). The subjects were paid for their services.

D. Procedure

After hearing levels had been determined, the subjects were seated in the listening room. The subjects were tested individually. The appropriate background noise (Fig. 6) was present from the beginning of the sessions. The subjects were told that they were exposed to conditions in which traffic sounds (cars and trains passing by) could be heard either for the entire time period of 45 s, or for a portion of this time period. The beginning and end of each 45-s condition was indicated on the monitor of their personal computer. After each condition the subjects responded to the question ''How annoying would you find the sound in the preceding period if you were exposed to it at home on a regular basis?'' They were instructed that while rating the sounds, they had to take into account everything that they heard in the 45-s time period. Moreover, they were encouraged to use the whole range of the rating scale with values from 0 (''not annoying at all'') to 9 (''extremely annoying'').

Six subjects started with the windows-open conditions, and the other six started with the windows-closed conditions. Before these experimental blocks, the subjects received six representative sound fragments to familiarize them with the differences among the conditions. Both in the training blocks and in the experimental blocks, the presentation order of the sound fragments was randomized. To enhance the reliability of the results, each condition was presented twice for rating in separate blocks.

For annoyance scores greater than 4, both in the training block and in the second experimental block of each facade attenuation type, the subjects had to respond to five questions that informed about the causes of the expected annoyance. The preselected causes included were the perception of ''loudness'' (in Dutch: ''luidheid'') and other specific sound characteristics such as ''heavily pounding'' or ''banging'' (the Dutch word ''bonkend'' was used), and ''squealing,'' ''shrilly,'' or ''squeaky'' (the Dutch word ''snerpend'' was used). The other causes included were the feeling of ''insecurity'' or ''unsafety'' (in Dutch: ''onveiligheid'') and the reaction of ''startle'' (in Dutch: ''schrik''). For each of these five questions, there were five response alternatives: quite correct, considerably correct, I don't know, not entirely correct, or not at all correct. In all conditions, the rating tasks were self-paced: The subjects were allowed to spend as much time for responding as they considered necessary.

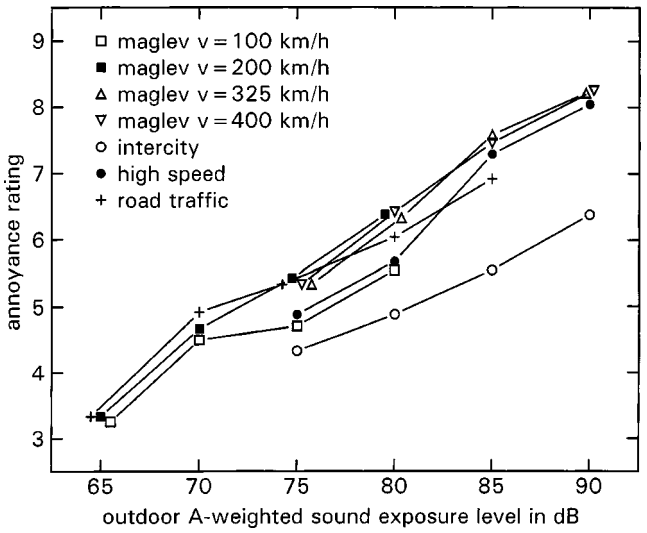


FIG. 7. Mean indoor annoyance ratings in the conditions which simulated open windows, for the various sound fragments as a function of the outdoor ASELs.

III. RESULTS

As explained in Sec. I, the annoyance is related to outdoor levels, and the results are presented for the two facade attenuation types separately. The responses of the subjects were considered reliable if the correlation coefficients, r, computed between the first and second ratings for each subject separately, were higher than 0.5. There was one subject who in the condition which simulated open windows did not fulfill this criterion. The data of the subject were replaced by those of a new subject with r-values that were considerably higher than the required criterion.

A. Annoyance in the windows-open conditions

The r-values, computed between the first and second ratings of the 29 sound fragments for each subject separately, ranged between 0.54 and 0.82 ($M=0.72$, $s.d.=0.09$). Analyses of variance performed for the annoyance scores from three subsets of the data that are defined in Sec. III A 1, showed that the mean scores obtained in the first measurements were not significantly different from those obtained in the second measurements ($p>0.05$), and that there were no significant first and second order interaction effects between replication and the stimulus variables ($p\geqslant 0.10$). In the more detailed presentation of the results below, it was therefore decided to average across subjects and replications.

Figure 7 shows the annoyance scores, averaged across

TABLE I. Annoyance ratings obtained in the conditions which simulated open windows. M=mean annoyance ratings, σ_m=standard error of the mean.

			Outdoor ASEL (dB)					
Sound source	V (km/h)		65	70	75	80	85	90
Maglev	100	M	3.3	4.5	4.7	5.5	⋯	⋯
		σ_m	0.31	0.32	0.29	0.26	⋯	⋯
	200	M	3.3	4.7	5.4	6.4	⋯	⋯
		σ_m	0.32	0.29	0.30	0.28	⋯	⋯
	325	M	⋯	⋯	5.3	6.3	7.6	8.2
		σ_m	⋯	⋯	0.28	0.30	0.21	0.22
	400	M	⋯	⋯	5.3	6.4	7.5	8.3
		σ_m	⋯	⋯	0.27	0.23	0.26	0.20
Intercity	120–140	M	⋯	⋯	4.3	4.9	5.5	6.4
		σ_m	⋯	⋯	0.42	0.33	0.31	0.31
High speed	300	M	⋯	⋯	4.9	5.7	7.3	8.0
		σ_m	⋯	⋯	0.39	0.34	0.27	0.27
Road traffic	80	M	3.3	4.9	5.3	6.0	6.9	⋯
		σ_m	0.32	0.29	0.28	0.26	0.22	⋯

subjects and replications, as a function of outdoor ASEL for each sound source separately. The mean values and the standard error of the mean are given in Table I.

In addition to main effects of ASEL and sound source, there was an interaction effect of these two variables. The statistical significance of these effects was tested in analyses of variance.

1. Analyses of variance

Data set I. The annoyance ratings for the road-traffic sounds and the sounds of the maglev train driving by at the two lower speeds, each presented at ASELs of 65, 70, 75, and 80 dB, were subjected to an analysis of variance [12 (subjects)×3 (sound type)×4 (sound level)×2 (replication), all variables within subjects]. The ratings significantly increased with increasing sound level [$F(3,33)=82.0$, $p<0.000\,001$]. The relatively small effect of sound type on the annoyance was just statistically significant [$F(2,22)=3.7$, $p<0.05$]. The mean scores suggest that the increase in the annoyance with sound level was smaller for the maglev train passing by at a velocity of 100 km/h than the increase obtained for the other two sound conditions. The interaction between sound type and sound level, however, was not significant [$F(6.66)=1.70$, $p>0.13$].

Data set II. Next, the annoyance ratings for the sounds of the maglev train passing by at the two higher driving speeds and the annoyance ratings for the intercity and high-speed trains, each presented at ASELs of 75, 80, 85, and 90 dB, were subjected to an analysis of variance [12 (subjects)×4 (sound type)×4 (sound level)×2 (replication), again with all variables within subjects]. The highly significant effect of sound level on the ratings found in the previous analysis was confirmed [$F(3,33)=63.0$, $p<0.000\,001$]. The ratings were also significantly affected by sound type [$F(3,33)=17.1$, $p<0.000\,01$]. This effect can be mainly ascribed to the relatively low annoyance ratings obtained for the intercity trains. The significant interaction between sound type and sound level [$F(9,99)=2.38$, $p<0.02$] can be largely explained by the annoyance caused by the high-speed train: for lower ASELs of 75 and 80 dB, its annoyance was in between that of the intercity train and the maglev train, whereas for the higher ASELs of 85 and 90 dB, the annoyance was no longer different from that caused by the maglev train. An analysis of variance in which the annoyance ratings for the high-speed train were excluded, suggested that the difference in annoyance between the maglev train and the intercity-train sounds tended to increase with increasing ASEL [$F(6,66)=2.15$, $p<0.06$].

Data set III. Inclusion of the annoyance ratings for road traffic in the analysis for ASELs of 75, 80, and 85 dB did not yield new information. In fact, of specific interest is the answer to the question whether the increase in the annoyance caused by the road-traffic sounds with increasing sound level is significantly less than that of the maglev-train sounds. The results of a separate analysis on the ratings for just these conditions showed that this interaction effect was not statistically significant [$F(4,44)=1.2$, $p>0.32$]. Moreover, this analysis showed that averaged across the three sound levels, there was no significant difference among the annoyance caused by the road-traffic sounds and the sounds of the maglev train passing by at the two different speeds [$F(2,22)=2.1$, $p>0.14$].

2. Summary of the main effects

Taking for granted the small and hardly significant difference in annoyance between (a) the sounds of the maglev train passing by at the speed of 100 km/h and (b) the sounds of the maglev train passing by at the speed of 200 km/h and the sounds of road traffic, three main conclusions may be drawn. At comparable ASELs (1) the annoyance was practically independent of the driving speed of the maglev train, (2) the annoyance caused by the maglev train was not different from the annoyance caused by road traffic, but (3) considerably higher than the annoyance caused by the intercity trains. The limited number of sound recordings of the high-speed train, together with the poor sound quality of one of these recordings that negatively affected the realistic character of two of the four sound fragments, do not allow us to draw firm conclusions about the annoyance caused by this sound source.

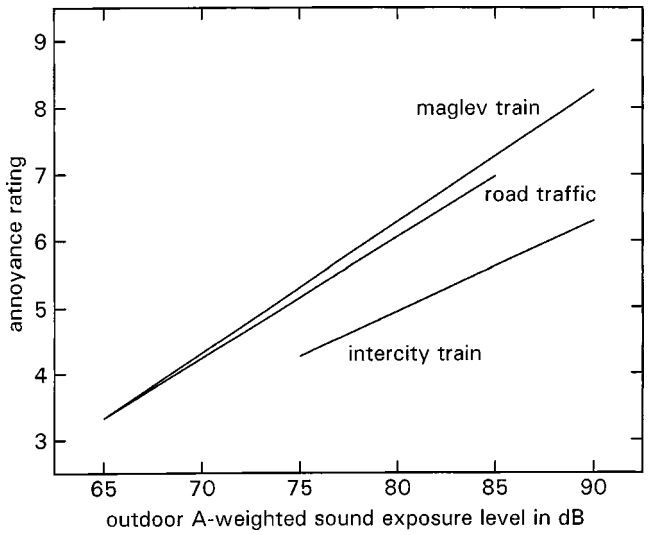


FIG. 8. Dose-response relations for maglev train, road traffic, and intercity train, in the conditions which simulated open windows. The solid lines are linear regression functions.

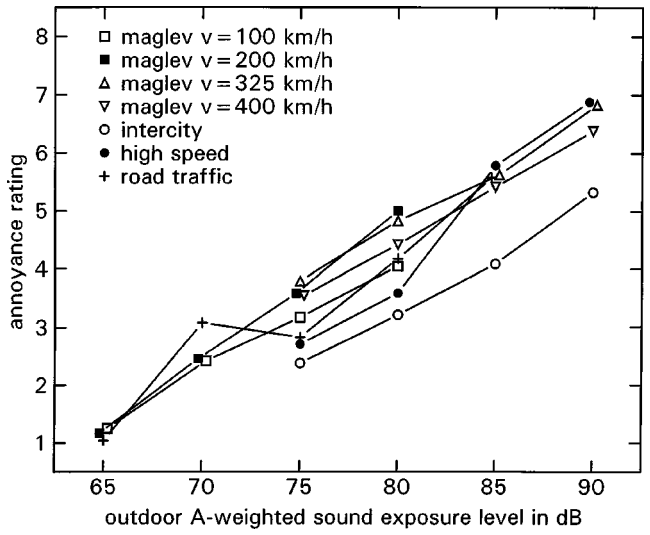


FIG. 9. Mean indoor annoyance ratings in the conditions which simulated closed windows, for the various sound fragments as a function of the outdoor ASELs.

Figure 8 shows the three dose-response relations. The relations were obtained by linear fits of the 16 mean ratings for the maglev train ($y=-9.47+0.197L_{AE}$, $r=0.98$), the five mean ratings for road traffic ($y=-7.14+0.166L_{AE}$, $r=0.98$), and the four mean ratings for the intercity trains ($y=-5.95+0.136L_{AE}$, $r=0.995$). The dose-response relation for the high-speed train (not shown in Fig. 8) is given by $y=-11.85+0.222L_{AE}$ ($r=0.99$).

With road traffic as the reference, Fig. 8 demonstrates that the bonus for the intercity trains varies from about 5 dB at $L_{AE}=75$ dB to about 9 dB at $L_{AE}=90$ dB. The difference in annoyance between the two relevant train types can be quantified as well: The types are equally annoying if, dependent on sound level, the ASEL of the maglev-train passby-sound is 5–10 dB lower than that of the intercity-train passby-sound.

B. Annoyance in the windows-closed conditions

The r-values, computed between the first and second ratings of the 29 sound fragments for each subject separately, ranged between 0.57 and 0.88 ($M=0.74$, $s.d.=0.09$). Analyses of variance performed for the annoyance scores from three data sets defined in Sec. III B 1 showed that the mean scores obtained in the first measurements were not significantly different from those obtained in the second measurements ($p>0.85$). In only one of the three subsets, there were small and hardly interpretable first order interaction effects between replication and the stimulus variables ($p\geqslant 0.02$). For example, the second ratings for the sounds of the intercity train and the maglev train ($v=400$ km/h) were slightly higher than the first ratings, whereas for the sounds of the high-speed train and the maglev train ($v=325$ km/h), the second ratings were slightly lower than the first ratings. Second order interaction effects between replication, sound level, and sound type were not significant ($p\geqslant 0.14$). In the more detailed presentation of the results below, it was therefore decided to average across subjects and replications.

Figure 9 shows the annoyance scores, averaged across subjects and replications, as a function of outdoor ASEL for each sound source separately. The mean values and the standard error of the mean are given in Table II.

Again, there were main and interaction effects of ASEL and sound source. The statistical significance of these effects was tested in analyses of variance.

1. Analyses of variance

Data set I. The annoyance ratings for the road-traffic sounds and the sounds of the maglev train driving by at the two lower speeds, each presented at ASELs of 65, 70, 75, and 80 dB, were subjected to the same analysis of variance that had been used in Sec. III A 1 for data set I.

The ratings significantly increased with increasing sound level [$F(3,33)=88.2$, $p<0.000\,001$]. The effect of sound type on the ratings was not significant [$F(2,22)=1.1$, $p>0.35$]. The interaction between sound type and sound level was statistically significant [$F(6.66)=3.27$, $p<0.007$]. This interaction effect is caused by the somewhat erratic relation between the annoyance ratings and sound level for the road-traffic sounds presented at ASELs of 70 and 75 dB. (In Sec. IV A, it will be explained that the unexpectedly high annoyance caused by the road-traffic sounds at $L_{AE}=70$ dB is probably related to the casual presence of relatively much low-frequency energy in the truck passby sound.) An analysis of variance in which the annoyance ratings for the road-traffic sounds were excluded showed that, for the maglev-train sounds, the increasing difference in annoyance between the two driving speeds with sound level was not statistically significant [$F(3,33)=1.97$, $p>0.14$].

Data set II. Next, the annoyance ratings for the sounds of the maglev train passing by at the two higher driving speeds and the annoyance ratings for the intercity and high-speed trains, each presented at ASELs of 75, 80, 85, and 90 dB, were subjected to an analysis of variance with the design that has also been used in Sec. III A 1 for data set II. Again, the annoyance ratings were highly affected by sound level

TABLE II. Annoyance ratings obtained in the conditions which simulated closed windows. M=mean annoyance ratings, σ_m=standard error of the mean.

			Outdoor ASEL (dB)					
Sound source	V (km/h)		65	70	75	80	85	90
Maglev	100	M	1.3	2.4	3.2	4.0	⋯	⋯
		σ_m	0.27	0.29	0.30	0.38	⋯	⋯
	200	M	1.2	2.5	3.6	5.0	⋯	⋯
		σ_m	0.23	0.34	0.35	0.30	⋯	⋯
	325	M	⋯	⋯	3.8	4.8	5.6	6.8
		σ_m	⋯	⋯	0.27	0.31	0.26	0.26
	400	M	⋯	⋯	3.5	4.4	5.4	6.4
		σ_m	⋯	⋯	0.26	0.33	0.34	0.29
Intercity	120–140	M	⋯	⋯	2.4	3.2	4.1	5.3
		σ_m	⋯	⋯	0.31	0.31	0.32	0.34
High speed	300	M	⋯	⋯	2.7	3.6	5.8	6.9
		σ_m	⋯	⋯	0.38	0.30	0.31	0.27
Road traffic	80	M	1.0	3.1	2.8	4.2	5.6	⋯
		σ_m	0.29	0.40	0.36	0.35	0.34	⋯

$[F(3,33)=99.6,\ p<0.000\,001]$. The effect of sound type was highly significant as well $[F(3,33)=18.3,\ p<0.000\,008]$. This effect can be mainly ascribed to the relatively low annoyance ratings for the intercity trains. The significant interaction between sound type and sound level $[F(9,99)=3.47,\ p<0.001]$ can be explained largely by the annoyance caused by the sounds from the high-speed train: for the lower ASELs of 75 and 80 dB, the ratings were still close to those of the intercity-train sounds, whereas for the higher ASELs of 85 and 90 dB, the ratings were almost equal to those for the maglev train.

Data set III. Inclusion of the annoyance ratings for road traffic in the analysis for ASELs of 75, 80, and 85 dB did not yield new information. As in Sec. III A 1, of specific interest is whether the increase in the annoyance caused by the road-traffic sounds with increasing sound level was significantly less than that of the maglev-train sounds. The results of a separate analysis on the ratings for just these conditions showed that this interaction effect was not statistically significant $[F(4,44)=1.6,\ p>0.20]$. Moreover, this analysis showed that averaged across the three sound levels, there was no significant difference among the annoyance caused by the road-traffic sounds and the sounds of the maglev train passing by at the two different speeds $[F(2,22)=2.9,\ p>0.08]$.

2. Summary of the main effects

Again, three main conclusions may be drawn. At comparable outdoor ASELs (1) the annoyance was practically independent of the driving speed of the maglev train, (2) the annoyance caused by the maglev train was not different from the annoyance caused by road traffic, but (3) considerably higher than the annoyance caused by the intercity trains. Due to the imperfections noted in Sec. II A, it would be premature to draw firm conclusions about the annoyance caused by the high-speed train.

Figure 10 shows the three dose-response relations. The relations were obtained by linear fits of the 16 mean ratings for the maglev train ($y=-12.5+0.213\mathrm{L_{AE}}$, $r=0.99$), the five mean ratings for road traffic ($y=-11.9+0.203\mathrm{L_{AE}}$, $r=0.96$), and the four mean ratings for the intercity trains ($y=-12.3+0.194\mathrm{L_{AE}}$, $r=0.995$). The dose-response relation for the high-speed train (not shown in Fig. 10) is given by $y=-19.5+0.294\mathrm{L_{AE}}$ ($r=0.99$).

With road traffic as the reference, Fig. 10 demonstrates that the bonus for the intercity trains is equal to about 5 dB. In contrast with the results obtained in the conditions which simulated open windows, this bonus is independent of sound level. The difference in annoyance between the two relevant train types can be quantified as well: The types are equally annoying if the ASEL of the maglev-train passby sound is 6 dB lower than that of the intercity train passby sound.

C. Causes of the expected annoyance

As indicated in Sec. II D, the subjects had also been asked about the causes of their expected annoyance. Recall that the five questions were asked only (1) in the blocks in which the sound fragments were rated for the second time, and (2) if the annoyance score was greater than 4. Collecting

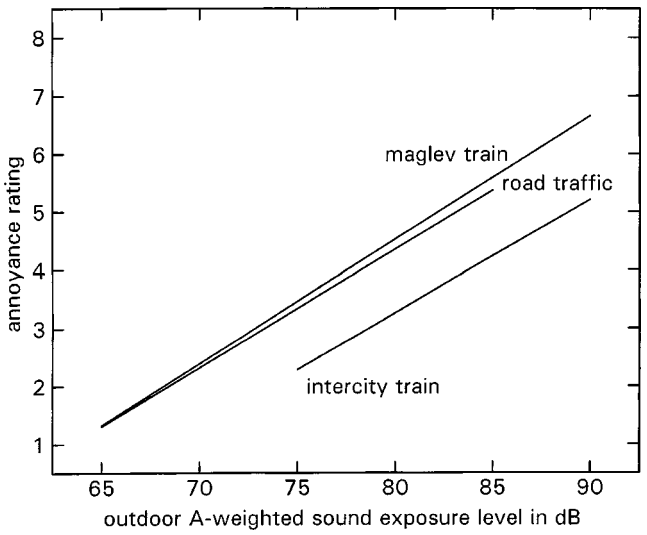


FIG. 10. Dose-response relations for maglev train, road traffic, and intercity train, in the conditions which simulated closed windows. The solid lines are linear regression functions.

information about causes of the annoyance in conditions in which the subjects only expected to be a little or moderately annoyed was considered to be irrelevant.

Overall loudness. For all sound types loudness was selected as a cause of the annoyance. This was most prominent[2] for the sound fragments presented at the higher sound levels (outdoor $L_{AE} \geqslant 75$ dB) in the conditions which simulated open windows. In the conditions which simulated closed windows, a majority of the subjects had selected this cause only for the sound fragments that had been presented at, say, outdoor $L_{AE} \geqslant 85$ dB.

Specific sound characteristics. A heavily pounding sound as a general cause of the annoyance was frequently mentioned only for the high-speed train passby sound presented at the higher sound levels of 85 and 90 dB. This held true both for the windows-open and for the windows-closed conditions. The shrilly sound character of a few passages of the maglev and the high-speed trains as a cause of the annoyance was chosen in the conditions which simulated open windows. In the windows-closed conditions, the shrillness of the sound was no longer an important cause of the annoyance, which can be understood from the effective facade attenuation of the pertinent high-frequency components.

Feeling of insecurity. In the conditions which simulated open windows, the majority of the subjects (with $n > 6$) selected insecurity or unsafety as one of the causes of the annoyance resulting from the sounds of the maglev train (outdoor $L_{AE} \geqslant 85$ dB) passing by at velocities of 325 and 400 km/h. For the same sounds in the conditions which simulated closed windows, the relevance of the feeling of insecurity was much lower.

Startle. Independent of the facade attenuation type, the startle reaction to the sound as a cause of the annoyance was frequently mentioned in the conditions in which the ASEL of the sound of the maglev train passing by at a velocity of 325 or 400 km/h was 85 or 90 dB.

IV. DISCUSSION

A. Contribution of CSEL to the prediction of the annoyance

In studies on the annoyance caused by impulse sounds produced by small, medium-large, and large firearms (Vos, 2001, 2003), it was shown that for conditions which simulated closed windows, the predictability of the indoor annoyance from outdoor ASELs was significantly improved by including the outdoor C-weighted sound exposure level (CSEL; L_{CE}) as a second variable. In these studies the difference between CSEL and ASEL ranged between about 0 and 30 dB.

It is of interest to explore whether for the traffic sounds investigated in the present study, addition of CSEL would result in an increase of the predictability of the annoyance as well. For each sound fragment, the outdoor spectrum was estimated from the spectrum determined at the ears of the subjects in the conditions which simulated open windows (Figs. 2–5), by adding the corresponding frequency-dependent facade attenuation (Fig. 1). These outdoor spectra allowed the calculation of ASEL and CSEL.

TABLE III. Difference between outdoor CSEL and ASEL in decibel, for various experimental conditions.

		Outdoor ASEL (dB)					
Sound source	v (km/h)	65	70	75	80	85	90
Maglev	100	5.9	5.8	5.8	5.5	⋯	⋯
	200	3.9	4.0	4.0	4.2	⋯	⋯
	325	⋯	⋯	3.3	3.3	3.3	3.5
	400	⋯	⋯	3.0	3.1	3.1	3.1
Intercity	120–140	⋯	⋯	1.3	1.0	0.7	0.0
High speed	300	⋯	⋯	0.7	0.7	7.1	7.1
Road traffic	80	8.1	13.8	3.3	4.5	3.5	⋯

Table III shows the differences between CSEL and ASEL for all 29 conditions. For the sounds of the maglev train passing by at speeds of 100, 200, 325, and 400 km/h, the difference $L_{CE} - L_{AE}$ was independent of ASEL and equal to about 6, 4, 3, and 3 dB, respectively. For the sounds of the intercity trains $L_{CE} - L_{AE}$ was equal to about 1 dB. For the high-speed train, $L_{CE} - L_{AE}$ was equal to 1 dB for the sound fragments that were presented at ASELs of 75 and 80 dB, and equal to 7 dB for the other sound fragments. Due to the selective application of the high-pass filter to the first two sound fragments, however, the validity of the differences among these $L_{CE} - L_{AE}$ values must be questioned.

As a result of the small differences between CSEL and ASEL, as found for the sounds from the maglev and intercity trains, it is unlikely or even impossible that addition of CSEL as a second variable would lead to a significant and unequivocally interpretable improvement of the annoyance prediction.[3]

For the road-traffic sounds the differences $L_{CE} - L_{AE}$ ranged between 3 and 14 dB (see Table III). In Sec. II A, it was explained that the sound fragments presented at ASELs of 65 and 70 dB, included passby sounds recorded at a distance of 60 m from the road, and that the sound fragments presented at ASELs of 75 and 80 dB included passby sounds recorded at a distance of 25 m. It should be emphasized that the passby sounds used in the various sound fragments had not been recorded simultaneously. As a result, in addition to recording distance (and sound level), the fragments may also be different with respect to the individual vehicles passing by. Moreover, even the fragments presented at ASELs of 65 and 70 dB did not exactly consist of the same components.

Figure 11 shows that as a result of this procedure, the sound fragment presented at $L_{AE} = 70$ dB contains much more energy in the 63-Hz $\frac{1}{3}$-octave band than the sound fragment presented at $L_{AE} = 75$ dB. This low-frequency energy has a considerable effect on CSEL only. The spectral difference shown in Fig. 11 is caused by inclusion of different trucks. The overall spectral contents of the various passenger cars in the two sound fragments were almost equal.

The somewhat erratic relation between the annoyance ratings and sound level for the road-traffic sounds in the condition which simulated closed windows (Fig. 9) can in part be explained by the corresponding $L_{CE} - L_{AE}$ values. This was confirmed in a multiple linear regression analysis. By adding $L_{CE} - L_{AE}$ as a second predictor of the mean annoyance ratings, the explained variance in the ratings signifi-

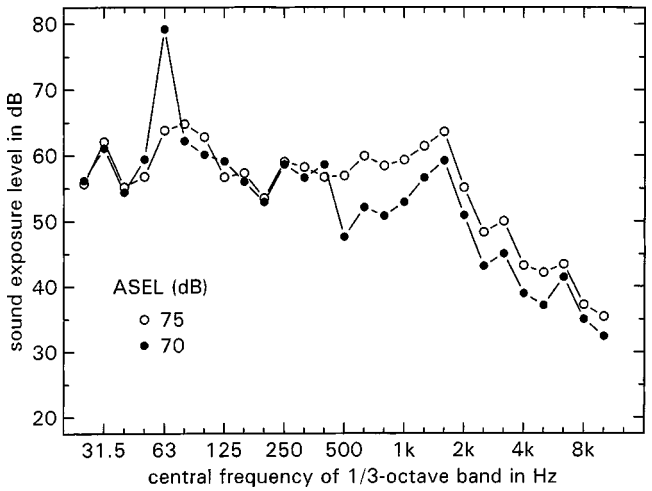


FIG. 11. Sound exposure levels in the various $\frac{1}{3}$-octave bands for road-traffic sounds, as determined at the ears of the subjects in two conditions which simulated open windows. The spectra correspond to overall outdoor ASELs as indicated.

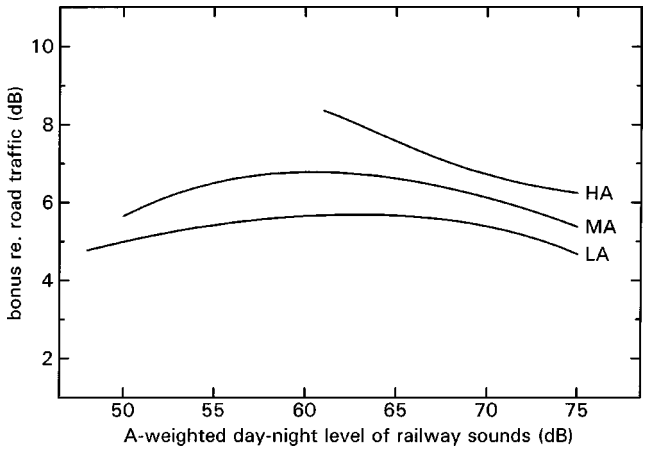


FIG. 12. Railway bonus derived from dose-response relationships reported in Miedema and Oudshoorn (2001), as a function of the day–night level of railway sounds, for three different annoyance measures.

cantly ($p<0.02$) increased from 91.0% to 99.7%. In a similar regression analysis performed on the ratings obtained in the conditions which simulated open windows (Fig. 7), addition of $L_{CE}-L_{AE}$ as a second variable did not result in a significant ($p>0.19$) increase in the predictability of the annoyance. This is consistent with the findings reported in Vos (2001).

B. Railway bonus

Both in the windows-open and in the windows-closed conditions, the annoyance caused by the intercity trains was considerably lower than that caused by road traffic, provided that the ASELs were the same. Averaged across the two facade attenuation types, the bonus for the intercity trains was equal to about 6 dB. Support for a railway bonus has been found in field surveys conducted more than 20 years ago (e.g., see Heimerl and Holzmann, 1979; Schümer-Kohrs *et al.*, 1981; Knall and Schümer, 1983; Fields and Walker, 1982). From the dose-response relationships for road-traffic and railway sounds obtained in a recent meta-analysis (Miedema and Oudshoorn, 2001), it can be revealed that for A-weighted day–night levels between 50 and 70 dB, the railway bonus varies between 5 and 8 dB.

Figure 12 shows the railway bonus as a function of the day–night level of railway sounds for three different annoyance measures given in Miedema and Oudshoorn (2001). The lower curve in Fig. 12 was derived from the relationships with the community response expressed as the percentage of respondents who were at least a little annoyed (LA). The two higher curves in Fig. 12 were derived from the relations with the response expressed as the percentages of respondents who were at least moderately (MA) or at least highly annoyed (HA). It can be concluded that the railway bonus is only slightly affected by the day–night level and the annoyance measure, and that 6 dB is the typical value.

As a result, the bonus for the intercity trains obtained in the present experiment corresponds well with the mean bonus found in field surveys for residential areas with moderate to high exposure levels. It should be emphasized that in contrast with the preliminary loudness data reported in Fastl and Gottschling (1996) and in Gottschling and Fastl (1997), the results of the present experiment do not support application of such a bonus to the maglev-train sounds.

The satisfying correspondence between the railway bonuses obtained in our experiment and in various field surveys supports the validity of the present laboratory study. One might argue that the 6-dB bonus is merely the result of specific features of the experimental method or the sound fragments. One such feature might be the stimulus duration, affecting ASEL of the two sources in a different way. However, it is unlikely that a change in stimulus duration would affect the size of the bonus, as can be understood from the following argument.

For the same density of the road-traffic sounds, a doubling of the stimulus duration results in a 3-dB increase in ASEL. With still one passage of the intercity train, a doubling of the total stimulus duration (i.e., the relevant rating period) has no effect on its ASEL. Relative to the condition with the shorter duration, the annoyance caused by the road-traffic sounds with the longer stimulus duration may be expected to remain the same. The annoyance caused by the railway sound, however, is expected to decrease as a result of the favorably judged large increase of the time period without noise (Vos, 1992a, b; Vos and Geurtsen, 1992, 1995). If this decrease in annoyance is equal to the change in annoyance produced by a 3-dB shift in the sound exposure level of the intercity train, there will be no change in the previous railway bonus at all.

Again for the same density of the road-traffic sounds, halving of the stimulus duration (from 45 to 22.5 s) results in a 3-dB decrease in ASEL. With still a single passage of the intercity train, halving of the total stimulus duration has no effect on its ASEL. Relative to the condition with the longer duration, the annoyance caused by the road-traffic sounds may be expected to remain the same. The annoyance caused by the railway sound, however, is expected to increase as a result of the lack of the favorably rated quiet period. Once more, if this increase in annoyance is equal to the change in annoyance that results from a 3-dB increase in the sound

exposure level of the intercity train, the railway bonus in this hypothetical stimulus configuration will be equal to the bonus obtained in the present experiment.

C. Issues for future research

In Sec. IV B it was shown that there is sufficient support for a bonus for the more conventional railway sounds. Although at present there is little direct evidence of the reasons for this railway bonus (e.g., see Fields and Walker, 1982), the more plausible explanations may have to do with the relatively long quiet periods between passbys, as mentioned in Sec. IV B, and with attitudinal variables such as ''fear'' and ''importance'' (Fields, 1993; Miedema and Vos, 1999). Both a low fear level associated with railways (trains do not fall out of the sky or run into people's houses) and a strong belief that the railways are economically or otherwise important for the local area or some broader community might reduce the annoyance.

In the present laboratory study, the annoyance caused by the maglev-train sounds was higher than the annoyance caused by the sounds from the intercity trains. It might be hypothesized that this effect is at least in part due to the fact that our listeners were not familiar[4] with the maglev train. With the passage of time, residents could potentially develop positive attitudes toward maglev trains, resulting in a decrease of the difference in annoyance between the two train types. For example, information about various measures of precaution might reduce fear and, subsequently, noise annoyance. For exploring the contribution of the nonacoustic factors described above,[5] the present experiment should be enlarged, amongst other things by providing information about the various sound sources and by accompanying the sounds with realistic images of the passing vehicles and trains.

V. CONCLUSIONS

At equal outdoor A-weighted sound exposure levels, (1) the annoyance was virtually independent of the speed of the maglev-train passages, (2) the annoyance caused by the maglev-train passby sounds was hardly different from the annoyance caused by the road-traffic sounds, and (3) the annoyance caused by the intercity train passages was considerably lower than that caused by the maglev-train and road-traffic sounds. These results (4) held true both for open and for closed windows.

Moreover, it was concluded that (5) the sounds might be expected to be equally annoying if the outdoor sound levels of the maglev-train passbys are at least 5 dB lower than those of the intercity train passbys, (6) the differences between the outdoor C-weighted and A-weighted sound exposure levels were too small for expecting a significant increase in the predictability of the annoyance by adding this difference in sound level as a second variable, and (7) in addition to perceived loudness, startle reactions and feelings of insecurity might play a role in the annoyance caused by the sounds of the maglev train.

ACKNOWLEDGMENTS

This research was financed by Project Group Zuyderzeeline (Ministry of Transport, Public Works, and Water Management, and Ministry of Housing, Spatial Planning, and the Environment), The Netherlands. The author is grateful to Keith Wilson for his comments on an earlier draft of this paper.

[1]More information can be obtained on www.dot.gov/affairs/maglevbg.htm and other websites.

[2]For example, restricting ourselves to those conditions in which more than 6 of the 12 subjects had answered the questions, loudness had been indicated as one of the causes of the annoyance in 80%–100% of the cases.

[3]For the total set of 29 conditions, the annoyance increased with increasing outdoor ASEL and increasing $L_{CE}-L_{AE}$. By addition of L_{CE}-L_{AE} as a second predictor of the annoyance, the explained variance slightly but significantly increased by 4 percent points in the condition which simulated open windows, and by 5 percent points in the condition which simulated closed windows. Separate analyses performed on the ratings for the 16 maglev train sounds only, however, showed that in the windows-closed listening mode, addition of $L_{CE}-L_{AE}$ did not lead to a significant improvement of the predictions ($p>0.14$), and in the windows-open listening mode it led to an improvement in the explained variance as small as one percent point ($p<0.02$). Moreover, in both listening modes, the regression weight for $L_{CE}-L_{AE}$ was negative, implying that the annoyance would decrease with increasing $L_{CE}-L_{AE}$, which is in contrast with the previous results.

[4]Our subjects were told that they were exposed to sounds of passing cars and trains. Although they were not explicitly asked to identify the source of the maglev-train sounds, there is no reason to believe that they would not have been able to assign the maglev-train sounds to the category of ''railway-like'' sounds. In spite of the fact that, just as our participants, none of the subjects in the studies reported by Fastl and Gottschling (1996) and by Gottschling and Fastl (1997) had heard the maglev-train sounds before, all of them had identified the sounds as ''train noise.'' Identifying the maglev-train sounds as ''train noise,'' however, does not mean that our subjects were familiar with these sounds to the same degree as they are with the more conventional intercity trains.

[5]More issues for additional research are described in the minutes of the Expert Meeting ''Annoyance caused by maglev trains,'' held on February 27th 2003 at TNO Human Factors in Soesterberg, The Netherlands. For a copy of these minutes, the reader is referred to www.zuiderzeelijn-info.nl.

de Graaff, E., Holties, H., Kurze, U. J., Martner, O., Meißner, F. H., and Nürnberger, H. (**2001**). *Acoustical research on maglev train for the Zuiderzeelijn*, Report No. 50694/2, M+P Raadgevende Ingenieurs, Den Bosch, The Netherlands, and Müller-BBM, Planegg, Germany.

Fastl, H., and Gottschling, G. (**1996**). ''Subjective evaluation of noise immissions from Transrapid,'' in *Proceedings Internoise 1996* (Liverpool, UK), Vol. 4, pp. 2109–2114.

Fields, J. M. (**1993**). ''Effect of personal and situational variables on noise annoyance in residential areas,'' J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M., and Walker, J. G. (**1982**). ''Comparing the relationships between noise level and annoyance in different surveys: a railway vs. aircraft and road traffic comparison,'' J. Sound Vib. **81**, 51–80.

Gottschling, G., and Fastl, H. (**1997**). ''Akustische Simulation von 6-Sektionen-Fahrzeugen des Transrapid,'' [Acoustical simulation of 6-section-vehicles of the Transrapid] in *Fortschritte der Akustik-DAGA 1997* (Bad Honnef, Germany), DPG-GmbH, pp. 254–255.

Heimerl, G., and Holzmann, E. (**1979**). ''Ermittlung der Belästigung durch Verkehrslärm in Abhängigkeit von Verkehrsmittel und Verkehrsdichte in einem Ballungsgebiet (Straßen- und Eisenbahnverkehr),'' [Determination of the annoyance caused by traffic noise as a function of transportation system and traffic density in the area of population concentration (road and railroad traffic)] Kampf dem Lärm **26**, 64–69.

Knall, V., and Schümer, R. (**1983**). ''The differing annoyance levels of rail and road traffic noise,'' J. Sound Vib. **87**, 321–326.

Miedema, H. M. E., and Oudshoorn, C. G. M. (**2001**). ''Annoyance from

transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals,'' Environ. Health Perspect. **109**(4), 409–416.

Miedema, H. M. E., and Vos, H. (**1999**). ''Demographic and attitudinal factors that modify annoyance from transportation noise,'' J. Acoust. Soc. Am. **105**, 3336–3344.

Neugebauer, D., and Ortscheid, J. (**1997**). ''Geräuschbewertung des Transrapid,'' [Noise evaluation of the Transrapid] *in Fortschritte der Akustik-DAGA 1997* (Bad Honnef, Germany), DPG-GmbH, pp. 403–404.

Quehl, J. (**1999**). ''Schienenbonus für Transrapid?—Eine semantische Klanguntersuchung zur Uebertragbarkeit des Schienenbonus auf die Magnetschnellbahn,'' [Railway bonus for Transrapid—A semantic investigation of timbre regarding the application of the railway bonus to the maglev] Z. Lärmbekämpfung **46**(5), 163–174.

Schümer-Kohrs, A., Schümer, R., Knall, V., and Kasubek, W. (**1981**). ''Vergleich der Lästigkeit von Schienen- und Straßenverkehrslärm in städtischen und ländlichen Regionen,'' [Comparison of annoyance by railway and road traffic noise in urban and rural areas] Z. Lärmbekämpfung **28**, 123–130.

Versfeld, N. J., and Vos, J. (**1997**). ''Annoyance caused by sounds of wheeled and tracked vehicles,'' J. Acoust. Soc. Am. **101**(5), Part 1, 2677–2685.

Versfeld, N. J., and Vos, J. (**2002**). ''A-weighted equivalent sound level as a predictor of the annoyance caused by road traffic consisting of various proportions of light and heavy vehicles,'' J. Sound Vib. **253**(2), 389–399.

Vos, J. (**1992a**). ''Noise annoyance around irregularly employed shooting ranges: the expected effect of various training schedules,'' *in Proceedings of the 6th International FASE Congress 1992* (Swiss Acoustical Society, Zürich, Switzerland), pp. 355–358.

Vos, J. (**1992b**). ''Annoyance caused by low frequency sounds from artillery fire: the expected effect of various training schedules,'' J. Low Freq. Noise Vib. **11**(2), 47–51.

Vos, J. (**1995**). ''A review of research on the annoyance caused by impulse sounds produced by small firearms,'' in *Proceedings Internoise 1995*, Newport Beach, CA, Vol. 2, pp. 875–878.

Vos, J. (**2001**). ''On the annoyance caused by impulse sounds produced by small, medium-large, and large firearms,'' J. Acoust. Soc. Am. **109**, 244–253.

Vos, J. (**2003**). ''A- and C-weighted sound levels as predictors of the annoyance caused by shooting sounds, for various facade attenuation types,'' J. Acoust. Soc. Am. **113**, 336–347.

Vos, J., and Geurtsen, F. W. M. (**1992**). *Geluidhinder rond onregelmatig gebruikte schietbanen: een studie naar het verwachte effect van verschillende tijdschema's* [Noise annoyance around irregularly employed shooting ranges: the expected effect of various training schedules], Report IZF 1992 A-13, TNO Human Factors, Soesterberg, The Netherlands.

Vos, J., and Geurtsen, F. W. M. (**1995**). *Geluidhinder rond onregelmatig gebruikte schietbanen: een vervolgstudie naar het verwachte effect van concentratie van schietactiviteiten op de hinder* [Noise annoyance around irregularly employed shooting ranges: a further study on the expected effect of concentration of training activities on annoyance], Report TNO-TM 1995 A-44, TNO Human Factors, Soesterberg, The Netherlands.

Auditory target detection in reverberation[a)]

Patrick M. Zurek[b)]
Sensimetrics Corporation, 48 Grove Street, Somerville, Massachusetts 02144

Richard L. Freyman and Uma Balakrishnan
Department of Communication Disorders, University of Massachusetts, Amherst, Massachusetts 01003

(Received 20 June 2003; revised 23 December 2003; accepted 30 December 2003)

Measurements and theoretical predictions of auditory target detection in simulated reverberant conditions are reported. The target signals were pulsed $\frac{1}{3}$-octave bands of noise and the masker signal was a continuous wideband noise. Target and masker signals were passed through a software simulation of a reverberant room with a rigid sphere modeling a listener's head. The location of the target was fixed while the location of the masker was varied in the simulated room. Degree of reverberation was controlled by varying the uniform acoustic absorption of the simulated room's surfaces. The resulting target and masker signals were presented to the listeners over headphones in monaural-left, monaural-right, or binaural listening modes. Changes in detection performance in the monaural listening modes were largely predictable from the changes in target-to-masker ratio in the target band, but with a few dB of extra masking in reverberation. Binaural detection performance was generally well predicted by applying Durlach's [in *Foundations of Modern Auditory Theory* (Academic, New York, 1972)] equalization-cancellation theory to the direct-plus-reverberant ear signals. Predictions in all cases were based on a statistical description of room acoustics and on acoustic diffraction by a sphere. The success of these detection models in the present well-controlled reverberant conditions suggests that they can be used to incorporate listening mode and source location as factors in speech-intelligibility predictions.
[DOI: 10.1121/1.1650333]

PACS numbers: 43.55.Hy, 43.66.Ba, 43.66.Dc, 43.66.Pn [LRB] Pages: 1609–1620

I. INTRODUCTION

Despite the fact that almost all of our everyday listening takes place against background noise in acoustically reflective environments, there have been no systematic published studies of masked detection in reverberant conditions. Measurements of target detection in reverberation, combined with other stimulus and listening variables, would provide interesting tests of the understanding of auditory detection that has been developed over the years based on studies using anechoic signals. The angular separation between target and masker sources, for example, is an important variable for both monaural and binaural detection (Doll *et al.*, 1992; Gilkey and Good, 1995; Saberi *et al.*, 1991), and the acoustic and psychoacoustic factors underlying the changes in anechoic detection resulting from angular separation are reasonably well understood (Good *et al.*, 1997). Our primary motivation for conducting the present study was to extend the test of such understanding to well-controlled reverberant conditions.

Another motivation for the present study stemmed from the desire to understand and predict speech reception in reverberant conditions. In contrast to the scarcity of work on masking, there have been numerous studies of speech reception in reverberation. These have led to various methods of intelligibility prediction such as the useful-to-detrimental ratio (Lochner and Burger, 1964), the articulation loss of consonants (Peutz, 1971), and the speech transmission index (Houtgast and Steeneken, 1985). Our goal for the present study with respect to intelligibility prediction was to develop a characterization of reverberant target detection that would allow predictive indices to be extended to include effects of listening-mode (monaural-left, monaural-right, or binaural) and source direction, as earlier work did for anechoic conditions (Zurek, 1993). We will report on the application of the present results to intelligibility prediction in a later paper.

Given the lack of prior work on auditory detection in reverberation, we felt it prudent to begin by foregoing real reverberant environments in favor of well-controlled and easily manipulated acoustic simulations of the room and listener. This choice allowed us to explore a range of acoustic conditions, including some that are not easily achieved in real rooms, and to relate masking results carefully to signal properties that result from room reverberation, source location, and acoustic diffraction by a model listener's head.

II. THEORY

Our goal is to formulate a model that relates masked target detection to stimulus, listener, and room-acoustics variables. Because the exact acoustic response at a point in a room is complex and sensitive, in detail, to the specific room-source geometry, it is not feasible to consider a model that takes such a response as its characterization of a room. It is because of this complexity that acousticians have sought a statistical characterization of the important properties of rooms with a few parameters (Beranek, 1954). In this paper we use the results from statistical room acoustics (SRA) to

[a)]Portions of this work were presented at the 141st meeting (June 2001) and 143rd meeting (June 2002) of the Acoustical Society of America.
[b)]Electronic mail: pat@sens.com

describe reverberation characteristics and to test the applicability of the statistical approach to masked target detection in a few simulated room/source configurations. Because we use a controlled simulation, however, we can analyze the actual stimuli generated by the specific simulated configurations. This stimulus-based analysis can also provide predictions for detection in the monaural-left and monaural-right modes.

Several acoustic factors are expected to play major roles in target detection in reverberation. For monaural detection, the primary factor is the change in power of either the target or the masker, or both, that results from changes in source location, room acoustics, or listening with the left ear versus the right ear. Based on a long history of psychoacoustic research showing linear increases in signal threshold with masking noise level (e.g., Hawkins and Stevens, 1950), we expect these power changes to affect detection performance directly. Our preliminary model of detection in the monaural-left and monaural-right listening modes simply and only reflects these changes in target and masker power.

Binaural detection will, of course, also be directly affected by power changes in the targets and maskers at the ears. In addition, binaural detection will be degraded by the decreased interaural coherence of the signals caused by reverberation. Improvements in detection with changes in the interaural parameters of target and masker, called binaural unmasking, have been explored in many detailed psychoacoustic studies [see Durlach and Colburn (1978) for a review]. In order to generate initial predictions of binaural detection in reverberation, we apply Durlach's (1963, 1972) equalization-cancellation theory of binaural unmasking, a model that has had very good success in summarizing a wide variety of data.

A. Predictions for monaural detection based on statistical room acoustics

We need to describe the signals at the two ears in terms of room-acoustic parameters, source locations, and head diffraction effects. To that end, first consider one source in a room with no listener present. The total steady-state acoustic power received at a point d_T m from the source is the sum of the direct power and the reverberant power [see, for example, Beranek (1954)],

$$P\{T,D+R,o\}=G_T(f)\left[\frac{Q_T(f)}{4\pi d_T^2}+\frac{4}{A(f)}\right], \tag{1}$$

where $P\{\cdot\}$ is the power spectrum of the signal specified by its arguments (see footnote 1 for explanation of the notation); $Q_T(f)$ is the directivity factor of the source (calculated with respect to the direction to the receiving point); $A(f)=S\alpha(f)/[1-\alpha(f)]$ is the room constant; S is the surface area of the room; $\alpha(f)$ is the average acoustic absorption coefficient; f is cyclic frequency; and $G_T(f)$ is a factor proportional to source power. (The notation denotes the target source, but the same expression holds for the masker source, with "M" replacing "T"). The first and second terms in the brackets on the right-hand side of Eq. (1) give the relative power levels of the direct sound and the reverberation, respectively. The ratio of these two terms is the direct-to-reverberant ratio:

$$\rho_T(f)=\frac{Q_T A}{16\pi d_T^2}. \tag{2}$$

The direct and reverberant components have different spatial characteristics. The direct component arrives at the receiving point from only one direction, while reverberation results from multiple reflections from surfaces in the environment. Reverberation is therefore assumed to be isotropic.

To find the power levels at a listener's ears, we must take account of diffraction of both the direct and reverberant signal components. If the center-head position is used as the origin of a spherical coordinate system, then diffraction of the direct wave reaching ear j ($j=1,2$) can be described by a head-related transfer function, defined as

$$H_j(f,d_T,\theta_T,\phi_T)=\frac{F_{D,j}(d_T,\theta_T,\phi_T)}{F_{D,o}(d_T)}, \tag{3}$$

where $F_{D,j}$ and $F_{D,o}$ are the Fourier transforms of the signals received at ear j and at the center-head position, respectively, from an anechoic source at distance d_T, azimuth θ_T, and elevation ϕ_T. Note that the normalization by $F_{D,o}$ in Eq. (3) removes the primary effect of distance from the source to the center-head reference point on H_j. The fact that the ears can be at different distances from the source means that H_j will still be dependent on d_T. Writing Eq. (3) in terms of target power spectra gives

$$|H_{j,T}|^2 P\{T,D,o\}=P\{T,D,j\}, \tag{4}$$

where $H_j(f,d_T,\theta_T,\phi_T)$ is abbreviated to $H_{j,T}$. Interaural differences in the direct target wave are found by taking the log magnitude and phase of the ratio $H_{1,T}/H_{2,T}$.

The effect of diffraction on the reverberant sound can be characterized in a similar way. The diffuse-field diffraction, Ψ, of the sound reaching ear j will be

$$\Psi(f)=\frac{P\{T,R,j\}}{P\{T,R,o\}}, \tag{5}$$

the ratio of the power spectrum at ear j to the power spectrum at the center-head position from a perfectly diffuse source. We assume, based on the approximately left–right symmetric positions of the ears on the head, and on the assumed isotropic nature of reverberation, that diffuse-field diffraction is the same for ears 1 and 2.

With the assumption that the D and R components are statistically independent, the total target power at either ear, $P\{T,D+R,j\}$, can be specified in terms of its direct and reverberant components,

$$P\{T,D+R,j\}=P\{T,D,j\}+P\{T,R,j\}, \tag{6}$$

which, by inserting Eqs. (3) and (5), can be rewritten as

$$P\{T,D+R,j\}=P\{T,D,o\}|H_{j,T}|^2+P\{T,R,o\}\Psi. \tag{7}$$

By identifying the reference center-head direct and reverberant target powers, $P\{T,D,o\}$ and $P\{T,R,o\}$, with the respective direct and reverberant components in Eq. (1), we can rewrite Eq. (7) as

$$P\{T,D+R,j\}=G_T\left(\frac{Q_T|H_{j,T}|^2}{4\pi d_T^2}+\frac{4\Psi}{A}\right). \tag{8}$$

The counterpart expression for the total masker power at ear j is

$$P\{M,D+R,j\}=G_M\left(\frac{Q_M|H_{j,M}|^2}{4\pi d_M^2}+\frac{4\Psi}{A}\right). \tag{9}$$

The ratio of total target power [Eq. (8)] to total masker power [Eq. (9)] provides the basis for the SRA prediction of monaural detection performance.

B. Predictions for monaural detection based on stimulus measurements

Predictions for monaural-left and monaural-right detection can also be based directly on the target-to-masker ratios in the stimuli presented to the two ears. Such stimulus-based predictions (called STIM) are computed directly from the digital representations of the target and masker signals after processing through the room/sphere simulation. This simulation is described in detail in Sec. III A.

C. Predictions for binaural detection based on statistical room acoustics and the EC model

Predictions for the binaural listening mode are made by applying Durlach's (1963, 1972) equalization and cancellation (EC) model to the ear signals described by SRA. These predictions will be labeled SRA+EC. The EC model was developed by Durlach to account for the observed improvements in detectability when a binaural target is presented with interaural parameters that are different from those of the masker. A schematic diagram of the EC model is shown in Fig. 1. The inputs to the model are the two ear signals, which are passed through identical peripheral bandpass filters. Although the model can be expanded to multiple bands, only a one-band version is considered here. It is assumed that that band is centered on the target's center frequency, and that the power spectra of the target and masker are both flat within the band. The EC model postulates that the auditory periphery (prior to binaural interaction) introduces amplitude and time-delay errors, ε_j and δ_j, respectively, in its representation of the bandpass signals. It is further assumed that: the amplitude and time errors are independent of one another; errors are independent between ears; and all errors are independent of the stimulus. The time error δ_j is assumed to be a normally distributed random variable with mean equal zero and standard deviation σ_δ; the amplitude error ε_j is assumed to be a random variable (of arbitrary distribution) with a mean equal zero and a standard deviation σ_ε.

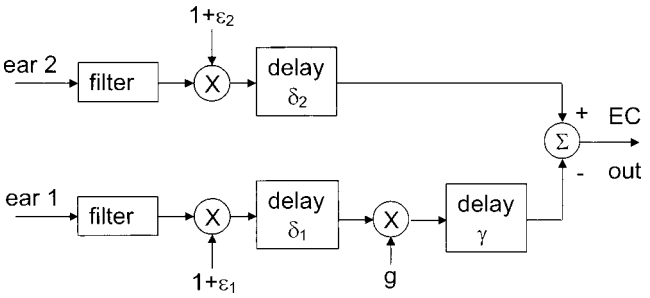


FIG. 1. Block diagram of Durlach's (1972) equalization and cancellation (EC) model of binaural unmasking.

The *equalization* part of the EC model consists of a gain g and a delay γ applied to one of the filtered signals. The *cancellation* part of the model consists of the subtraction of the equalized narrow-band signals from the two sides. In the model as implemented here, the values of the equalizing delay and gain are chosen, without constraints, to maximize the expected target-to-masker ratio after cancellation.

SRA+EC predictions for binaural unmasking are based on the expected target-to-masker ratio out of the EC device. With $F_1(f)$ and $F_2(f)$ being the Fourier transforms of the target signals after bandpass filtering at the two ears, the output of the cancellation device $F_{EC}(f)$ can be written as

$$F_{EC}=(1+\varepsilon_2)F_2e^{-i2\pi f\delta_2}-g(1+\varepsilon_1)F_1e^{-i2\pi f(\delta_1+\gamma)}. \tag{10}$$

The long-term target power out of the EC device is found by taking the expectation of $|F_{EC}|^2$ over both the random errors and the random variations in the stimulus which, after some manipulation, results in

$$P\{T,EC\}=(1+\sigma_\varepsilon^2)[P\{T,2\}+g^2P\{T,1\}] - 2ge^{-(2\pi f\sigma_\delta)^2}\,\mathrm{Re}[C\{T,1\times2\}e^{-i2\pi f\gamma}], \tag{11}$$

where $C\{T,1\times2\}$ is the cross spectrum between the target signals at the two ears. A counterpart expression to Eq. (11) exists for the masker. The target-to-masker ratio out of the EC device, $Z_{EC}(f)$, is then the ratio of those two terms:

$$Z_{EC}=\frac{P\{T,EC\}}{P\{M,EC\}}=\frac{k[P\{T,2\}+g^2P\{T,1\}]-2g\,\mathrm{Re}[C\{T,1\times2\}e^{-i2\pi f\gamma}]}{k[P\{M,2\}+g^2P\{M,1\}]-2g\,\mathrm{Re}[C\{M,1\times2\}e^{-i2\pi f\gamma}]}, \tag{12}$$

where $k=(1+\sigma_\varepsilon^2)e^{(2\pi f\sigma_\delta)^2}$ is a convenient term incorporating the variances of both the amplitude and time errors. A value of $k=1$ implies error-free processing and (potentially) the maximal noise reduction possible for the given stimulus characteristics. As k increases above unity the subtractive terms in both the numerator and denominator on the right-hand side of Eq. (12) have less influence, the main result of which is reduced cancellation of the masker.

Equation (12) can be restated in terms of the reverberant ear signals by replacing the power-spectrum terms (e.g., $P\{T,j\}$) with the total direct+reverberant power spectra (e.g., $P\{T,D+R,j\}$) described in Sec. II A. For the cross-spectrum terms, the assumed independence of the D and R components mentioned above allows us to write

$$\begin{aligned} C\{T,D+R,1\times 2\} &= C\{T,D,1\times 2\}+C\{T,R,1\times 2\} \\ &= H_{1,T}H_{2,T}^{*}P\{T,D,o\} \\ &\quad +\Psi P\{T,R,o\}W(f), \end{aligned} \tag{13}$$

where * indicates complex conjugate and where the cross-spectrum between reverberant ear signals is expressed as the product of the reverberant power at either ear, $\Psi P\{T,R,o\}$, and the normalized correlation, $W(f)$, between reverberant components at the ears. Using Eqs. (8), (9), and (13) in Eq. (12) we obtain

$$Z_{EC}=\frac{kG_T[(Q_T|H_{2,T}|^2/4\pi d_T^2+4\Psi/A)+g^2(Q_T|H_{1,T}|^2/4\pi d_T^2+4\Psi/A)]-2g\ \mathrm{Re}\{G_T(Q_TH_{1,T}H_{2,T}^{*}/4\pi d_T^2+4\Psi W/A)e^{-i2\pi f\gamma}\}}{kG_M[(Q_M|H_{2,M}|^2/4\pi d_M^2+4\Psi/A)+g^2(Q_M|H_{1,M}|^2/4\pi d_M^2+4\Psi/A)]-2g\ \mathrm{Re}\{G_M(Q_MH_{1,M}H_{2,M}^{*}/4\pi d_M^2+4\Psi W/A)e^{-i2\pi f\gamma}\}}. \tag{14}$$

Factoring out $4/A$ and using the expression for direct-to-reverberant ratio in Eq. (2) gives

$$Z_{EC}=\frac{G_T[k(\rho_T|H_{2,T}|^2+\Psi)+kg^2(\rho_T|H_{1,T}|^2+\Psi)-2g\ \mathrm{Re}\{(\rho_TH_{1,T}H_{2,T}^{*}+\Psi W)e^{-i2\pi f\gamma}\}]}{G_M[k(\rho_M|H_{2,M}|^2+\Psi)+kg^2(\rho_M|H_{1,M}|^2+\Psi)-2g\ \mathrm{Re}\{(\rho_MH_{1,M}H_{2,M}^{*}+\Psi W)e^{-i2\pi f\gamma}\}]}. \tag{15}$$

To summarize, Eq. (14) or (15) gives the narrow-band target-to-masker power ratio at the output of the EC device as a function of target and masker locations, SRA parameters, head-diffraction variables, and EC model parameters. SRA+EC prediction for the improvement in binaural detection in a test condition relative to a reference condition is found by taking the ratio of Z_{EC} for the test condition to Z_{EC} for the reference condition, always using maximizing values of g and γ. Note that Eq. (14) or (15) can also be used to generate the same SRA predictions for monaural detection as described above [the ratio of Eq. (8) to Eq. (9)] by setting the cross-ear and non-test-ear terms to zero.

D. Fixed functions and parameters

The expressions given above give target-to-masker ratios in terms of general quantities. Some of these quantities will be experimental parameters of the study described below, while others will be fixed. The study uses a simulation in which a listener is modeled acoustically as a rigid sphere with a radius of 0.0875 m, a choice that determines several of the quantities. In particular, the direct-wave diffraction functions $H_{j,T}$ and $H_{j,M}$ are those described by Morse (1981). The effect of a rigid-sphere diffraction in a diffuse field, Ψ, was described by Waterhouse (1963); this function for the sphere used here was calculated from his Eq. (1.2) and is plotted in Fig. 2. The normalized zero-lag correlation between reverberant signals at the two ears, W, is a function similar to those measured and computed by Lindevald and Benade (1986). The function used here was found computationally for the specific placement of ear positions on the sphere used here. This function is also plotted in Fig. 2.

Other fixed parameters are the source directivities factors and the EC model error parameters. Q_T and Q_M were both made equal 1 to give omnidirectional simulated sources. The EC model's error parameters were chosen to fit the experimental data described below; the specific parameter values were $\sigma_\varepsilon=0.15$ and $\sigma_\delta=85\ \mu$s.

III. METHODS

Behavioral measurements of target detection thresholds were made by presenting normal-hearing listeners with targets and maskers processed through a room/sphere simulation. Experiment 1 was an exploration of target-frequency effects, comparing detection in an anechoic condition with detection in one level of reverberation. In experiments 2 and 3, the target center frequency was fixed at 500 Hz and the degree of acoustic absorption in the room simulation was the main experimental parameter. In all experiments the target

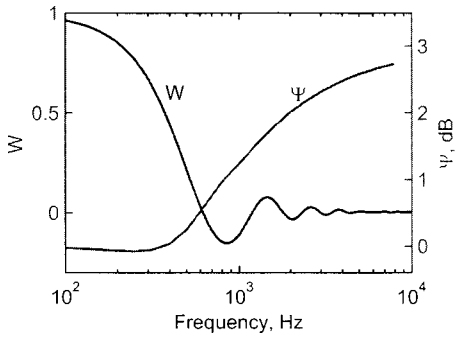


FIG. 2. The normalized interaural correlation function for reverberant sound, W (left axis), and the diffuse-field diffraction function, Ψ (right axis).

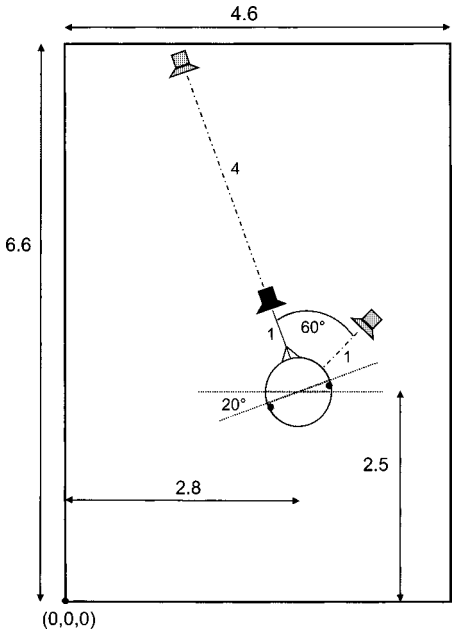


FIG. 3. Schematic plan diagram of the room and sphere. The height of the room was 2.6 m. The sphere center and the sound sources were all located 1.2 m above the floor. The sphere, which had a radius of 8.75 cm, is drawn 4.5 times larger than it would be if drawn to scale.

source was located directly ahead of the sphere at a distance of 1 m in the simulation, while masker location was varied.

A. Simulations

Simulations were used to generate approximations to the target and masker signals that would exist at a listener's ears in a specified acoustic environment. An image-method simulation of sound transmission in a rectangular room (Allen and Berkley, 1979; Peterson, 1986; Desloge, 2002) was used to generate impulse responses from a source at any position to two sensing points on a rigid sphere modeling a listener's head. For all the conditions reported here, the $[x,y,z]$ dimensions of the simulated room were [4.6,6.6,2.6] m (see Fig. 3) and the acoustic absorption coefficient, α, which was an experimental parameter, was uniform and frequency independent on the six room surfaces. The radius of the rigid sphere modeling the head was 8.75 cm and the sensing points modeling the ears were located symmetrically on the left and right sides of the sphere, 10° to the rear of diametrically opposed points in the equatorial plane (Duda and Martens, 1998). The normal bisector of that diameter defines 0°. For all conditions, the target source was located at 0°, 1 m from the sphere's center. The masker source was at one of the three source positions shown in Fig. 3. The center position of the sphere (at [2.8,2.5,1.2]) was chosen to be off-center with respect to all three dimensions of the room, while the 20° orientation was chosen to make the interaural axis parallel with the floor and ceiling (the typical orientation with respect to those surfaces) but not with any of the walls. The simulated target and masker sources had omnidirectional radiation characteristics.

The impulse responses computed from the room simulation incorporated room reflections for 0.5 s following the initial direct impulse (11 025 points at the 22 050-Hz sampling rate used throughout this study). Initial observations indicated that noise processed through such a room simulation possessed an unnaturally regular spectral response, which was manifested perceptually as a ''mechanical'' sound quality. In order to reduce this spectral regularity, a random time jitter (uniformly distributed over ±10%) was introduced to all reflections of order higher than 5. This random jitter was effective in eliminating the objectionable quality.

TABLE I. Acoustic parameters resulting from the values of acoustic absorption α used in the room simulation. Direct-to-reverberant ratios are those at the center position with sphere absent. SRA-predicted ratios are computed from Eq. (2) with $Q_T=1.0$. STIM ratios were computed from the room-simulation impulse response (with sphere absent) filtered through the $\frac{1}{3}$-octave at 500 Hz.

		Direct-to-reverberant ratio (dB) at center position				
α	T_{60} (s)	SRA ($d=1$ m)	STIM ($d=1$ m, $\theta=0°$)	STIM ($d=1$ m, $\theta=60°$)	SRA ($d=4$ m)	STIM ($d=4$ m, $\theta=0°$)
0.10	1.063	−5.8	−5.7	−4.5	−17.8	−17.7
0.15	0.708	−3.8	−4.8	−2.5	−15.8	−14.8
0.25	0.425	−1.0	−1.9	0.8	−13.0	−11.7
0.40	0.266	2.0	1.5	4.0	−10.0	−8.3
0.70	0.152	7.4	6.9	10.8	−4.6	−3.2
1.0	0			∞		

A cross-check between room-acoustic theory (SRA) and the room simulation (STIM) can be made in terms of reverberant power levels or, equivalently, in terms of direct-to-reverberant ratios because the same direct-wave diffraction function is used in SRA predictions and in the simulation. Table I gives center-position direct-to-reverberant ratios measured in the simulation output (STIM) and predicted (SRA) in the $\frac{1}{3}$-octave at 500 Hz for the three source locations and several of the absorption coefficients used in this study. SRA predicts a reverberant power level that depends only on room surface area and absorption and is independent of position in the room. Predicted SRA direct-to-reverberant ratios for this no-sphere condition therefore depend on source distance (by virtue of the effect on direct-wave power) but not direction. Also given in Table I is the reverberation time predicted by SRA, which is $T_{60}=0.16V/(S\alpha)$, where V is room volume. There will be variation in the reverberant sound levels measured across position in a room due to random interference from all contributing reflections. Schroeder (1969) derived the standard deviation σ in the decibel level of reverberant power measured through a band of B Hz to be $\sigma \doteq 5.57/\sqrt{1+BT_{60}/4}$. For $B=110$ Hz at 500 Hz and the range of T_{60} used here, σ ranges from 1 dB (at $\alpha=0.1$) to 2.5 dB (at $\alpha=0.7$). The differences between SRA and STIM reverberant levels do not exceed this expected degree of variation. Therefore, for these simulations with no sphere present, SRA provides a satisfying description of the direct-to-reverberant power ratios measured in the output of the simulation. Further comparisons of SRA and STIM with sphere present will be made below in connection with the detection data.

B. Signals

Listeners were presented with stimuli consisting of a continuous masker to which pulsed targets were added. The

target to be detected was a 300-ms segment of a $\frac{1}{3}$-octave band of noise that was passed through the room/sphere simulation. The masker was continuous broadband noise, also passed through the room/sphere simulation. Tokens of the $\frac{1}{3}$-octave noise targets were generated by creating two independent white noises, low-pass filtering them at $0.1155 \times CF$ (=center frequency), multiplying them by quadrature sinusoids at CF, and summing the resultant noises (Davenport and Root, 1958). An extended length of this noise was divided into 20 individual 300-ms samples and a 50-ms linear rise/fall time was applied to each sample. These noise samples were then convolved with the simulation impulse responses for the target position and allowed to decay for as long as 100 ms beyond the 300-ms nominal duration. A linear 50-ms fall time was applied after 350 ms to insure that all targets decayed to zero by 400 ms. Thus, total target duration ranged from 300 ms in anechoic conditions to 400 ms in reverberant conditions.

Because the target signals were gated before they were passed through the room/sphere simulation, the reverberant power in the target signals increased over the duration of the target. According to SRA, reverberant power builds up (and decays) with an exponential time constant of $T_{60}/13.8$. This build-up of reverberant power had a small effect on overall target power for the room parameters used here. The worst case for the stimuli used here was with $\alpha=0.1$ (see Table I), for which the exponential time constant was 77 ms. For that case, the predicted reverberant target energy in the final 200 ms before input stimulus termination was only 0.5 dB below the steady-state level [200 ms was used as a typical estimate of integration time for detection (Green *et al.* 1957)]. Empirical evidence that this difference is negligible is shown below in comparisons of the SRA predictions with STIM power levels in the actual stimuli.

The other reasonable option for constructing brief reverberant target signals would have been to use tokens of noise in which reverberation was fully built up. We chose to include reverberant build-up because it provides a more-realistic simulation of natural target signals such as speech, which are composed of short-duration segments with varying spectral content.

Different sets of targets and maskers were used in three experiments. Experiment 1 was primarily an exploration of target-frequency effects comparing detection in an anechoic condition with detection in one simulated reverberant environment, modeled on a classroom at the University of Massachusetts. In this experiment the masker was generated by shaping a white noise to produce approximately equal masker levels at target threshold across frequency for the baseline reference condition, which was monaural-left listening with both target and masker at 0°, 1 m in an anechoic room. The amplitude adjustments within the octave bands were determined after preliminary threshold measurements with white noise masking and were applied before processing through the room simulation. The applied gain was −18 dB below 125 Hz, 0 dB at 125 Hz, +6 dB at 250 and 500 Hz, and 0 dB from 1 kHz through 8 kHz.

In experiments 2 and 3, the target center frequency was fixed at 500 Hz and acoustic absorption was the main experimental parameter. In these two experiments the masker was unshaped white noise. In all cases, maskers were generated by convolving 20 s of noise with simulated room/sphere impulse responses. The approximate middle 18 s of the noise was presented repeatedly during the experiment to create the continuous maskers with fully built-up reverberation.

The targets were presented through a D/A converter (TDT DA3) running at 22.05 kHz, attenuated (TDT PA4), mixed with the masker (TDT SM3), low-pass filtered at 8.5 kHz (TDT FT6), routed to a headphone buffer (HB5), and presented through in-phase TDH-49 headphones with MX 41/AR cushions. The masker was presented from a Dell computer integrated sound board and attenuated (TDT PA4), before being mixed with the target.

C. Procedure

Targets in the anechoic condition ($\alpha=1$) were presented at a fixed level of 43 dB SPL (measured with the headphone on a 6-cc coupler). This is also the level of the direct component of the target in both ears in reverberant conditions. Identical attenuator settings were used throughout the experiments, so target levels with $\alpha<1.0$ were higher due to the addition of reverberant energy. The only exception was in experiment 3, where a reverberant target was masked by an anechoic masker and higher-than-usual masker levels were needed to mask the target. In that case, 5 dB of additional target attenuation was applied to avoid uncomfortably loud masker levels.

Masker levels (at the input to the simulation) required to mask the fixed-level targets were estimated using a 4AFC adaptive procedure.[2] Each trial contained four temporal intervals that were marked by lights, with the target presented during one randomly chosen interval. The subject pressed a button corresponding to the interval he/she thought contained the target. Feedback was presented immediately following each trial. The level of the continuous masker was adapted after each trial according to a two-correct-up, one-incorrect-down stepping rule that estimated the 70.7% correct point on the psychometric function (Levitt, 1971), which corresponds to a d' of approximately 1.5 (Macmillan and Creelman, 1991). The initial step size was 8 dB and was halved on the first three reversals of the adaptive track to reach a final step size of 1 dB. The masker levels presented on the last six of ten total reversals were averaged to estimate the masker level required for threshold. If the standard deviation for the six reversals exceeded 2.5 dB, the run was discarded and repeated at a later time.

In experiment 1, three adaptive runs were averaged to estimate the final threshold. In general, thresholds were stable across runs. In cases where the across-run standard deviation exceeded 2.5 dB, additional runs were added until that criterion was met. In experiments 2 and 3, the procedure was modified slightly. Thresholds were estimated from the average of two adaptive tracks unless the across-run standard deviation exceeded 2.0 dB; additional runs were completed as necessary to meet this criterion. On four occasions, an outlier data point was discarded from the threshold computation in order to make the remaining three or four runs meet the criterion. Thus, data points presented for all conditions of

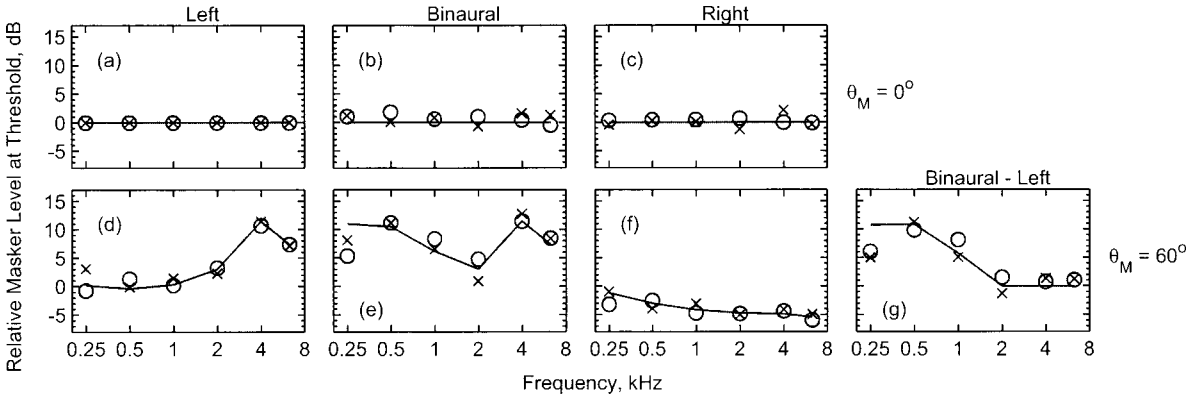


FIG. 4. Relative masker level at threshold for monaural-left, monaural-right, and binaural listening as a function of center frequency of the $\frac{1}{3}$-octave-band target in a simulated anechoic environment. The symbols show measured thresholds for two listeners; the solid lines in the monaural-left and monaural-right listening conditions show SRA predictions. The solid line in the binaural listening panels gives the SRA+EC predictions. All measurements and predictions are expressed relative to the thresholds in the condition in panel (a).

all experiments had across-reversal standard deviations of less than 2.5 dB within each run and across-run standard deviations of less than 2.5 dB.

D. Subjects

Five adults with normal hearing confirmed by audiometry participated in the experiments. Individual subjects' results are differentiated by different symbols in the data plots. The subject whose data are plotted as diamonds was a 36-year-old male; the other four subjects were females aged 22–23 years. Although data from only two subjects were taken in each experimental condition, their results were consistent enough that we felt that running additional subjects was not worthwhile.

IV. RESULTS

All thresholds were normalized to a common reference condition, which was monaural-left listening in an anechoic environment with target and masker sources co-located at 0° and 1 m distance. Each subject's thresholds were normalized by his or her own reference threshold at each frequency. Standard deviations across replicated runs, averaged for each subject over the conditions in an experiment, varied from 0.57 to 1.06 dB. In some cases two threshold measurements were taken and in other cases three or more. Assuming two measurements, a conservative estimate of 1 dB for the standard deviation, and normally distributed error, the 95% confidence interval around the means is ±1.4 dB.

The data plots for monaural detection in Figs. 4–7 that will be described contain SRA predictions (solid lines) from Eqs. (8) and (9), using functions and parameter values as stated above, as well as STIM predictions[3] (dashed lines). The binaural data plots include SRA+EC predictions for binaural detection using EC error parameters of $\sigma_\varepsilon = 0.15$ and $\sigma_\delta = 85\,\mu$s, which were chosen by visual inspection of fits to all the data in Figs. 4–7. These parameter values are comparable to those used in previous work (Durlach, 1972). In all cases, the predictions are based on changes in target-to-masker ratio in the target band from the reference to the test condition.

A. Experiment 1

Figure 4 shows threshold masker level as a function of the center frequency of the $\frac{1}{3}$-octave-band target for monaural-left, monaural-right, and binaural listening in a simulated anechoic environment with target and masker both at a distance of 1 m and with the target located at 0°. Because

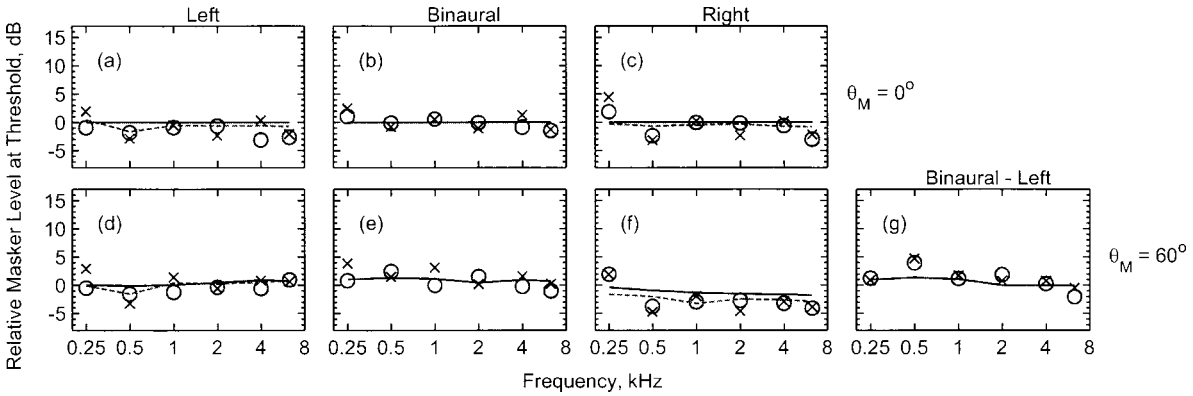


FIG. 5. Relative masker level at threshold for monaural-left, monaural-right, and binaural listening as a function of center frequency of the $\frac{1}{3}$-octave-band target in a simulated reverberant environment. The symbols show measured thresholds for two listeners; the solid and dashed lines in the monaural-left and monaural-right listening conditions show SRA and STIM predictions, respectively. The solid line in the binaural listening panels give the SRA+EC predictions. All measurements and predictions are expressed relative to the thresholds in the condition in Fig. 4(a).

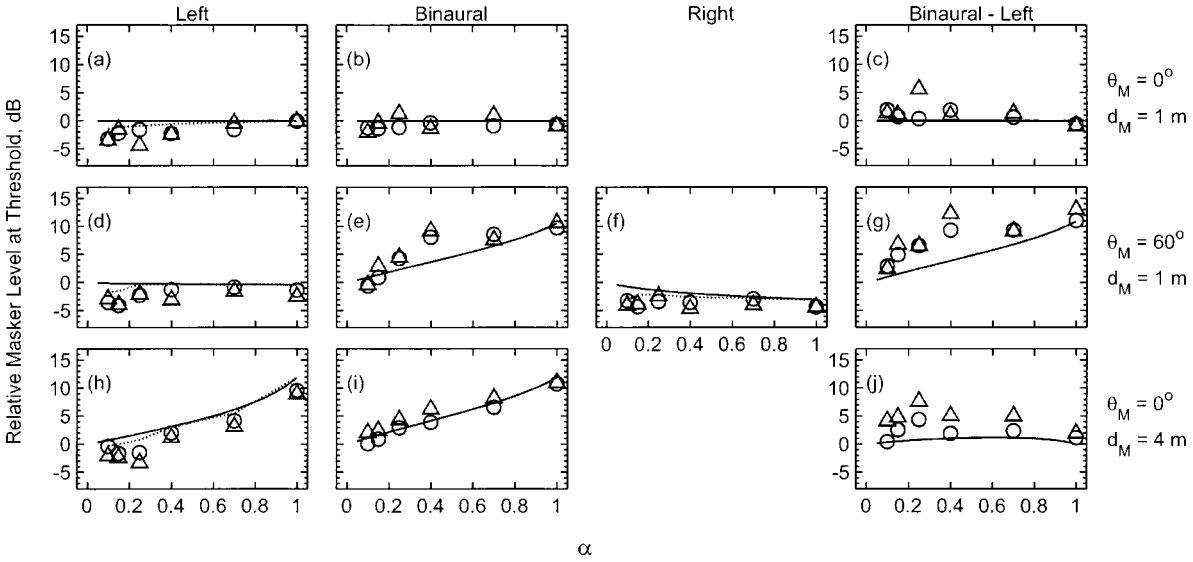


FIG. 6. Relative masker level at threshold as a function of absorption coefficient α. The target frequency was fixed at 500 Hz. The symbols show measured thresholds for two listeners; the solid and dashed lines in the monaural-left and monaural-right listening conditions show SRA and STIM predictions, respectively. The solid line in the binaural listening panels gives the SRA+EC predictions. All measurements and predictions are expressed relative to the 500-Hz thresholds for anechoic, monaural-left detection.

masker level at threshold is plotted in the figures, larger ordinate values signify lower target-to-masker ratios. In this figure, data for two subjects are shown with target direction $\theta_T = 0°$ and with masker direction $\theta_M = 0°$ or 60°. Panel (a)—monaural-left listening with co-located target and masker—represents the reference condition and so the thresholds are 0 dB by definition. Thresholds for the other two listening conditions with $\theta_M = 0°$, monaural-right and binaural, are very close to those measured with monaural-left listening.

The lower panels in Fig. 4 show the changes in threshold in all three listening modes that result from moving the masker 60° to the right. For monaural-left listening the masker is partially shadowed by the sphere, and so detection is improved. Because the shadowing effect increases with frequency, the threshold improvement increases with frequency. An acoustic "illumination" effect is seen in monaural-right listening (panel f) where the right ear receives an increase in masker level (relative to $\theta_M = 0°$ in panel c) as large as 5 dB at high frequencies.

The solid lines in Fig. 4 show the SRA predictions for monaural listening and the SRA+EC predictions for binaural listening. Because the same diffraction functions were used in generating the stimuli and in the SRA predictions, the STIM predictions for anechoic conditions are identical to the SRA predictions.

The threshold pattern seen in Fig. 4(e) for binaural listening presumably reflects both the acoustic effects just mentioned as well as psychoacoustic effects of having both ear signals available simultaneously. The usual interpretations in detection cases such as this are that thresholds at high frequencies (>1 kHz) are improved primarily by the ability to monitor the ear with better target-to-masker ratio, while thresholds at low frequencies (≤1 kHz) are improved prima-

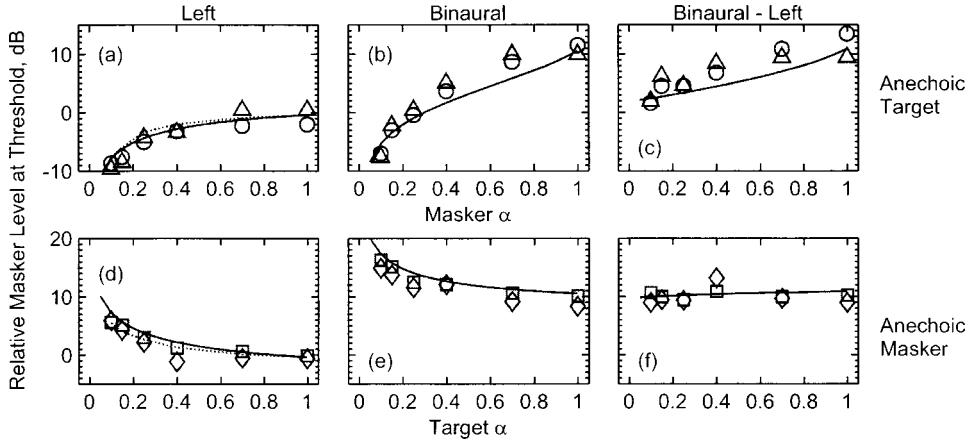


FIG. 7. Relative masker level at threshold as a function of absorption coefficient α. The target frequency was fixed at 500 Hz. In the upper row the target was anechoic and the masker's α was varied. In the lower row the masker was anechoic and α of the target was varied. The symbols show measured thresholds for two listeners; the solid and dashed lines in the monaural-left and monaural-right listening conditions show SRA and STIM predictions, respectively. The solid line in the binaural listening panels gives the SRA+EC predictions. All measurements and predictions are expressed relative to the 500-Hz thresholds for anechoic, monaural-left detection.

rily through binaural unmasking (Zurek, 1993). The former interpretation is easily tested, and confirmed in this case, by noting that the binaural advantage—the difference between binaural and monaural-left threshold—plotted in Fig. 4(g) is approximately zero above 1 kHz. The latter interpretation is supported by the good fit of the EC model predictions at 1 kHz and below, which result principally from binaural interaction. The discrepancy between predictions and data at 250 Hz in Figs. 4(e) and (g) might be attributable to individual listener variability, which is substantial in the low-frequency region for binaural detection tasks [see Fig. 50 of Durlach and Colburn (1978)].

Figure 5 shows data parallel to Fig. 4 but in a reverberant environment simulated with $\alpha=0.14$, which results in $T_{60}=760$ ms and $\rho_T=\rho_M=-4.1$ dB (at 1 m). The primary effect of this degree of reverberation is a reduction in the large changes seen in the anechoic condition from moving the masker to 60°. For monaural listening, the advantage from the acoustic shadow in monaural-left listening is nearly gone, while there is still some degrading effect seen in monaural-right listening.

Figure 5 also shows some deviations of the monaural threshold data from the predictions. Specifically, for both monaural-right listening conditions [panels (c) and (f)], threshold masker level is slightly larger at 250 Hz and slightly smaller at 500 Hz than the predictions. In addition, there is a tendency for monaural-right thresholds with $\theta_M=60°$ [panel (f)] to be lower than the SRA predictions at 500 Hz and above. The STIM predictions, which are 1–2 dB lower than the SRA predictions at all frequencies, provide a better fit in that region.

The large binaural advantages seen with $\theta_M=60°$ under anechoic conditions [Fig. 4(e)] are nearly completely removed under this degree of reverberant [Fig. 5(e)]. The only hint of a binaural advantage in Fig. 5(g) is at 500 Hz, where it is about 4 dB.

The condition for the top row of panels in Fig. 5 has the target and masker co-located at 0°, 1 m in a reverberant room. In the steady state (after reverberation has fully built up), there would be no difference between the SRA and STIM predictions, which would both be 0 dB. The small differences between the SRA and STIM predictions in panels (a) and (c) of Fig. 5 show the effect of incomplete build up of target reverberant power. The predicted value of this difference is about 0.2 dB for $\alpha=0.14$ used here. Obtained differences are comparable to that value except at 500 Hz in the monaural-left condition, where it is about 1.5 dB.

With $\theta_M=60°$, SRA and STIM predictions are also very close for the monaural-left condition, with a discrepancy at 500 Hz similar to that just described. This difference is not necessarily solely attributable to build up of target reverberation in this case because the target and masker are at different locations, so there could be steady-state SRA-STIM differences as well as transient differences. A consistent SRA-STIM difference can be seen across frequency in panel (f), suggesting an effect of room/source/sphere geometry that is not captured by the SRA model supplemented with direct-wave diffraction.

B. Experiment 2

Figure 6 shows masker-level thresholds measured at 500 Hz only, as a function of the absorption coefficient α. The target was again located at $\theta_T=0°$, $d_T=1$ m while the masker was either co-located with the target at $\theta_T=0°$, $d_T=1$ m (top row of panels); shifted in direction only to $\theta_T=60°$, $d_T=1$ m (middle row); or moved in distance only to $\theta_T=0°$, $d_T=4$ m (bottom row). [In the interests of time, thresholds for only one of the monaural modes (left) were measured when sources were symmetrically located with respect to the median plane of the sphere, where differences between the two monaural modes are expected to be negligible.] Results for the monaural listening conditions (panels a, d, f, and h of Fig. 6) indicate that increasing reverberation (decreasing α) leads to a larger degradation in detection than is predicted by either SRA or STIM. Thresholds at $\alpha\leqslant 0.25$ are 2–4 dB worse than predicted by SRA. About 1 or 2 dB of this discrepancy is accounted for by the STIM prediction, but there are still differences between STIM and data as large as 3 dB in some cases.

The results in Fig. 6(b) for binaural listening with $\theta_M=0°$ show no change in threshold relative to the monaural anechoic reference condition. However, comparison of binaural to monaural-left thresholds in Fig. 6(c) shows a small binaural advantage of a few dB at small α's. Binaural listening with $\theta_M=60°$ in Fig. 6(e) shows a strong dependence on α. The direction of this dependence is expected from the fact that the target and masker signals become more spatially diffuse as α decreases, leading to greater similarity in the interaural differences of target and masker that are responsible for binaural unmasking. Note again that the measured binaural advantages shown in panel (g) are clearly larger than predicted by SRA+EC.

The largest discrepancy between binaural thresholds and SRA+EC predictions in this study occurs in Fig. 6(e) with $\alpha=0.4$, where listeners perform about 5 dB better than predicted. This discrepancy was explored in auxiliary measurements on two additional listeners who gave results that fell between the solid line and the results of the original listeners.

Panels (h), (i), and (j) of Fig. 6 show the effect of increasing the sphere-to-masker distance to 4 m while keeping the directions of both the target and the masker fixed at 0°, and the distance to the target fixed at 1 m. This change by a factor of 4 in masker distance is expected from the inverse-square law to require 12 dB greater masker level for anechoic listening ($\alpha=1$). Thresholds in the binaural condition come close to that prediction while the monaural-left thresholds are about 2 dB worse than predicted. The reason for this discrepancy is unknown. At $\alpha=1$ there is no difference between the stimuli for this condition and the reference condition, other than a scaling of the masker. Again, because monaural-left thresholds are worse than predicted by both SRA and STIM, the measured binaural advantages plotted in Fig. 6(j) are larger than predicted.

C. Experiment 3

We next sought to determine whether the effects of reverberation on target detection resulted principally from re-

verberation of the target, the masker, or both, by keeping one source anechoic and varying the reverberation of the other. The data in Fig. 7 show target detection for monaural-left and binaural listening when either the target (upper row) or the masker (lower row) was anechoic and the other was reverberant to varying degrees. [These are not realistic conditions for sources that are at equal distances and that have equal directivity factors; however, such conditions can be approximated with one source (the anechoic one) being very close and/or very directional, and the other source (the one subject to room reverberation) being more distant and/or less directional]. In both cases the masker was positioned at $\theta_T = 60°$, $d_T = 1$ m and the target source, as always, was at $\theta_T = 0°$, $d_T = 1$ m.

Panels (a) and (b) of Fig. 7 show monaural-left and binaural detection thresholds, respectively, for an anechoic target as a function of the absorption coefficient used in generating the masker.

The decline in monaural-left thresholds with decreasing α is the result of the reverberant power increase in the masker. Binaural thresholds also reflect this change in masker power, in addition to a binaural advantage that increases with α. Panel (c) of Fig. 7 plots this binaural advantage, which ranges from 12 dB at $\alpha = 1$ to about 2 dB at $\alpha = 0.1$. Koenig *et al.* (1977), using a similar condition of in-phase (anechoic) target and a masker with very low direct-to-reverberant ratio, measured a binaural advantage of about 3 dB.

The lower panels in Fig. 7 show the reverse configuration, where the masker is anechoic and the degree of reverberation of the target is varied. In this case, it is target power that increases with decreasing α, so monaural-left thresholds [Fig. 7(d)] improve roughly in the same pattern that they declined for the reverse configuration in Fig. 7(a). Binaural thresholds again reflect the change in target power with increasing reverberation, along with a binaural advantage [Fig. 7(f)] that is approximately constant at 10 dB as a function of target reverberation. This result, which may seem surprising at first glance, is accurately predicted by the SRA+EC model. Qualitatively, the model predicts a large and constant binaural advantage in this case because it is the interaural coherence of the masker that is most important for masker cancellation and hence binaural unmasking. Because in this condition the masker is anechoic and perfectly coherent, the degree to which it can be cancelled is independent of target reverberation. The small variation in the SRA+EC prediction with target α results from the change in target power after cancellation, which changes with degree of target reverberation.

V. DISCUSSION

Overall, the results of this study of auditory target detection using well-controlled simulations of reverberant environments are reasonably well predicted by existing models. Monaural detection is largely predictable from the changes in the within-band target-to-masker power ratio caused by a combination of acoustic diffraction effects and reverberation. Binaural detection is predicted well by the combined SRA+EC model, which takes into account the acoustic effects on the signals reaching the ears and also the ability of the binaural auditory system to enhance target detection.

A small but clear discrepancy between the detection data and the theoretical predictions was observed for monaural detection in the most-reverberant conditions (Fig. 6). Monaural thresholds for detecting a $\frac{1}{3}$-octave target masked by a broadband noise were 2–3 dB worse in extreme reverberation than in an anechoic condition. Two possible explanations of this result immediately come to mind. The first is that monaural detection might rely on cross-frequency comparisons of power levels between the target and adjacent masker bands, and that such comparisons are made less reliable by reverberation. The second possible explanation is that detection might be better for targets that are turned on more abruptly, in which case the slower rise of reverberant targets might lead to poorer thresholds. These two hypotheses, which are not mutually exclusive, remain to be tested.

Another noteworthy discrepancy between theory and data was seen with $\alpha = 0.4$ in Fig. 6(e). In that case the two listeners' binaural thresholds were about 5 dB better than those predicted by the SRA+EC model. Subsequent measurements with two additional listeners also exceeded theoretical predictions, but by only 2–3 dB. The masker stimulus with $\alpha = 0.4$ is a mixture of a coherent direct component and an incoherent reverberant component at a direct-to-reverberant ratio (at the center-head position) of 2.0 dB. The EC model's predicted threshold improvement of about 4 dB stems from the fact that at this direct-to-reverberant ratio, approximately 61% of the masker power is coherent. Cancellation of that coherent part gives a 4-dB reduction in noise power. The observation that all four listeners performed better than this prediction, some by as much as 5 dB, discloses a clear shortcoming of the EC model.

Aside from that large discrepancy, all other SRA+EC predictions for binaural detection are quite accurate, even those at $\alpha \leqslant 0.25$ where predicted monaural thresholds deviate clearly from the data. A possible explanation for the superiority of binaural predictions is that different mechanisms operate in monaural and binaural detection. In the former case, as mentioned above, detection may involve cross-frequency comparisons or temporal factors that are not incorporated into our SRA and STIM predictions, which are based only on target-to-masker ratio. Binaural detection, on the other hand, is based on interaural phase and level differences within the target band. Target-induced changes in interaural differences are lessened by reverberation just as they are by other means of decorrelation. The EC model is therefore as successful at predicting reverberation effects as it is at predicting the effects of other forms of decorrelation (Durlach, 1972).

The greater impact of reverberation on monaural than binaural detection thresholds results in sizeable binaural advantages (or monaural disadvantages). One expects that the binaural advantage must go to zero as direct-to-reverberant power ratio approaches zero. However, in the largest degrees of reverberation used here there is still a clear benefit from binaural interaction. The binaural advantages in Figs. 6(g) and (j) remain a few dB at $\alpha \leqslant 0.25$ where center-position direct-to-reverberant ratios (Table I) are all negative. It is

possible that such binaural detection advantages underlie the small binaural advantages in speech intelligibility that have been measured in strongly reverberant conditions (e.g., Nábělek and Pickett, 1974).

The comparisons between SRA and STIM levels that were possible here provided examples of the size of errors involved in applying SRA to a few specific simulated room/source configurations. The errors in predicted reverberation levels were typically less than 2 dB and were always within the expected degree of variation (Schroeder, 1969). For most purposes, the simplification afforded by SRA will be worth the cost of such small errors.

The current simulation of a reverberant room combined with a rigid-sphere model of the listener appears to capture the primary factors affecting both monaural and binaural target detection in reverberation. There will, of course, be discrepancies between the detailed acoustic response of a sphere and actual human heads (and bodies). However, we expect that such differences will be manifested as simple changes in target-to-masker ratio and interaural coherence. The simulation of a rectangular room used here is likewise an extreme simplification of actual listening spaces. But again, we suggest that this simple simulation is adequate to evince the primary factors underlying auditory target detection in reverberation, namely, power changes in the ear signals and decreased coherence between them.

In summary, the results of the present study of auditory target detection in reverberation are reasonably well predicted by existing models of monaural and binaural detection. The accuracy of these models is sufficient for our intended application to predicting speech intelligibility. Accounting for source direction effects will be a generally useful extension to speech-intelligibility-prediction schemes, while the ability to predict monaural and binaural listening-mode effects will be useful in applications to hearing-impaired listeners, with and without hearing aids.

ACKNOWLEDGMENTS

This work was supported by NIH Grant Nos. DC01625 and DC00100. The room/sphere simulation used here was developed originally by P. M. Peterson and was later enhanced by M. P. O'Connell and J. G. Desloge, all at the Communications Biophysics Group at the Massachusetts Institute of Technology. The authors thank Dr. Desloge for making this software available and for modifying it further to correct the problem discussed in the text. The authors also thank Steve Colburn, Nat Durlach, Bruce Henning, Brad Libbey, Barbara Shinn-Cunningham, Les Bernstein, and an anonymous reviewer for their comments on earlier versions of the manuscript.

[1]To reduce confusion among the signal quantities dealt with in this paper a special notation is used. Direct and reverberant components are symbolized with *D* and *R*, respectively; target and masker with *T* and *M*, respectively; the two ears with *1* and *2*; and the center-head reference point with *o*. For example, the Fourier transform of the direct signal at ear 2 would be $F_{D,2}$. Power spectra are denoted by $P\{\ \}$ with the signal whose power spectrum is considered specified in the arguments. $P\{T,R,o\}$, for example, is the power spectrum of the reverberant component of the target at the center-head reference point. Cross spectra are denoted by $C\{\ \}$ with the cross-spectrum signal components coupled by an "×." For example, $C\{T,D\times R,j\}$ is the cross-spectrum between the direct and reverberant target components at ear *j*.

In some cases, signals at the ears or at the output of the EC device are considered without regard to whether they are direct or reverberant and so have the *D* or *R* variable omitted (e.g., F_1, F_{EC} etc; $P\{T,EC\}$ is the power spectrum of a target in the output of the EC device).

Because all signals are expressed in the frequency domain, the frequency variable *f* will be omitted for them; for clarity, however, it will be shown for other quantities when they are first introduced but omitted thereafter.

[2]The target level was held fixed because of the desire to test hearing-impaired listeners eventually with the same stimuli; comparisons to normal-hearing listeners are simplified if the target level is the same for both groups. A concern with an adaptive procedure that controls masker level is that the listener could intentionally err in order to reduce the overall sound level. However, we believe that this was not a factor in our experiments because the sound levels were typically about 65 dB SPL, and never exceeded 75 dB SPL. We feel that with these stimulus levels it is unlikely that the listeners were motivated by loudness discomfort to adopt this strategy.

[3]As mentioned previously, the target stimuli were gated on and off prior to passage through the simulation and so contain reverberant build up and decay. To estimate target power levels we slid a 200-ms analysis window across each target signal and took the maximum level out of that sliding window as the power level.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–951.

Beranek, L. L. (**1954**). *Acoustics* (McGraw–Hill, New York).

Davenport, W. B., and Root, W. L. (**1958**). *An Introduction to the Theory of Random Signals and Noise* (McGRaw–Hill, New York).

Desloge, J. G. (**2002**). Personal communication.

Doll, T. J., Hanna, T. E., and Rissotti, J. S. (**1992**). "Masking in three-dimensional auditory displays," Hum. Factors **34**, 255–265.

Duda, R. O., and Martens, W. L. (**1998**). "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory. Vol. II*, edited by J. Tobias (Academic, New York).

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

Gilkey, R. H., and Good, M. D. (**1995**). "Effects of frequency on free-field masking," Hum. Factors **37**, 835–843.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. H. Anderson (Lawrence Erlbaum, Mahwah, NJ).

Green, D. M., Birdsall, T. G., and Tanner, W. P. (**1957**). "Signal detection as a function of signal intensity and duration," J. Acoust. Soc. Am. **29**, 523–531.

Hawkins, Jr., J. E., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Koenig, A. H., Allen, J. B., Berkley, D. A., and Curtis, T. H. (**1977**). "Determination of masking-level differences in a reverberant environment," J. Acoust. Soc. Am. **61**, 1374–1376.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lindevald, I. M., and Benade, A. H. (**1986**). "Two-ear correlation in the statistical sound field of rooms," J. Acoust. Soc. Am. **80**, 661–664.

Lochner, J. P. A., and Burger, J. F. (**1964**). "The influence of reflections on auditorium acoustics," Sound Vib. **1**, 426–454.

MacMillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, UK).

Morse, P. M. (**1981**). *Vibration and Sound* (American Institute of Physics, New York).

Nábělek, A. K., and Pickett, J. M. (**1974**). "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids," J. Acoust. Soc. Am. **56**, 628–639.

Peterson, P. M. (**1986**). ''Simulation of the impulse response between a single source and multiple, closely-spaced receivers in a reverberant room,'' J. Acoust. Soc. Am. **80**, 1527–1529(L).
Peutz, V. M. A. (**1971**). ''Articulation loss for consonants as a criterion for speech transmission in a room,'' J. Audio Eng. Soc. **19**, 915–919.
Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**). ''Free-field release from masking,'' J. Acoust. Soc. Am. **90**, 1355–1370.
Schroeder, M. R. (**1969**). ''Effect of frequency and space averaging on the transmission responses of multimode media,'' J. Acoust. Soc. Am. **46**, 277–283.
Waterhouse, R. V. (**1963**). ''Diffraction effects in a random sound field,'' J. Acoust. Soc. Am. **35**, 1610–1620.
Zurek, P. M. (**1993**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Needham Heights, MA).

Time-reversal processing for an acoustic communications experiment in a highly reverberant environment

James V. Candy,[a] Alan W. Meyer, Andrew J. Poggio, and Brian L. Guidry
University of California, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, California 94551

(Received 24 May 2002; revised 16 July 2003; accepted 5 December 2003)

Time-reversal (T/R) communications is a new application area motivated by the recent advances in T/R theory. Although perceived by many in signal processing as simply an application of matched-filter theory, a T/R receiver offers an interesting solution to the communications problem for a reverberant channel. In this paper, the performance of various realizations of the T/R receiver for an acoustic communications experiment in air is described along with its associated processing. The experiment is developed to evaluate the performance of point-to-point T/R receivers designed to extract a transmitted information sequence propagating in a highly reverberant environment. It is demonstrated that T/R receivers are capable of extracting the transmitted coded sequence from noisy microphone sensor measurements with zero-symbol error. The processing required to validate these experimental results is discussed. These results are also compared with those produced by an equivalent linear equalizer or inverse filter, which provides the optimal solution when it incorporates all of the reverberations. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1646397]

PACS numbers: 43.60.Dh, 43.28.We, 43.28.Tc [JCB] Pages: 1621–1631

I. INTRODUCTION

In communicating critical information over noisy channels in hostile environments whether they are in a chaotic battlefield, or in an urban environment, or in emergencies evolving from natural disasters like earthquakes, tornadoes or floods, or on the floor of a securities exchange, or even in the clandestine operations of national security—the information sequence must be reliably received and extracted or serious consequences could result. A typical communications problem and system is shown in Figs. 1(a) and 1(b), where we see the host station communicating through a medium to its client stations as well as a potential receiver realization. Here the transmitted information code is subjected to a variety of noise and signal distortions, corrupting the fidelity of the information being transmitted, and reducing the effective capacity of the channel.

Significant effort is continually directed toward researching new techniques for correcting or negating the effects of these disturbances. Time reversal (T/R) is thought to be a known methodology in the communications area, and, unfortunately, it is merely considered by many as just a method to implement the well-known matched filter.[1,2] However, its introduction from the theoretical acoustics viewpoint (see Refs. 3 and 4 and all the references therein) has demonstrated that T/R can provide enhanced signal levels in a reverberant environment especially when implemented with an array of transceivers.

The ever-increasing demand for greater channel capacity continually necessitates better solutions for channel equalization. Equalization is defined as the removal of channel distortion to extract the transmitted information. The current state of the art in channel equalization either requires *a priori* knowledge of the channel or the use of a known training sequence and adaptive filtering.[5,6] Most current solutions to the equalization problem require a detailed channel model that typically does not include a multipath. If the assumed model within the processor does not at least capture the dominant characteristics of the channel, then the received information is still highly distorted and possibly useless. Developments in T/R signal processing indicate that the potential for compensating the transmission channel while mitigating the need for detailed *a priori* knowledge of the channel characteristics is possible.[7–20] The T/R receiver does not require an explicit model for its realizations. It implicitly compensates for multipath distortion by using the equivalent of an imbedded phase conjugation technique for the equalization. This is the primary theoretical property of T/R that can be exploited for the point-to-point communications problem. A successful development of a T/R receiver can increase channel bandwidth, thereby enabling the proportional increase in the volume of information.

From a signal processing perspective, T/R processing appears to be an application of *matched filtering* in which the output signal-to-noise ratio (SNR) is maximized. This T/R replicant is then cross-correlated with the received signal to produce the optimal filtered output.[1,2] However, it becomes more complicated in the spatiotemporal case in which the optimal matched filter must not only match the transmitted temporal function, but also the corresponding spatiotemporal channel medium impulse response or *Green's function*. Contemporary research in spatial division multiple access (SDMA) depends on the spatiotemporal Green's function. Consequently, T/R becomes the optimal implementation for SDMA.[21,22] It has been shown that time-reversal techniques are applicable to spatiotemporal phenomena that satisfy a wave-type equation possessing the time reversal invariance property.[3,4] Time reversal is the dynamic *broadband* analog

[a]Electronic mail: candy1@llnl.gov

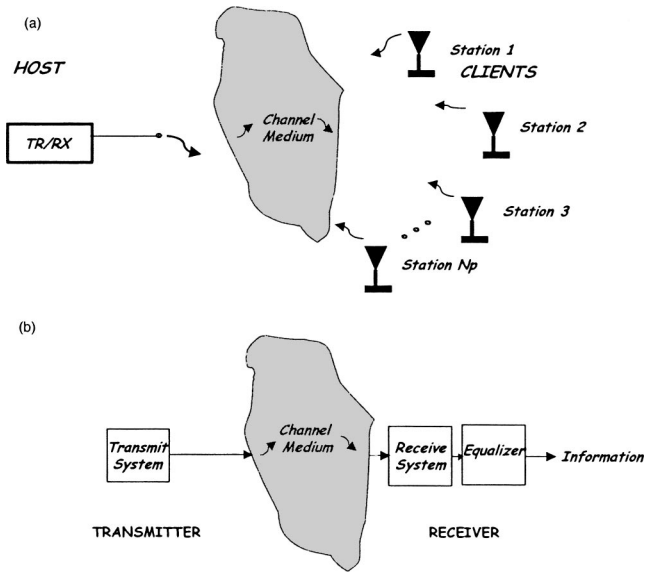


FIG. 1. Typical communications channel with transmitter/receivers. (a) Physical environment. (b) Communications system with receiver/equalizer.

of the well-known phase conjugate mirror (optics) used to focus *narrowband* monochromatic waves. It represents the ''optimal'' spatiotemporal matched filter in the sense of maximizing the output SNR. It is essentially a technique that can be used to ''remove'' the aberrations created by an inhomogeneous or random channel. In communications, it is an open question of whether or not a T/R receiver can overcome the inherent uncertainty created by the medium providing the enhancement required to extract the transmitted information sequence. Although the array aspects of T/R are ignored in this paper by considering only point-to-point communications, it is clear that the future of wireless transceivers will incorporate small sensor arrays to take advantage of spatial diversity enabling the T/R approach to provide a potential solution.[21,22]

Perhaps the initial idea of T/R communications evolved from the seminal work of Parvulescu[7] (in 1967 and published in 1995 due to security issues). Here the underlying Green's function of the ocean is first estimated using a pilot or probe pulse, time reversed and retransmitted through the medium to focus and achieve a high SNR gain. The realization of this receiver is captured by that of *T/R Receiver I* described in the next section. Following this inaugural work, Jackson[8] developed the formal theory of T/R (phase conjugation) in underwater acoustics (again *T/R Receiver I*), while the work of Dowling[9] was aimed at developing receivers capable of compressing coded acoustic pulses in a fading multipath channel. Here the receiver realization is equivalent to post-processing the received data with the estimated Green's function (*T/R Receiver III*). Model-based methods developed by Hermand[10] and Brienzo[11] followed by eliminating the pilot and using modeled Green's functions (*T/R Receiver III*). Subsequent experiments to demonstrate the performance of T/R in the ocean channel followed with the work of Kuperman[12] (*T/R Receiver I*), Hermand[13] (*T/R Receiver III*), Hodgkiss[14] and Edelmann[15] (*T/R Receiver I*). Recently Rouseff[16] proposed a passive T/R approach in underwater acoustic communications using a realization equivalent to *T/R Receiver IV*. In a recent focusing study Yon[17] implemented the *T/R Receiver I* for acoustics in a room. Noncoherent communications studied by Smith,[18] Heinemann[19] in underwater acoustics using a T/R receiver realization (*T/R Receiver I*) demonstrates the effectiveness of T/R communications in a slowly (temporal) changing environment. In the ultrasonic regime Derode[20] discusses the communications problem in terms of focusing and transmitting codes to receivers (*T/R Receiver I*). The work we present in this paper unifies these designs and introduces another possible realization, *T/R Receiver II*. The feasibility and performance of these various T/R receiver realizations in a noisy, highly reverberant environment is investigated and compared to an optimal equalizer in this paper.

First the realizations of various T/R receivers are discussed and applied to noisy microphone measurements in a highly reverberant environment. The development of various processing techniques using T/R receivers to extract the transmitted information sequence and evaluate the performance on real data is discussed and compared to the performance of an equivalent linear equalization receiver. In Sec. II, the underlying T/R theory relative to the communications problem is briefly discussed along with a brief development of T/R receivers, the optimal linear equalizer and other essential components of an acoustics communications system. The development of an experiment to assess the feasibility and performance of the T/R receivers is described in Sec. III along with the associated signal processing. The receiver performance on the raw measurement data to extract the information sequence is discussed in Sec. IV as well as the underlying processing. Finally, we summarize these results and discuss future efforts.

II. T/R COMMUNICATIONS

In this section we develop a suite of ''point-to-point'' time-reversal receivers to recover a transmitted information sequence or code from a set of receiver measurements in a highly reverberant, temporally stationary environment. Here the emphasis is on the processing provided by the various realizations to extract the information signal and characterize their performance. It should also be noted that no attempt is made to optimize the receiver and configure it in a ''real-time'' operational mode. Rather, the experiment is performed to demonstrate the feasibility of using ''time reversal'' to extract the information sequence.

A. Time-reversal background

The detection of a transmitted information sequence can be transformed to the problem of maximizing the output signal-to-noise ratio, $\mathrm{SNR}_{\mathrm{out}}$, at the receiver of a communications system. The underlying system model for the communications problem is given by

$$z(t)=s_{\mathrm{out}}(t)+n_{\mathrm{out}}(t)=g(r;t)*[s(t)+n(t)], \tag{1}$$

for $z(t)$, the noisy measurement, $s_{\mathrm{out}}(t)$, the output signal consisting of the convolution of $s(t)$, the transmitted signal (information) and $g(r;t)$, the spatiotemporal channel response. The output noise, $n_{\mathrm{out}}(t)$, is also the convolution with the input noise, $n(t)$, an additive random (white) zero

mean, noise of variance, σ_n^2. The matched-filter problem based on the model of Eq. (1) is *given* a "known" signal, $s(t)$, in additive white noise, *find* the filter response, $f(t)$, that maximizes the $\mathrm{SNR}_{\mathrm{out}}$, defined by

$$\max_f SNR_{\mathrm{out}} \equiv \frac{\xi_{s_{\mathrm{out}}}}{E\{n_{\mathrm{out}}^2(t)\}} = \frac{|f(t)*s(t)|^2}{\sigma_{n_{\mathrm{out}}}^2}, \tag{2}$$

for $\xi_{s_{\mathrm{out}}}$ defined as the output signal energy. The solution to this problem is classical and reduces to applying the Schwartz inequality[1,2] to the numerator, that is,

$$|f(t)*s(t)|^2 \leqslant \xi_f \times \xi_s, \tag{3}$$

for ξ. the energy of $f(t)$ and $s(t)$. When $f(t)$ is related to $s(t)$ by a constant, say unity, then this relation is satisfied with equality at some time T,

$$f(t) = s(T-t), \tag{4}$$

the reversed, shifted signal or replicant. The matched-filtering operation, $mf(t)$, is then

$$mf(t) = E\{f(t)*z(t)\} = E\{s(T-t)*z(t)\} = C_{sz}(T-t), \tag{5}$$

where C_{sz} is the cross-correlation function of the known signal, $s(t)$, and the measurement, $z(t)$.

For time reversal, the matched-filter problem is identical to that posed above with a "known" Green's function of the medium replacing the known signal.[3,4,7] The Green's function, $g(\mathbf{r},\mathbf{r}_0;t)$, is the result of a point-to-point communications link between a *host* station (source) at $\mathbf{r}_0$ to a *client* station (receiver) at $\mathbf{r}$. In the T/R case, the matched-filter solution is again found by maximizing, $\mathrm{SNR}_{\mathrm{out}}$, leading to the modified numerator,

$$|f(t)*g(\mathbf{r},\mathbf{r}_0;t)|^2 \leqslant \xi_f \times \xi_g, \tag{6}$$

that is satisfied with equality at some time T, if

$$f(t) = g(\mathbf{r},\mathbf{r}_0;T-t). \tag{7}$$

Thus, for T/R, the optimal matched-filter solution is the time-reversed Green's function from the host station-to-client station (source-to-receiver) or *vice versa*. Comparing these results with the usual matched-filter solution above, the Green's function of the channel is reversed rather than the transmitted replicant signal. Note that since T/R theory requires reciprocity,[3,4] the result of Eq. (6) is valid for *both* transmission and reception, that is, $g(\mathbf{r},\mathbf{r}_0;T-t) \leftrightarrow g(\mathbf{r}_0,\mathbf{r};T-t)$. Note also that if an array is included to sample the spatial field or transmit a wave, then these results include the focus at client station (receiver) position, $\mathbf{r}$, yielding the optimal, *spatiotemporal* matched-filter solution,[2] $g(\mathbf{r}_\diagup,\mathbf{r}_0;T-t)$ at sensor position, $\mathbf{r}_\diagup$.

B. Time-reversal receivers

In this section, a set of realizations of the T/R receiver is developed that can be implemented to solve the point-to-point communications problem. The T/R operation can be performed either on *transmission* or *reception* using the channel Green's function or a known pilot signal leading to four potential realizations. These realizations are shown in Figs. 2 and 3.

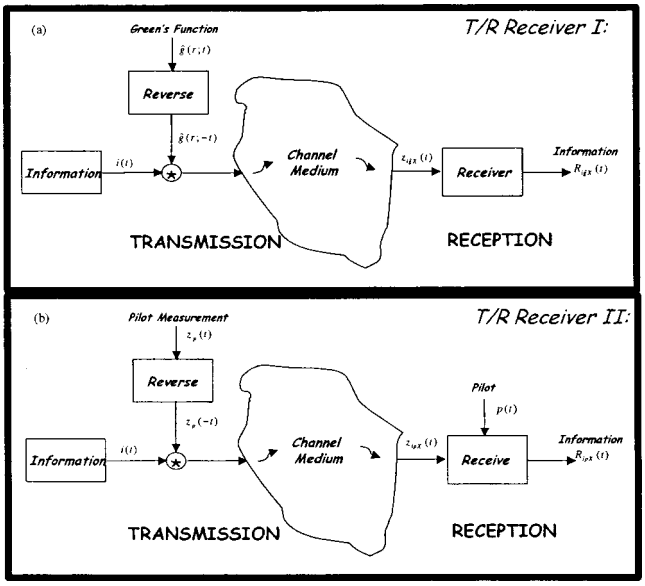


FIG. 2. T/R transmission-based receivers: (a) T/R receiver I: estimated Green's function and reversal on *transmission*. (b) T/R receiver II: pilot measurement and reversal on *transmission*.

First, we define the following measurements: the *pilot* and the *information* to proceed with our development. The *pilot measurement* is defined as

$$z_p(t) \equiv g(\mathbf{r};t)*p(t), \quad \text{(pilot measurement)}, \tag{8}$$

where z_p is the pilot measurement, which can also be contaminated with additive noise; g is the spatiotemporal Green's function of the channel; and p is the known, transmitted pilot signal. The purpose of the pilot in this case, as well as that of equalization,[1,2] is to estimate the Green's

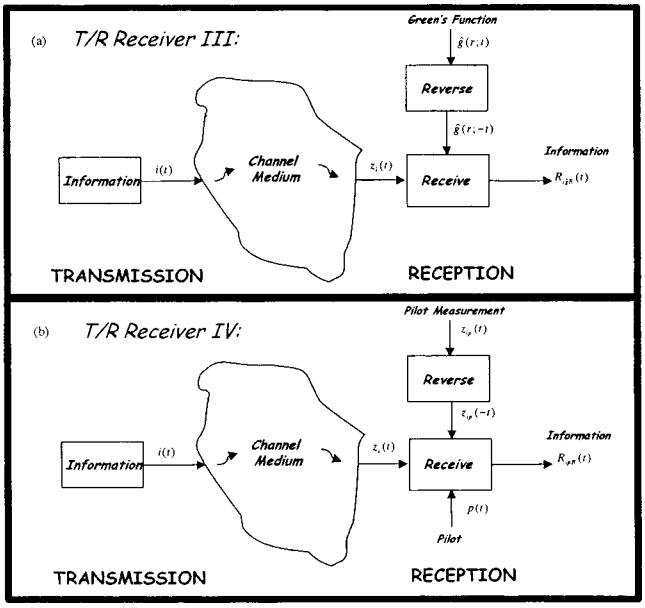


FIG. 3. T/R reception-based receivers: (a) T/R receiver III: estimated Green's function and reversal on *reception*. (b) T/R receiver IV; pilot measurement and reversal on *reception*.

function, $\hat{g}(\mathbf{r};t)$, which is used in the final receiver design to mitigate the distortion effects created by the channel medium and unknown transfer characteristics of the measurement system. The pilot is not only used to estimate the Green's function, but also to implement the receiver by convolution [see Fig. 2(b)] on reception, as discussed below.
The *information measurement* is defined by

$$z_i(t) \equiv g(\mathbf{r};t) * i(t) \quad \text{(information measurement)}, \tag{9}$$

where z_i is the information measurement; and i is the transmitted information signal.

Realizations of the T/R receiver on *transmission* follow directly from linear systems theory and we consider the following two cases: one using the Green's function of the channel medium, and one using the pilot measurement. The first realization we develop is probably the most advantageous by performing T/R during transmission thereby mitigating the distorting effects of the medium directly.

1. T/R receiver I: Estimated Green's function ($\hat{g}$) T/R receiver realization on TRANSMISSION (Refs. 7, 8, 12, 14, 15, 17–20)

The first realization of a point-to-point T/R receiver is shown in Fig. 2(a) in which the Green's function of the channel medium is first estimated (see the discussion in Sec. II C), reversed, and convolved with the information signal on transmission. On reception, the input to the T/R receiver is defined as $z_{i\hat{g}X}(t)$, with the subscript representing the information signal (i), convolved with the estimated Green's function ($\hat{g}$), on transmission (X). Therefore, we have

$$\begin{aligned} z_{i\hat{g}X}(t) &\equiv z_i(t) * \hat{g}(\mathbf{r};-t) = g(\mathbf{r};t) * \hat{g}(\mathbf{r};-t) * i(t) \\ &= C_{g\hat{g}}(t) * i(t), \end{aligned} \tag{10}$$

where $C_{g\hat{g}}(t)$ is the cross-correlation of $g(\mathbf{r};t)$ with $\hat{g}(\mathbf{r};t)$. The output of the T/R receiver is simply

$$R_{i\hat{g}X}(t) = z_{i\hat{g}X}(t). \tag{11}$$

Clearly, if $C_{g\hat{g}}(t) \approx C_{gg}(t)$ and $C_{gg}(t)$ is impulsive, then $i(t)$ is recovered directly by this realization. However, in reality, we have that the autocorrelation function, $C_{g\hat{g}}(t)$, acts as a window function filtering or smearing $i(t)$. It is also interesting to note in this particular realization that if a sensor array replaces the single sensor transmitter used for point-to-point communications, then the optimal spatial–temporal matched filter is additionally achieved with the added gain in SNR afforded by the array as well as the focusing capability of the T/R operation.[3,4] Next we consider a second realization using a similar structure.

2. T/R receiver II: Pilot (p) measurement (z_p) T/R receiver realization on TRANSMISSION

An alternate realization of the T/R receiver is to use the pilot measurement, as shown in Fig. 2(b). Here a known pilot signal is selected to excite the channel medium providing $z_p(t)$, which is reversed and convolved with the information signal on transmission to generate, $z_{ipX}(t)$, as input to the T/R receiver defined by

$$z_{ipX}(t) = z_i(t) * z_p(-t) = C_{gg}(t) * p(-t) * i(t). \tag{12}$$

Convolution with the known pilot signal on reception, $p(t)$, is therefore necessary to recover the information at the output of the receiver given by $R_{ipX}(t)$, that is,

$$R_{ipX}(t) = z_{ipX}(t) * p(t) = C_{gg}(t) * C_{pp}(t) * i(t), \tag{13}$$

where the $C_{\cdot}$ are the respective autocorrelations of g and p. [For completeness, another possible interpretation of the receiver output is given by $R_{ipX}(t) = C_{gp}(t) * C_{gp}(-t) * i(t)$, which follows directly from the commutative property of the convolution operation.] Note that if both C_{gg} and C_{pp} are impulsive, then $i(t)$ is recovered directly. Realistically, we can think of the information signal as being filtered or smeared by both correlations.

Next, we consider the realizations of the T/R receiver on reception. Again we have two cases to consider: one with the Green's function and one with the pilot measurement realizations.

3. T/R receiver III: Green's function ($\hat{g}$) T/R receiver realization on RECEPTION (Refs. 9–11, 13)

On reception, (R), the estimated Green's function is reversed and convolved with the receiver input, $z_i(t)$, to give

$$R_{i\hat{g}R}(t) = z_i(t) * \hat{g}(\mathbf{r};-t) = C_{g\hat{g}}(t) * i(t), \tag{14}$$

which is mathematically identical to $R_{i\hat{g}X}(t)$ of Eq. (11), with the exception that the mitigation of the channel medium is performed on reception rather than transmission. The structure of the T/R receiver for this realization is shown in Fig. 3(a). Finally, we consider the case of using the pilot measurement on reception.

4. T/R receiver IV: Reversed pilot (p) and measurement (z_p) T/R receiver realization on RECEPTION (Ref. 16)

As in the previous realization, we start with the receiver input and convolve it with the reversed pilot measurement as well as the known pilot signal to generate the receiver output, $R_{ipR}(t)$, where

$$R_{ipR}(t) = z_i(t) * z_p(-t) * p(t) = C_{gg}(t) * C_{pp}(t) * i(t), \tag{15}$$

which is mathematically equivalent to $R_{ipX}(t)$ of Eq. (13). The realization of this receiver is shown in Fig. 3(b).

C. Channel medium Green's function estimation

As we discussed in Sec. II B, the channel Green's function is an integral part of the two T/R receiver realizations. It can be estimated from the pilot measurement of Eq. (8) and is similar to the operations used for equalization,[1,2] but is much better conditioned numerically for solution, since the forward, $g(\mathbf{r};t)$, rather than the inverse, $g^{-1}(\mathbf{r};t)$, is required for T/R. The estimated Green's function is used in the realizations to mitigate the distortion effects created by the channel medium and unknown transfer characteristics of the measurement system.

The estimate, $\hat{g}(\mathbf{r};t)$, can be obtained using the *optimal* Wiener solution obtained from the pilot measurement by solving the minimum mean-squared error (MSE) problem

$$\min_{g} J = E\{\varepsilon^2\}, \quad \text{for } \varepsilon(t) \equiv z(t) - \hat{z}(t)$$

$$= z(t) - \hat{g}(\mathbf{r};t) * p(t), \tag{16}$$

and ε is defined as the error. The solution of this problem leads to the well-known Wiener filter[6] given by

$$\hat{\mathbf{g}} = \mathbf{C}_{pp}^{-1}\mathbf{c}_{zp}, \tag{17}$$

where $\mathbf{C}_{pp}$ is an $M \times M$ correlation matrix and $\mathbf{c}_{zp}$ is a $M \times 1$ cross-correlation vector. The estimated or filtered measurement is

$$\hat{z}(t) = \hat{g}(\mathbf{r};t) * p(t) = \sum_{m=0}^{M-1} \hat{g}(\mathbf{r};m)p(t-m) = \hat{\mathbf{g}}^T\mathbf{p}(t), \tag{18}$$

for

$$\hat{\mathbf{g}}^T \equiv [\hat{g}(\mathbf{r};0) \cdots \hat{g}(\mathbf{r};M-1)] \quad \text{and}$$

$$\mathbf{p}^T(t) \equiv [p(t) \cdots p(t-M+1)].$$

Since the correlation matrix of Eq. (17) is Toeplitz,[6] we choose to use the Levinson–Wiggins–Robinson (LWR) recursion to efficiently perform the inversion (order M^2).

Another approach to the Green's function estimation is to utilize the fact that the autocorrelation of a signal is the convolution with itself time reversed, $[C_{xx}(t) = x(t) * x(-t)]$. Therefore, if we simply convolve the pilot measurement, $z_p(t)$, with the reversed pilot, that is,

$$\hat{g}(\mathbf{r};t) = z_p(t) * p(-t) = [g(\mathbf{r};t) * p(t)] * p(-t)$$

$$= g(\mathbf{r};t) * C_{pp}(t), \tag{19}$$

then

$$\hat{g}(\mathbf{r};t) \approx g(\mathbf{r};t), \quad \text{for } C_{pp}(t) \to \delta(t).$$

We call this the *reversed pilot signal* approach to Green's function estimation. Note that one candidate pilot signal is a chirp that approximately satisfies the impulsive-like autocorrelation function property.

D. Linear equalization receiver

Before leaving this section, we briefly discuss the development of a linear equalizer[1,2] (LE). Using the receiver model of Eq. (1) with the time-reversed Green's function replaced by an equalization filter, we have that

$$\hat{i}(t) = z_i(t) * h_{\mathrm{EQ}}(t) = [g(\mathbf{r};t) * h_{\mathrm{EQ}}(t)] * i(t). \tag{20}$$

For direct recovery of the transmitted coded information sequence, the equalizer must be an inverse filter, that is,

$$\hat{i}(t) = [g(\mathbf{r};t) * \hat{g}^{-1}(\mathbf{r};t)] * i(t) \approx i(t), \quad \text{for}$$

$$h_{\mathrm{EQ}}(t) = \hat{g}^{-1}(\mathbf{r};t). \tag{21}$$

Note that this implementation provides an optimal solution (in theory) to the equalization problem as long as the estimated Green's function contains all of the medium information (multipath, reverberation, inhomogeneities, etc.). However, in a highly reverberant environment, the size or order of the filter (number of weights) is prohibitive, limiting its realizability as well as the fact that the inverse problem is ill-conditioned in general.[23] We will ignore these constraints in this paper and estimate an inverse filter using all of the reverberations to provide a *bound* for the best we could hope to achieve.

Typical algorithms for linear equalizers in communications are implemented using a variety of adaptive signal processing algorithms ranging from the popular least mean square (LMS) technique to the more computationally intensive recursive least squares (RLS) approach.[5,6] The main reason for selecting adaptive processing is that the channel is assumed to be random and time varying. In this problem, it is assumed that the channel is *not* varying in time, since our application is a stairwell. Therefore, we further assume that the underlying statistics are stationary. It is well known that for stationary processes adaptive processing techniques converge to the Wiener solution of Eq. (17) with the inputs and outputs switched to obtain the inverse rather than the forward impulse response, that is, we obtain the equalization filter, given by

$$\hat{\mathbf{g}}^{-1} = \mathbf{C}_{zz}^{-1}\mathbf{c}_{pz}, \tag{22}$$

using the LWR algorithm as before [see Eq. (17)].

Thus, both the T/R and equalizer-based receivers must estimate the channel medium or inverse channel medium Green's functions, respectively, using a pilot or probe excitation prior to processing the transmitted information data.

III. SIGNAL PROCESSING FOR THE T/R COMMUNICATION EXPERIMENT

In this section we describe the experimental environment and processing to gather the receiver data. We lay out the steps required to extract the coded information sequence using the various T/R receiver realizations and the optimal LE. The experiments are performed in a stairway located between two floors consisting of three landings, high ceilings of corrugated steel, pipes, handrails, and other nonsound absorbing protrusions as well as ambient building noise. The geometry of this stairway is shown in Fig. 4, clearly indicating the potential for a highly reverberant environment.

The point-to-point experiments were performed using a point-like B&K 4296 sound source powered by a B&K 2716 amplifier and an Analogic 2020 arbitrary waveform generator for transmitting both the pilot chirp pulse swept from 0.1 to 2 kHz and the BPSK modulated code of 0.1 kHz bandwidth at a carrier frequency of 1.207 kHz. On reception, a B&K BP-2716 microphone and amplifier were used with a flat frequency response in our bands of interest along with an 8-bit LeCroy digitizer sampling at 10 kHz. The experiment was controlled using a laptop computer. A typical transmitted code received on the microphone with the corresponding spectrum is shown in Fig. 5. Here we observe the noisy transmitted coded message signal and carrier in Fig. 5(a) and the raw received microphone data in Fig. 5(b) with corresponding spectra in Fig. 5(c). Note that the response is dominated by a long reverberation response and noise—an ideal environment to test feasibility and the performance of the T/R approach.

All of the receiver realizations incorporate common functions such as synchronization, demodulation, and quan-

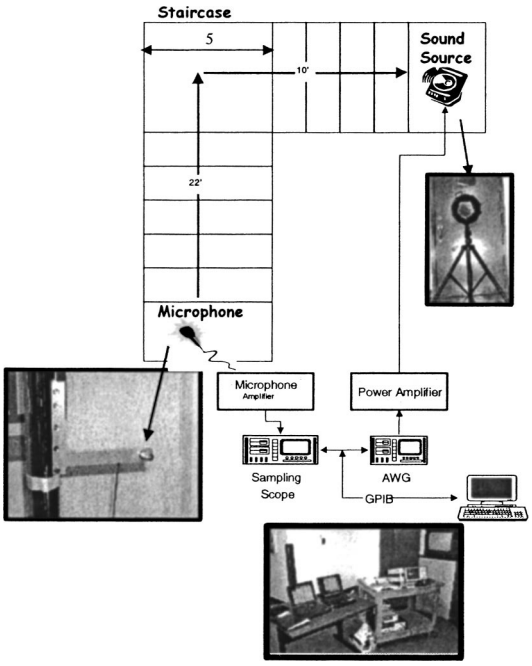


FIG. 4. Experimental environment and setup for T/R communications testing: stairway with three landings and high corrugated steel ceilings (10–12 ft) along with the equipment setup for point-to-point communications.

tization to extract the transmitted codes. Each of these operations is required for the type of coding and modulation scheme selected for this experiment.

Synchronization is the process of aligning and locally generating the extracted code between the receiver and the transmitter.[1,2] We transmit binary phase-shift keyed (BPSK) codes using the double sideband suppressed carrier (DSBSC) amplitude modulation (AM) technique.[1,2] The processed T/R

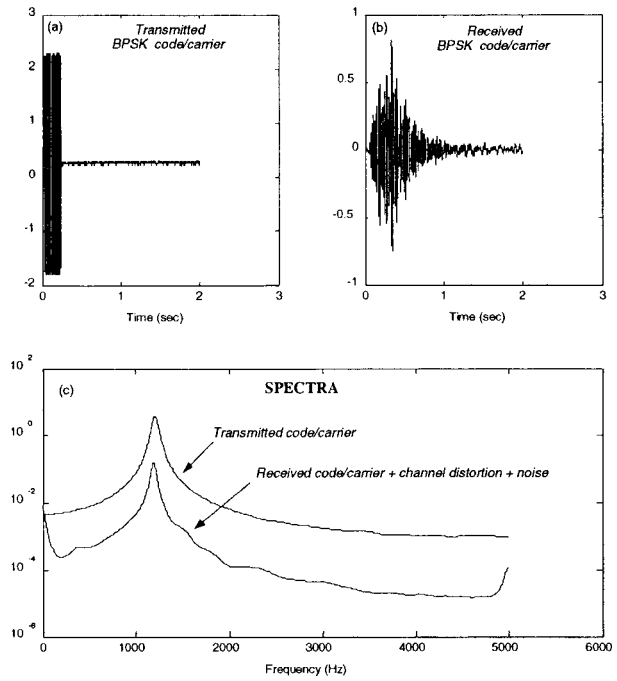


FIG. 5. Transmitted code sequence and stairwell response: (a) Transmitted BPSK coded sequence with carrier. (b) Received BPSK coded sequence with carrier (stairwell response). (c) Raw transmitted and received spectra.

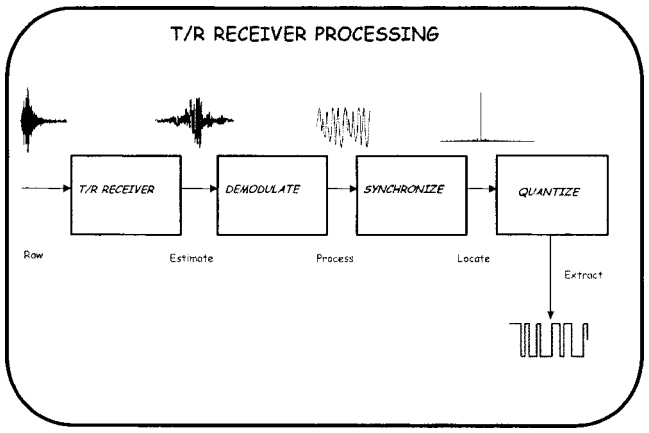


FIG. 6. Signal processing of experimental data: T/R reception (estimate code), demodulation (process), synchronization (locate code), quantization (extract code).

signal is then demodulated using an adaptive phase locked loop to align the carrier for demodulation.

As discussed in Sec. III A, the matched filter is the optimal processor to maximize the output SNR. If it is desired to optimally *detect* the transmitted information, then a matched-filter receiver of Eq. (5) can be constructed such that the cross-correlation of $i(t)$ and its estimate at the output of a T/R receiver $\hat{i}(t)$ achieve a maximum at some lag T, that is,

$$\hat{T}=\max C_{i\hat{i}}(k-T)|_{k=T}=C_{i\hat{i}}(0). \tag{23}$$

Note that $\hat{i}(t)$ can be any of the T/R receiver realizations depicted in Figs. 2 and 3. Therefore, in essence, all that is necessary is to "detect" that the transmitted information code has been recovered. This detection is accomplished by performing the matched-filter calculation above and locating the well-defined peak at $\hat{T}$. The known pilot signal can be used to estimate $\hat{T}$. It should also be noted that extracting the actual information sequence, as in the case of voice communications, requires the solution of the problem of estimating, $\hat{i}(t)$, from the reverberant transmission through the medium, which is the desired output of the T/R receiver.

The final processing step for any of these receivers is to *quantize* the estimated and detected BPSK coded information sequence. The *quantizer* simply limits the estimated coded signal to a above or below a specified threshold, that is,

$$i(t)=\begin{cases} +A, & \hat{i}(t)>\tau, \\ -A, & \hat{i}(t)\leqslant\tau, \end{cases} \tag{24}$$

where A is the quantization level or amplitude and τ is the preset threshold.

The overall processing is shown in Fig. 6, where we see that the *raw* signal is received by the T/R receiver and processed using one of the realizations discussed in the previous section to produce the *estimated* (modulated) code sequence. The data are then *demodulated* to remove the carrier providing the processed data for synchronization. Note that this step incorporates a phase-locked loop[1,2] to align the carrier phase of the receiver with the transmitted, perform the demodulation (multiplication), and low-pass filtering. A stan-

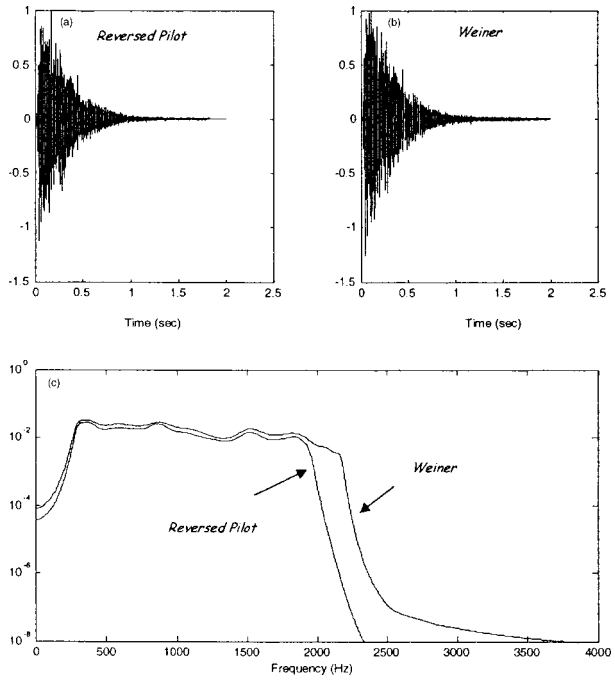


FIG. 7. Average Green's function estimation based on an ensemble of 30 measurements each: (a) Reversed pilot signal solution using chirp. (b) Optimal Wiener solution. (c) Spectra.

dard matched filter is then used to synchronize the processed data and temporally locate the onset of the code that is then quantized, extracted, and compared to the transmitted code for performance analysis.

Recall that to implement the T/R receivers I and III we must estimate the channel (stairway) Green's functions. We investigated the two approaches discussed previously: (1) the *optimal* Wiener solution; and the (2) *reversed pilot signal* convolution/correlation technique using a chirp pilot. The optimal solution is computationally demanding while the reversed pilot solution is simple, but suboptimal. The results are depicted in Fig. 7, where the performance of both methods are shown. From a statistical viewpoint, we ran 30 realizations of each measurement, estimated the underlying Green's functions and averaged them to obtain the curves shown in the figure. This approach enabled us to generate ensemble statistics to evaluate the quality of our measurements and make improvements as required. In Figs. 7(a) and 7(b) we see the average estimated Green's functions, $\hat{g}(\mathbf{r};t)$, along with their corresponding spectra in Fig. 7(c). It is clear that both spectra are almost identical in the transmission bands of interest and therefore we used the reversed pilot approach for our estimates in implementing the receivers.

The basic experimental approach is summarized as follows.

(1) *Excite* the medium with a chirp pilot sequence to estimate the channel medium (Green's function).
(2) *Transmit* the modulated information signal from the source (speaker) along with the required Green's function, pilot, etc. through the reverberant medium (stairway) to the microphone sensor.
(3) *Receive* (microphone) the noisy, reverberant sequence and digitize.

TABLE I. Experimental processing of T/R receiver data.

1.	*Estimate* the channel medium Green's function, $\hat{g}(\mathbf{r};t)$ of Eq. (17) or (19);
2.	*Process* the raw measurement data on (transmission) reception using the T/R receivers of Eqs. (10)–(15) to obtain the respective $R_\cdot(t)$;
3.	*Demodulate* the T/R receiver output, $R_\cdot(t)$;
4.	*Detect* the transmitted code using the matched-filter, $C_{i\hat{i}}(k-T)$ of Eq. (5) to locate its onset, $\hat{T}$, for synchronization;
5.	*Extract* (cut) the detected information sequence, $\hat{i}(\hat{T}-t)$; and
6.	*Quantize* to obtain the final estimate of the coded information sequence $\hat{i}(t)$ of Eq. (22).

(4) *Process* the raw data with the T/R receiver (software).
(5) *Demodulate* the processed data to estimate the information.
(6) *Locate the temporal onset of the code extracting it from the demodulated data*.
(7) *Quantize* the results for a further performance analysis.

This completes the description of the experimental setup, equipment, and processing. Next we discuss results.

IV. RESULTS

In this section we discuss the results and performance analysis of the T/R communication receivers and linear equalizer on an ensemble consisting of 30 time series per task. *All* of the plotted results are averaged over the ensemble. The typical processing procedure we follow in the experiment (Fig. 6) is shown in Table I.

As discussed previously, we estimated the Green's functions of the stairwell using the reversed pilot approach shown in Fig. 7 and demonstrated that the medium could support the frequency band between 0.1 to 2 kHz. We designed the BPSK code to occupy a bandwidth of 0.1 kHz centered at 1.207 kHz, as shown in Fig. 5. For this feasibility study, we chose 100 samples/symbol for high fidelity symbol data. With this information in mind, we developed an experimental computer simulation for design and implementation of the various receivers for processing and ran them over the 30 member ensembles.

It should also be noted that due to an experimental limitation of our arbitrary waveform generator, only 4000 samples were available for *transmission* into the medium including the samples created by convolution operations on transmission. The estimated Green's function response did not settle down until approximately 10 000 samples; therefore, we had to choose a subset of 4000 samples for receiver designs on transmission. We chose to precondition the transmitted code using a simulation of the experiment by performing the estimated Green's function convolution operation and then selecting the ''best'' 4000 sample set to find any that would satisfy the symbol error criterion. Once the best 4000 samples were selected, they were transmitted into

the stairwell and the T/R receiver output [Eq. (10)] was, in fact, the estimated code for this realization.

A. Performance analysis

We will use a *symbol error criterion* to evaluate the performance of each of the receiver realizations in this paper. *Symbol error* is defined as the percentage of symbols missed over the total transmitted. In our application, since a bit is represented by one symbol, bit error and symbol error are synonymous.

For our application we can vary the threshold of the quantizer [see Eq. (24)] over the processed T/R receiver output (after demodulation) and generate a performance curve defined by the percentage symbol error versus the threshold. This curve plays a role similar to the operating characteristic curve[24] in classical detection theory—thresholds can be selected to minimize symbol error. If zero-symbol error can be achieved by any particular threshold, then the receiver under investigation can recover the code perfectly. The range of the thresholds yielding zero-symbol error gives an indication of the performance capability of the receiver. The threshold range corresponding to zero-symbol error indicates the robustness of the receiver, that is, the larger the range, the more threshold values that can be selected to yield zero-symbol error. It is interesting to note that just as in classical detection theory,[24] the value of the threshold is also selected for the receivers based on some performance criterion (e.g., the Neyman–Pearson criterion leads to a fixed false alarm probability). Next we discuss the performance of the receivers and subsequently compare them to one another.

B. T/R receiver I performance

The realization of this T/R receiver uses the estimated Green's function convolved with the code *transmitted* into the stairwell medium [Fig. 2(a)]. This is one of the most common realizations applied in the literature.[7,8,12,14,15,17–20] After demodulation and synchronization, we see the total output (ensemble average) of the T/R receiver [Fig. 8(a)] and the corresponding code estimates produced on both simulated and estimated data [Fig. 8(b)]. It is interesting to note that the prediction of the estimate from the simulator, which is developed using the estimated Green's functions, closely matches that of the actual experiment. This code estimate is then quantized based on a selected threshold. Each estimated symbol (○) is then compared to the true transmitted symbol (∗), as shown in Fig. 8(b). For the selected threshold, the T/R receiver is able to achieve a zero-symbol error on the simulated data. However, in the figure, we show an estimate where the *threshold* was *not* optimally selected to achieve zero-symbol error (misses one symbol). After simulation and selection of the appropriate 4000 samples from the Green's function/information sequence convolution, zero-symbol error was achieved, as shown by the simulated curve in the figure. The overall performance of this T/R realization is shown later in Fig. 13 by the (···) line and compared to each receiver in the bar chart of Fig. 14.

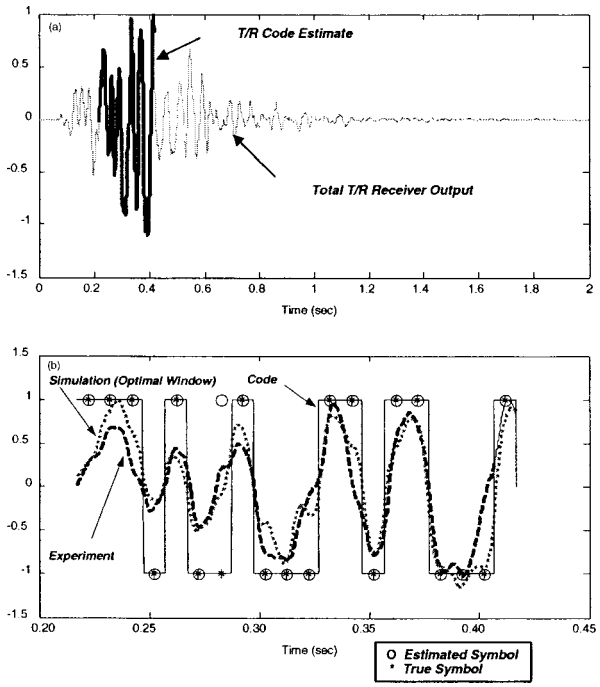


FIG. 8. T/R receiver I output: (a) T/R receiver output and code estimate. (b) True code, extracted code estimate, simulated code estimate using optimal 4000 sample window, true symbol (∗) and estimated symbol (○).

C. T/R receiver II performance

The realization of this T/R receiver uses the pilot measurement convolved with the code *transmitting* into the medium and then convolves the *received* output with the known pilot to recover the code [Fig. 2(b)]. We show the results of this realization in Fig. 9. Note again that these plots are ensemble averages. Again, after processing the raw data in Fig. 9(a), the receiver demodulates and detects the transmitted

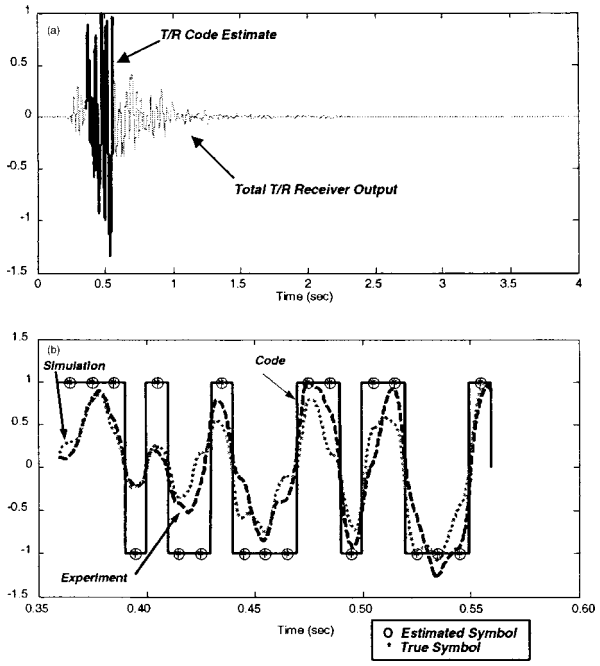


FIG. 9. T/R receiver II output: (a) T/R receiver output and code estimate. (b) True code, extracted code estimate, simulated code estimate, true symbol (∗), and estimated symbol (○).

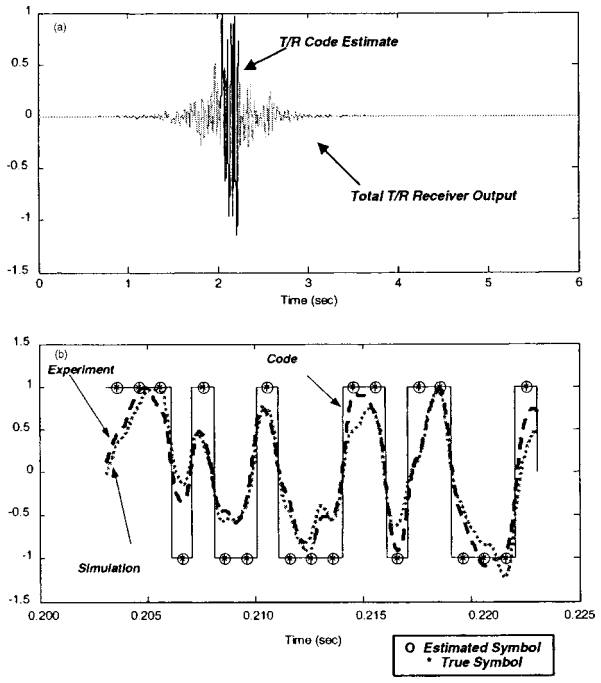


FIG. 10. T/R receiver III output: (a) T/R receiver output and code estimate. (b) True code, extracted code estimate, simulated code estimate, true symbol (∗), and estimated symbol (○).

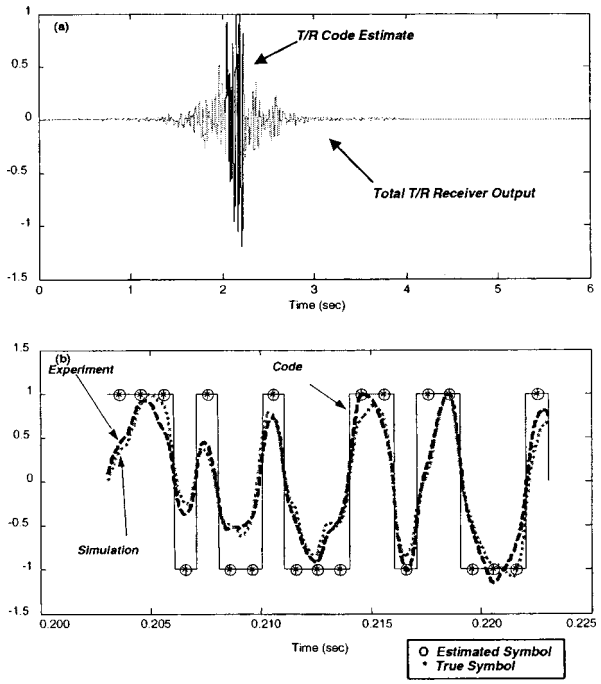


FIG. 11. T/R receiver IV output: (a) T/R receiver output and code estimate. (b) True code, extracted code estimate, simulated code estimate, true symbol (∗), and estimated symbol (○).

code. As shown in Fig. 9(b), both estimated and simulated results are in reasonable agreement. After quantization, the transmitted code can be recovered with zero-symbol error depending on the choice of threshold. The overall performance is shown later in Fig. 13 by the (-··-) line and compared to the other realizations in Fig. 14.

D. T/R receiver III performance

This T/R receiver uses the estimated Green's function convolved with the raw received code data [Fig. 10(a)] on *reception*. The performance results of this realization are shown by the ensemble plots in Fig. 10(b). After the usual processing, the experimental and simulated results track each other. For the selected threshold, a zero symbol is achieved with perfect code recovery. The overall performance is shown later in Fig. 13 by the (---) line and compared in Fig. 14.

E. T/R receiver IV performance

The final T/R receiver realization uses both the measured pilot medium response and known pilot convolved with the transmitted code measurement on *reception* for recovery. The results are again quite reasonable, as shown in Fig. 11, with the raw data in Fig. 11(a) and extracted code information in Fig. 11(b) as before. Both estimates of the code are consistent with the true transmitted code, as shown in Fig. 11(b). After quantization, the recovered code again achieves a zero-symbol error for the chosen threshold. Performance comparisons are available later in Fig. 13 by identifying the (-···-) line and Fig. 14.

F. LE receiver V performance

The T/R receivers can be compared to the optimal (MSE sense) linear equalizer design on *reception*. Recall that the LE is designed using the entire Green's function including all reverberations, etc. coupled to the Wiener MSE estimator of the inverse Green's function [Eq. (21)]. After tedious tuning of the LE [Fig. 12(a)], the code is extracted as before with

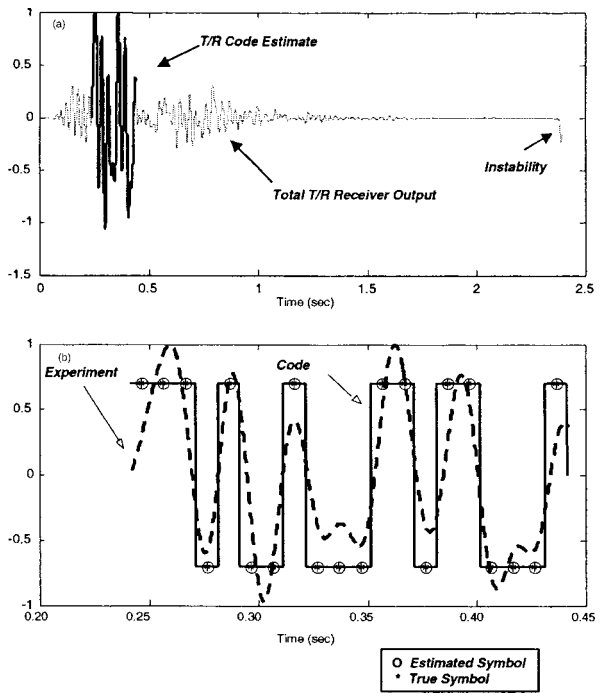


FIG. 12. LE receiver V output: (a) LE receiver output and code estimate. (b) True code, extracted code estimate, simulated code estimate, true symbol (∗), and estimated symbol (○).

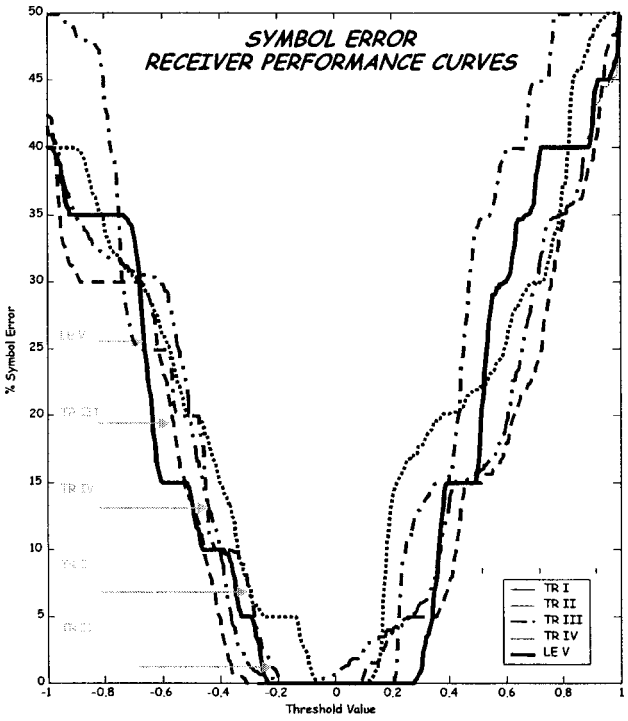


FIG. 13. Receiver performance curves: T/R receiver realizations I–IV and optimal linear equalizer V receivers.

zero-symbol error for the selected threshold [Fig. 12(b)]. The overall performance is shown in Fig. 13 by the solid (—) line. It is clear from the figure that the inverse filter has captured the reverberative Green's function providing the optimal solution, as indicated by the threshold range in the bar chart of Fig. 14. Unfortunately, the instability in solving the inverse problem inhibits the practical use of this approach in a timely manner.

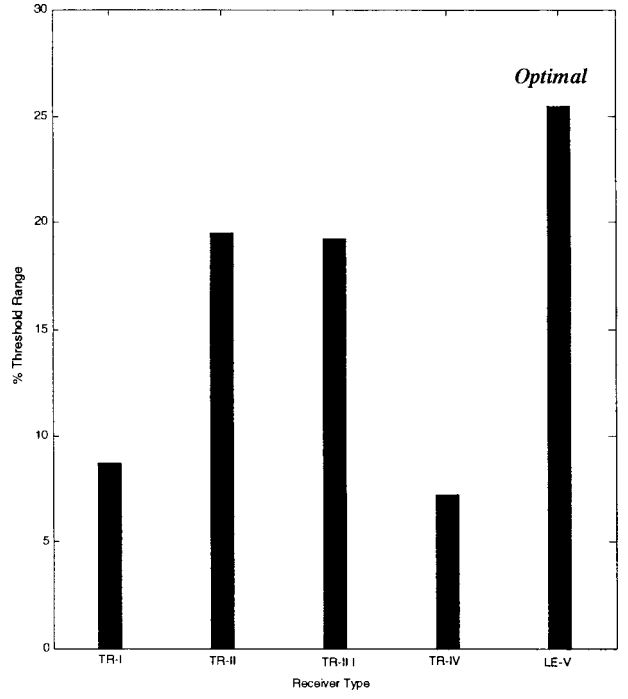


FIG. 14. T/R and LE receiver performance based on the threshold range for zero-symbol error performance.

G. Performance comparison

From the zero-symbol error curves in Fig. 13 and the threshold range bar chart in Fig. 14, it is clear that the T/R receivers can achieve zero-symbol errors over a reasonable range of the available thresholds for robust operations. This statement is supported by observing that the average curves in Fig. 13 of the T/R receivers closely approximate optimal performance. The percentage of available threshold range to achieve zero-symbol error in Fig. 14 is reasonable (~15%) when compared to the optimal (~25%). T/R receivers II and III appear to perform best in class (~20%) based on the zero-symbol error criterion.

The implementation advantages and simplicity of the T/R receivers offer an attractive alternative to the optimal. However, for a slowly time-varying channel, the T/R receivers may require more pilot transmissions to maintain the validity of the underlying channel Green's functions, which is the key to T/R receiver design in a highly reverberative environment. This completes the description of receiver performance; next we conclude our results and mention future work.

V. CONCLUSIONS

The feasibility of time-reversal (T/R) receiver designs for an acoustic point-to-point communication experiment is investigated in this paper. The receiver is applied to recover a (transmitted) coded information sequence in a highly reverberant environment with zero-symbol errors. The results of the T/R receiver are compared to the output of an optimal linear equalization receiver incorporating all of the reverberations as well. It is shown that it is feasible to design T/R receivers and expect near optimal performance in a highly reverberative environment. The design of four possible T/R receiver realizations using estimated Green's functions, pilot measurements, and BPSK codes with DSBSC-AM modulation and demodulation was demonstrated and their performance evaluated against the optimal linear equalizer. It was shown that it is possible to achieve zero-symbol error for each T/R design comparable to LE performance; thus, implying a practical application with obvious implementation advantages.[18,19] Thus, it is concluded that the T/R receiver can function reasonably well in a highly reverberant environment and is capable of detecting and recovering a transmitted information sequence with zero-symbol error.

Future efforts will be concentrating on the development and application of acoustic arrays to take full advantage of the additional gain afforded with a corresponding increase in the SNR as well as the focusing capability of T/R receiver designs.

ACKNOWLEDGMENTS

The initial technical discussions with Professor Simon Haykin of McMaster University kindled this idea of using T/R in communications environments. Dr. David Chambers of the LLNL has provided timely advice and direction during the course of this research. This work was performed under

the auspices of the U.S. Department of Energy contract by the Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.

[1] J. G. Proakis, *Digital Communications* (McGraw-Hill, New York, 1995).
[2] S. Haykin, *Communication Systems* (Wiley, New York, 2001).
[3] M. Fink and C. Prada, ''Acoustic time reversal mirrors,'' Inverse Probl. **17**, R1–R38 (2001).
[4] M. Fink, ''Time reversal in acoustics,'' Contemp. Phys. **37**, 95–109 (1996).
[5] S. Haykin, *Adaptive Filter Theory* (Prentice–Hall, Upper Saddle River, NJ, 2001).
[6] J. V. Candy, *Signal Processing: The Modern Approach* (McGraw-Hill, New York, 1988).
[7] A. Parvulescu, ''Matched-signal (MESS) processing by the ocean,'' J. Acoust. Soc. Am. **98**, 943–960 (1995).
[8] D. R. Jackson and D. R. Dowling, ''Phase conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181 (1991).
[9] D. R. Dowling, ''Acoustic pulse compression using passive phase-conjugate processing,'' J. Acoust. Soc. Am. **95**, 1450–1458 (1994).
[10] J.-P. Hermand and W. Roderick, ''Acoustic model-based matched filter processing for fading time-dispersive ocean channels: Theory and experiment,'' IEEE J. Ocean. Eng. **18**, 447–465 (1993).
[11] R. K. Brienzo and W. S. Hodgkiss, ''Broadband matched-field processing,'' J. Acoust. Soc. Am. **94**, 2821–2831 (1993).
[12] W. A. Kuperman, W. S. Hodgkiss, and H. C. Song, ''Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror,'' J. Acoust. Soc. Am. **103**, 25–40 (1998).
[13] J.-P. Hermand, ''Broad-band geoacoustic inversion in shallow water from waveguide impulse response measurements on a single hydrophone: theory and experimental results,'' IEEE J. Ocean. Eng. **24**, 41–66 (1999).
[14] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, ''A long range and variable focus phase-conjugation experiment in shallow water,'' J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[15] G. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, ''An initial demonstration of underwater acoustic communication using time reversal,'' IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[16] D. Rouseff, D. R. Jackson, W. L. Fox, C. D. Jones, J. A. Ritcey, and D. R. Dowling, ''Underwater acoustic communication by passive-phase conjugation: theory and experiment,'' IEEE J. Ocean. Eng. **26**, 821–831 (2001).
[17] S. Yon, M. Tanter, and M. Fink, ''Sound focusing in rooms: The time-reversal approach,'' J. Acoust. Soc. Am. **113**, 1533–1543 (2003).
[18] K. B. Smith, A. M. Abrantes, and A. Larraza, ''Examination of time-reversal acoustics in shallow water and applications to noncoherent underwater communications,'' J. Acoust. Soc. Am. **113**, 3095–3110 (2003).
[19] M. Heinemann, A. Larraza, and K. B. Smith, ''Experimental studies of applications of time-reversal acoustics to noncoherent underwater communications,'' J. Acoust. Soc. Am. **113**, 3111–3116 (2003).
[20] A. Derode, A. Tourin, J. de Rosny, M. Tanter, S. Yon and M. Fink, ''Taking advantage of multiple scattering to communicate with time-reversal antenna,'' Phys. Rev. Lett. **90**, No. 1, p. 014301-1–014301-4 (2003).
[21] A. F. Naguib, N. Seshadri, and A. R. Calderbank, ''Space-time coding and signal processing for high data rate wireless communications,'' IEE Signal Process. Mag. **17**, No. 3, pp. 76–92 (2000).
[22] A. J. Paulraj, and C. B. Papadias, ''Space-time processing for wireless communications,'' IEEE Signal Process. Mag. **14**, No. 6, pp. 49–83 (1997).
[23] M. Tanter, J.-L. Thomas and M. Fink, ''Time-reversal and the inverse filter,'' J. Acoust. Soc. Am. **108**, 223–234 (2000).
[24] H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (John Wiley, New York, 1968).

The Helmholtz equation least-squares method and Rayleigh hypothesis in near-field acoustical holography

Tatiana Semenova[a] and Sean F. Wu[b]
Department of Mechanical Engineering, Wayne State University, Detroit, Michigan 48202

(Received 6 May 2003; revised 6 November 2003; accepted 26 January 2004)

In this paper we present a numerical investigation of reconstructing time-harmonic acoustic pressure field in two dimensional space by using a series expansion—the so-called Helmholtz equation least-squares (HELS) method. Series expansion methods (or the Rayleigh methods) have been widely used in predicting the scattered acoustic pressure. With regularization, they can also be applied to reconstruction of acoustic pressure on the source surface from the measurements taken in the field, and HELS is the first such attempt for these problems. In this paper, we establish HELS in the framework of the Rayleigh methods and reveal its interrelationship with the Rayleigh hypothesis. In particular, to regularize a reconstruction problem, we use the method of quasisolutions, i.e., a Tikhonov regularization with an *a posteriori* choice of the regularization parameter. It is shown that without regularization HELS can still yield a satisfactory reconstruction of acoustic radiation from an arbitrary object when enough measurements are taken at sufficiently close range to the source. With regularization the number of measurements can be reduced and reconstruction accuracy be enhanced. It is concluded that HELS can be used to reconstruct acoustic radiation from a convex arbitrarily shaped vibrating object regardless of the validity of the Rayleigh hypothesis, although in practice the results will depend on the rate of convergence of the approximating sequence. [DOI: 10.1121/1.1687426]

PACS numbers: 43.60.Pt, 43.20.Rz, 43.40.Rj [EGW] Pages: 1632–1640

I. INTRODUCTION

Reconstruction of a time-harmonic acoustic field on the surface of a vibrating object based on the measurements taken in the near-field is a subject of near-field acoustical holography (NAH). Since its first appearance in 1980, NAH has incorporated a number of approaches for different applications. The Fourier transform-based NAH[1,2] is suitable for reconstructing acoustic radiation from source surfaces that contain a level of constant coordinate in a coordinate system in which the Helmholtz equation is separable. The integral formulation (based either on the Helmholtz integral equation[3] or on the single-layer potential[4]) and series expansion approaches[5,6] can be used to reconstruct acoustic radiation from arbitrarily shaped source surfaces.

The Helmholtz equation least-squares (HELS) is a series expansion method suggested for the reconstruction of acoustic radiation by Wang and Wu.[5] A number of experimental investigations[7,8] have shown that, if certain conditions are met, HELS can be a powerful alternative to the integral equation methods due to its simplicity and efficiency, as it requires fewer measurements in the reconstruction process. However, the question on how HELS can be implemented to reconstruct acoustic radiation from arbitrarily shaped objects has not been fully addressed. In this paper we attempt to fill this void.

A number of series expansion methods, sometimes referred to as the Rayleigh methods,[9] has been developed in the past for handling acoustic scattering problems. The convergence of the Rayleigh methods has been connected to the validity of the Rayleigh hypothesis, which is an assumption that the acoustic pressure can be continued analytically up to the source surface.[10] In scattering, it has been shown[11,12] that some of the Rayleigh methods such as the point-matching method (PMM) are applicable only if the Rayleigh hypothesis is valid, while others such as the adaptive point-matching method[12] and the least-squares approximation method (LSAM)[11,13,14,15] can be used regardless of the validity of the Rayleigh hypothesis. The Rayleigh methods can also be used to solve boundary-value problems for the Helmholtz equation, e.g., LSAM has been used to predict acoustic radiation from a finite cylinder.[16]

In scattering, some Rayleigh methods rely on the fact[17,11] that certain families of solutions to the Helmholtz equation are complete in $L^2(\Gamma)$ for any simple closed sufficiently smooth surface Γ. Such families can be used to approximate any solution of the Helmholtz equation uniformly on Γ. The difference between various Rayleigh methods lies in the way the expansion coefficients are found, which in turn determines their interrelationship with the Rayleigh hypothesis. When the Rayleigh hypothesis is not valid, some Rayleigh methods may fail to yield an approximation to the true solution on Γ, although such an approximation always exists.

The Rayleigh methods for reconstruction are based on the results proved by Isakov and Wu,[18] which include completeness of certain families of solutions to the Helmholtz equation for general domains and stability estimates for LSAM. In this paper, we show that HELS is a special case of LSAM, thus the dependence of HELS on the Rayleigh hypothesis is an artifact of discretization.

[a]Electronic mail: tsemenov@eng.wayne.edu
[b]Electronic mail: sean_wu@wayne.edu

It must be noted that there is a principal difference between the Rayleigh methods for reconstruction and those for the prediction of acoustic radiation or scattering. In predicting acoustic radiation or scattering, the nodes of the Rayleigh methods lie on the boundary of a vibrating object; while in reconstruction they lie on a measurement surface at some distance away from the object. Moreover, in a radiation or scattering problem, achieving a good approximation to the true solution on the boundary of an object guarantees a good approximation everywhere in its exterior.[19] Whereas in reconstruction, which is an ill-posed problem, achieving a good approximation on the measurement surface will not necessarily lead to an accurate reconstruction unless regularization is used.[18] Thus, the Rayleigh methods must be equipped with a regularization procedure in order to be used in reconstruction.

The method of quasisolutions,[20,21] which can be regarded as a Tikhonov regularization with an *a posteriori* choice of the regularization parameter,[19] was suggested by Isakov and Wu[18] to use with the Rayleigh methods for reconstruction of the acoustic pressure. A quasisolution, being an approximate solution to the original problem on a compact set, is sought to satisfy a prescribed bound. We want to emphasize here that this choice of the regularization method is by no means the best. It is used to justify the validity of HELS by Isakov and Wu[18] and to facilitate an examination of the interrelationship between HELS and the Rayleigh hypothesis in this paper. For this reason, we assume that a prescribed bound is known. A determination of this bound is beyond the scope of the present paper.

For simplicity, we consider reconstruction of the acoustic pressure using HELS in two dimensions, which allows us to locate the possible singularities in the analytic continuation of the solution across the source contour.[22,23] Our goal in the paper is to reveal the interrelationship of HELS with the Rayleigh hypothesis and examine numerically the impact of regularization on the approximating sequence.

II. SELECTION OF THE MATCHING POINTS IN THE RAYLEIGH METHODS FOR NAH

In this section we describe various Rayleigh methods that differ in the selection of the measurement points. In Sec. IV, we demonstrate numerically that the Rayleigh methods, when equipped with Tikhonov regularization, can be used to reconstruct the acoustic pressure on the source contour, provided that the approximating sequence converges sufficiently fast.

We consider a time-harmonic acoustic pressure u radiated from an arbitrary-shaped convex vibrating object Ω in $\mathbb{R}^2$, with a time-dependence factor of the form $\exp(-i\omega t)$. The acoustic pressure u satisfies the Helmholtz equation and Sommerfeld radiation condition,

$$\nabla^2 u + k^2 u = 0, \quad x \in \mathbb{R}^2 \backslash \overline{\Omega}, \tag{1}$$

$$\lim_{|x|\to\infty} |x|^{-1/2}\left(\frac{\partial u}{\partial |x|} - iku\right) = 0. \tag{2}$$

We assume that the acoustic pressure u satisfies a Dirichlet boundary condition on $\partial\Omega$. Other boundary conditions can be treated in a similar manner.[24,25]

In reconstruction, it is required to determine a solution u everywhere outside Ω and on the boundary $\partial\Omega$ based on the values it takes on a sufficiently smooth measurement contour $\Gamma = \partial\Omega_0$, where $\Omega_0 \supset \Omega$,

$$u(x) = \tilde{u}(x), \quad x \in \Gamma. \tag{3}$$

The problem of reconstructing the solution u on the boundary $\partial\Omega$ is recurrent in engineering applications, for example, in diagnosing the noise sources in a vibrating structure. Note that reconstruction problem possesses a unique solution u, i.e., the measured data (3) determine the acoustic pressure u on the source surface $\partial\Omega$ uniquely.[26]

In practice, the measured values of u on the right-hand side of Eq. (3) may be contaminated by noise. In the Rayleigh methods, an approximate solution u_N to Eqs. (1)–(3) is sought in the form of a linear combination of outgoing cylindrical waves, so that u_N satisfies the Helmholtz equation (1) and Sommerfeld radiation condition (2) exactly:

$$u_N(x) = \sum_{s=-N}^{N} a_s \phi_s(x), \tag{4}$$

where

$$\phi_s(x) = H_{|s|}^{(1)}(k|x|)e^{is\hat{x}}, \quad s = -N, \ldots, N, \tag{5}$$

where N is the expansion order, $(|x|, \hat{x})$ are polar coordinates of a point $x \in \mathbb{R}^2 \backslash \Omega$, $H_{|s|}^{(1)}$ are the Hankel functions of the first kind of order $|s|$, and a_s are the unknown coefficients to be determined by satisfying condition (3) in some sense.

In general, the coefficients a_s change with the expansion order N and differ from the coefficients A_s of the series solution to Eqs. (1) and (2) subject to boundary conditions on $\partial\Omega$,

$$u(x) = \sum_{s=-\infty}^{\infty} A_s \phi_s(x), \tag{6}$$

where A_s are determined by the following formula:[11]

$$A_s = \frac{i}{8}\int_{\partial\Omega}\left(u(y)\frac{\partial}{\partial\nu} - \frac{\partial u}{\partial\nu(y)}\right)H_s^{(2)}(k|y|)e^{-is\hat{y}}\, dS_y. \tag{7}$$

The series (6) converges absolutely and uniformly on compact subsets of the exterior of the minimum circle, i.e., the circle circumscribing $\partial\Omega$ with its center at the origin.[19] One of the methods in scattering and reconstruction, namely PMM,[27] is based on the assumption that the solution can be represented in the form (6) everywhere outside the Ω, i.e., that series (6) also converges uniformly inside the minimum circle up to the boundary $\partial\Omega$. This assumption is known as the Rayleigh hypothesis. If the boundary $\partial\Omega$, and the function u together with its normal derivative on $\partial\Omega$ are analytic, the validity of the Rayleigh hypothesis can be determined[23] by examining the locations of the singularities of the analytic continuation of u across $\partial\Omega$. When the singularities lie inside the maximum circle inscribing the boundary $\partial\Omega$ with its center at the origin, the Rayleigh hypothesis is valid.[10,28] When the singularities lie on or outside the maximum circle, the

Rayleigh hypothesis is invalid as the series (6) converges only outside of the circle on which the singularities lie, with its center at the origin. Possible singularities can be found[23] as the algebraic branch points of the Schwarz function of $\partial\Omega$. The actual singularities also depend on the values of the acoustic pressure and normal velocity on $\partial\Omega$.

A. LSAM

In LSAM, the coefficients a_s in expansion (4) are obtained by satisfying condition (3) in the least-squares sense,

$$\min_{a_{-N},\ldots,a_N} \|u_N - u\|^2_{L^2(\Gamma)}, \tag{8}$$

LSAM was first used in scattering by Meecham,[29] proved independently by Barantsev[30] and Yasuura,[13,11] and summarized by Ramm.[26] It also gives rise to the so-called "modified Rayleigh conjecture,"[25,31] which states that on Γ and on compact subsets of the exterior of $\partial\Omega$, the LSAM expansion converges to u in the mean. In backpropagation, a similar result is proved by Isakov and Wu.[18] Thus, LSAM can be used in the reconstruction of acoustic radiation even if the Rayleigh hypothesis is invalid. It has also been shown that on compact subsets of the exterior of the region bounded by Γ, it converges to u uniformly.[11] Moreover, the LSAM coefficients converge to coefficients (7).[30]

Discretizing Eq. (8), we arrive at

$$\int_\Gamma |u_N - u|^2 \, dS = \sum_{m=1}^{2M+1} \omega_m |u_N(x_m) - u(x_m)|^2 + \rho_M, \tag{9}$$

where x_m are integration nodes, ω_m depend on the quadrature formula used in discretization, and ρ_M is an error term. The following notation:

$$A_{mn} = \phi_{n-(N+1)}(x_m), \quad \xi_n = a_{n-(N+1)}, \quad b_m = u(x_m), \tag{10}$$

with $m = 1,\ldots,2M+1$, $n = 1,\ldots,2N+1$ allows us to rewrite $u_N(x_m)$ as a column of a matrix $u_N(x_m) = \sum_{n=1}^{2N+1} A_{mn}\xi_n = (A\xi)_m$ and to simplify the form of Eq. (9) to

$$\sum_{m=1}^{2M+1} \omega_m |(A\xi)_m - b_m|^2 + \rho_M. \tag{11}$$

Finally, we denote $W = \mathrm{diag}(\omega_m^{1/2})$, discard ρ_M in (11), and obtain a weighted least-squares problem,

$$\min_\xi \|S(A\xi - b)\|_2^2. \tag{12}$$

Although expansion (4) with coefficients determined by (8) can be used to approximate the solution u regardless of the validity of the Rayleigh hypothesis, the results may depend on the number of integration nodes and the quadrature formula used in discretization,[11,15] i.e., on the magnitude of the discarded term ρ_M in (11).

B. PMM

In PMM,[9] the coefficients a_s in expansion (4) are obtained by matching the approximate solution u_N to u_Γ at points $x_m \in \Gamma$, whose number is equal to the number of the unknown coefficients a_s,

$$u_N(x_m) = u(x_m), \quad x_m \in \Gamma, \quad m = 1,\ldots,2N+1. \tag{13}$$

The matching points are usually chosen equidistant, although they can be selected in a special way as in the adaptive point matching method.[12,32]

Using the notation (10), Eq. (13) can be rewritten in a matrix form,

$$A\xi = b. \tag{14}$$

In scattering, the adaptive PMM is proved to converge even if the Rayleigh hypothesis is invalid,[32] whereas the nonadaptive PMM may diverge, depending on the relative locations of the points x_m to the singularities of the analytic continuation of the acoustic pressure.[12,32,33] Therefore, when extended to the problem of reconstruction, the nonadaptive PMM may fail to provide an approximate solution u_N satisfying condition (3) when the measurement contour Γ lies partially inside the minimum circle.

Note that PMM is a special case of LSAM that can be obtained if we start from condition (8), divide Γ into $2N+1$ segments (not necessarily of equal length), and use the rectangular formula to approximate the integrals over each segment. However, the error term ρ_N of such a discretization may be large. Therefore there is no contradiction between the unconditional convergence of LSAM and conditional convergence of PMM depending on the validity of the Rayleigh hypothesis.

C. HELS

In HELS,[6] the coefficients a_s in expansion (4) are found by matching u_N to u_Γ at points $x_m \in \Gamma$ whose number is greater that the number of the unknowns a_s,

$$u_N(x_m) = u(x_m), \quad x_m \in \Gamma, \quad m = 1,\ldots,2M+1, \quad M > N. \tag{15}$$

In other words, a_s are determined by solving a matrix equation (14) in the least squares sense,

$$\min_\xi \|A\xi - b\|_2^2, \tag{16}$$

where the notation (10) is used.

The least-squares problem (16) can be solved via singular-value decomposition:

$$\xi = \sum_{j=1}^{\mathrm{rank}(A)} \frac{u_j^* b}{\sigma_j} v_j,$$

where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{2N+1}$ are singular values of A, u_j and v_j are left and right singular vectors of A, $j = 1,\ldots,2N+1$.

Note that in Eq. (15), no restrictions on the choice of the points x_m are imposed.[5] For convenience sake in practice, equidistant matching points are usually used.[5,8]

The early criticism of HELS[34] stemmed from the doubt whether matching u_N to u_Γ at a finite number of points on an arbitrarily shaped Γ would always lead to satisfying the condition (3) in some sense, i.e., whether HELS would work if the Rayleigh hypothesis is invalid and Γ is located partially inside the minimum circle. Indeed, the convergence of HELS should depend on the number and locations of the matching points, since HELS is a special case of LSAM. The interre-

lationship between HELS and the Rayleigh hypothesis is demonstrated via numerical examples in Sec. IV.

In HELS, the number of measurements is selected first and then the expansion order N is determined. Since only one set of measurements is used, the efficiency of the HELS algorithm is high.

III. REGULARIZATION

A. Regularization by the method of quasisolutions

In general, the Rayleigh methods in their original forms can only be used to approximate the acoustic pressure on and beyond a measurement contour Γ, since they solve the problem of approximating a radiating solution u by u_N on Γ. The ill-posed nature of a reconstruction problem is not embedded in the approximation problem; thus an attempt to backpropagate such a solution without regularization may fail. For applications of the Rayleigh methods to reconstruction problems, Isakov and Wu[18] propose regularization by quasisolutions. We will seek a quasisolution to the least-squares problem (12) among those whose norm is bounded on the source boundary $\partial\Omega$,

$$\|u_N\|^2_{L^2(\partial\Omega)} \leqslant \alpha^2, \quad \alpha = \|u\|_{L^2(\partial\Omega)} + \varepsilon, \quad \varepsilon > 0, \tag{17}$$

which is more readily implemented numerically than a bound on the solution in the whole reconstruction region[18] between the source boundary $\partial\Omega$ and the measurement curve $\Gamma = \partial\Omega_0$.

For an approximate solution u_N of the form (4), we can write the constraint (17) in a matrix form,

$$\xi^* B \xi \leqslant \alpha^2, \tag{18}$$

where B is a Hermitian positive definite matrix with the elements $B_{mn} = (\phi_m, \phi_n)_{L^2(\partial\Omega)}$, i.e., the Gram's matrix, whose elements can be calculated approximately by using a quadrature formula. Applying the Cholesky factorization $B = C^* C$, we arrive at a constraint in the standard form

$$\|C\xi\|_2^2 \leqslant \alpha^2. \tag{19}$$

The least-squares minimization problem (12) or (16) with the quadratic inequality constraint (19) is solved by generalized singular value decomposition (GSVD[35,36]). In this study, we use the built-in GSVD routine in Matlab.

The accuracy of reconstruction on the source contour $\partial\Omega$ depends critically on the value of the limiting constant α. In what follows, we assume that the right value of α is specified. The determination of ε in (17) is beyond the scope of the present paper.

B. Regularization by selecting an optimum expansion order for noisy input data

The regularization by quasisolutions may be omitted if the standoff distance, i.e., the distance between the measurement and reconstruction contours, is small and the number of expansion terms is not too large. Under these conditions, in order to prevent amplification of the measurement errors through the high-order terms in reconstruction when the input data are noisy, we must select an optimum expansion order by minimizing the integral error on the measurement contour[6] with respect to N,

$$R_N^2(\Gamma) = \int_\Gamma |u_N - u|^2 \, dS. \tag{20}$$

For simplicity, the integral error $R_N(\Gamma)$ is evaluated by a rectangular formula with a relatively small number of integration nodes. Wu[6] suggests to take the number of integration nodes in (20) equal to the number of the matching points in (15).

Since it is more meaningful to consider normalized quantities, we use the normalized integral error of approximation on Γ, $E_N^M(\Gamma)$, and the normalized integral error of reconstruction on $\partial\Omega$, $E_N^M(\partial\Omega)$, defined by the formulas

$$E_N^M(S) = \|u_N - u\|_{L^2(S)} / \|u\|_{L^2(S)}. \tag{21}$$

In reconstruction $u|_{\partial\Omega}$ is unknown, so it is impossible to calculate $E_N^M(\partial\Omega)$ and thus eliminate the high-order terms through which the errors are amplified by examining their effect directly. The underlying assumptions in regularization by determining an optimum expansion order are that (1) both the normalized integral error of reconstruction $E_N^M(\partial\Omega)$ and the normalized integral error of approximation $E_N^M(\Gamma)$ have a minimum; (2) they reach this minimum simultaneously at N_{opt}; and (3) N_{opt} is large enough to provide an acceptable accuracy.

We will see in Sec. IV D that in some cases $E_N^M(\partial\Omega)$ and $E_N^M(\Gamma)$ may attain a minimum at different values of N depending on the noise level and the standoff distance and, possibly, on the validity of the Rayleigh hypothesis. Thus, regularization by determining an optimum expansion order may sometimes fail. On the other hand, a Tikhonov regularization can take care of both the noise in the input data and the ill-posedness of a reconstruction problem.

IV. NUMERICAL EXAMPLES

A. Description of examples

We use HELS to reconstruct the acoustic pressure in two cases: (a) when the Rayleigh hypothesis is valid and (b) when the Rayleigh hypothesis is invalid.

In the case (a), the vibrating object is a circle $C = \{x^2 + y^2 = 1.5^2\}$ with the boundary condition $v|_C = \exp(-\hat{x}^2) - C_1$, where C_1 is a constant. The Rayleigh series converges in the exterior of C. We use HELS to reconstruct the acoustic pressure radiated from C on an ellipse $\partial\Omega = \{(x/2)^2 + y^2 = 2^2\}$ based on measurements taken on a similarly shaped ellipse $\Gamma = \{(x/2)^2 + y^2 = r_\Gamma^2\}$.

In case (b), we reconstruct the acoustic pressure radiated from an ellipse $\partial\Omega = \{(x/2)^2 + y^2 = 2^2\}$ with boundary data

$$u|_{\partial\Omega} = 1. \tag{22}$$

based on measurements taken on a similarly shaped ellipse $\Gamma = \{(x/2)^2 + y^2 = r_\Gamma^2\}$. Since the focal distance for the ellipse $\partial\Omega$ is greater than the semiminor axis, the Rayleigh hypothesis is invalid,[28] which can also be easily seen by evaluating the partial sums of the Rayleigh series on $\partial\Omega$ numerically.

The acoustic wave number k in these cases is equal to 2. The exact input data b and the noise contaminated ones $\tilde{b}$ are used in Eq. (16). The noisy data are produced by adding a uniformly distributed noise such that $\|b-\tilde{b}\|_2=\delta\|b\|_2$, where δ is the noise level.

We have selected the boundary conditions (22) for the sake of illustration: a Rayleigh method applied to the reconstruction problem with boundary conditions (22) yields a fast convergent approximating sequence. However, the rate of convergence is significantly slowed down when the boundary conditions with a wave number much larger than the acoustic wave number are used, e.g.,

$$u|_{\partial\Omega}=\cos l\hat{x}, \quad l\gg k. \tag{23}$$

The acoustic pressure is calculated on Γ by the Nyström method[38] based on the given acoustic pressure on $\partial\Omega$. Next, HELS is used to reconstruct the acoustic pressure on $\partial\Omega$ from the data on Γ and the reconstructed values are compared with those given by the boundary conditions. In what follows, we use Eq. (15) and take equidistant matching points.

B. HELS: when the Rayleigh hypothesis is valid

1. HELS and a truncated series

As illustrated in Fig. 1(a), if the input data are exact and the Rayleigh series converges to the boundary, the truncated Rayleigh series (cf. Sec. II C) can be used for reconstruction from any standoff distance, and no regularization is necessary. Figure 1(a) shows the normalized integral errors of approximation and reconstruction by HELS versus the expansion order N.

However, if the input data are noisy, the HELS coefficients differ from the series solution coefficients calculated by the integration formula (7). As a result, the normalized integral errors for the truncated series level off and are the same for all contours [see Fig. 1(b)], while the normalized integral error of reconstruction by HELS starts to grow beyond a certain N (see Fig. 2). This is because in calculating the coefficients of the series solution by integration, the noise affects all the coefficients equally, while in solving the least-squares problem (16), the noise affects the coefficients of the higher-order terms more than those of the lower-order terms.

2. The choice of the maximum number of terms

To determine the expansion order $N_{\max}$ sufficient to yield an acceptable reconstruction, Wu *et al.*[7] follow the work of Weinreich and Arnold[39] and take $N_{\max}$ as the smallest integer greater than kR, where R is the radius of the minimum circle. The wave functions with indices larger than $N_{\max}$ are found to make no significant contribution to the solution (cf. Sec. IV C 1). Figure 2 shows the normalized integral errors of approximation on Γ and reconstruction on $\partial\Omega$, which depicts clearly that the terms with indices higher than $kR=8$ can be omitted.

3. The choice of the number of measurements

On the basis of experimental and numerical studies with HELS, Wu *et al.*[7] suggest taking $M=1.4N_{\max}$. However, since PMM converges when the Rayleigh hypothesis is valid, a smaller $M=N_{\max}$ suffices[27] (cf. Sec. IV C 2).

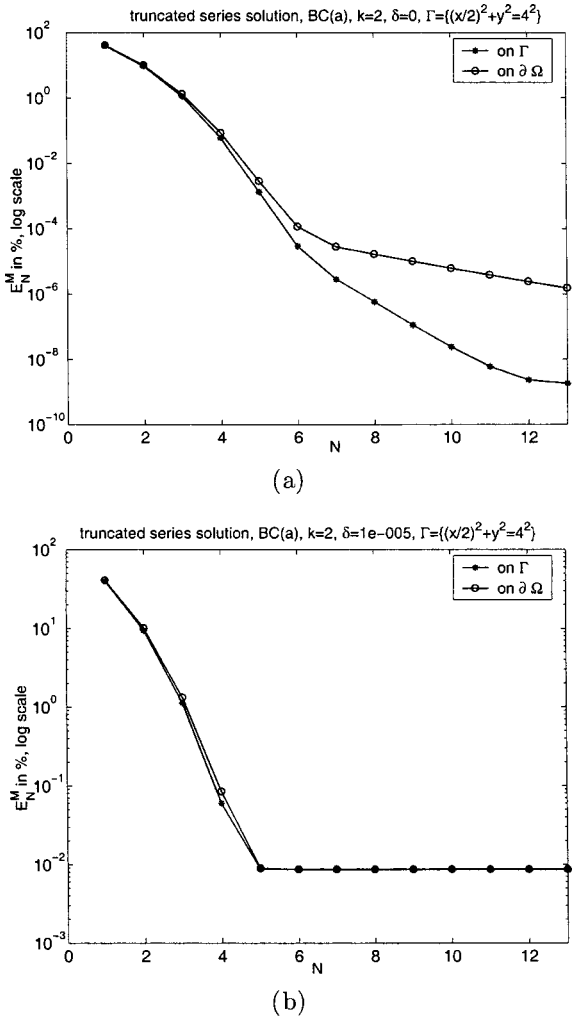


FIG. 1. Truncated series solution when the Rayleigh hypothesis is valid: the normalized integral error of approximation on the measurement contour $E_N^M(\Gamma)$ and the normalized integral error of reconstruction on the source boundary $E_N^M(\partial\Omega)$ versus the expansion order N. Case (1a): exact input data, case (1b): noisy input data with $\delta=10^{-5}$.

C. HELS: when the Rayleigh hypothesis is invalid

In order to examine the behavior of HELS approximating sequence with and without regularization by the method of quasisolutions, we consider reconstruction from a small standoff distance ($d=0.4$).

1. The choice of the maximum number of terms

The criterion for a sufficient expansion order[7] $N_{\max}\approx kR$ can only be used when the series (6) converges in the reconstruction region. When the Rayleigh hypothesis is not valid, taking $N_{\max}\approx kR$ may be insufficient. Indeed, this criterion is based on the assumption that for any boundary conditions the approximating sequence converges on Γ as fast as the series (6) converges outside of the minimum circle, which is not always true. In general, the rate of convergence of the approximating sequence depends on the shape of the object and the boundary conditions.

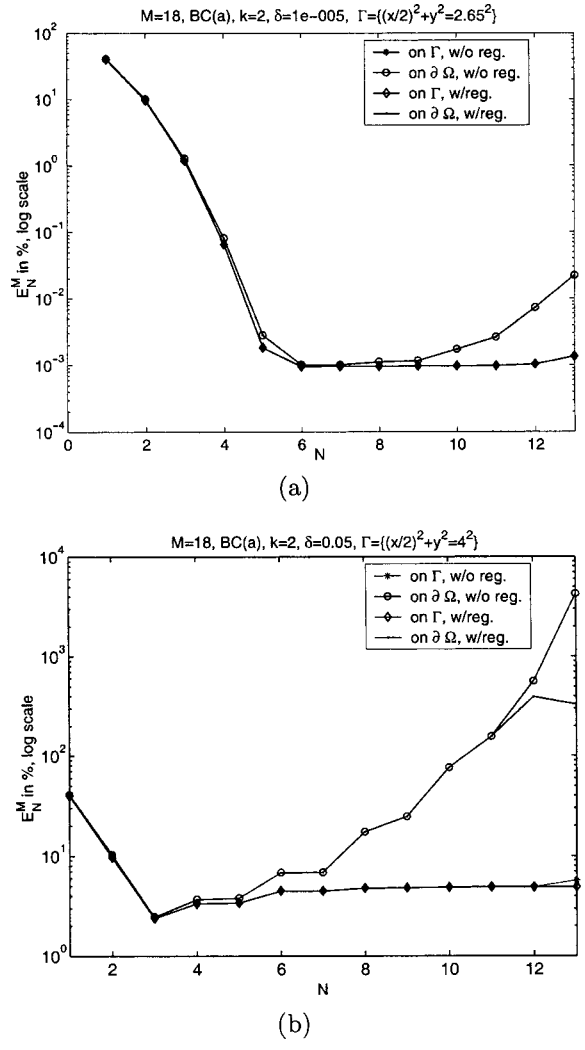


FIG. 2. HELS with and without regularization by quasisolutions when the Rayleigh hypothesis is valid: the normalized integral error of approximation on the measurement contour $E_N^M(\Gamma)$ and the normalized integral error of reconstruction on the source boundary $E_N^M(\partial\Omega)$ versus the expansion order N, given noisy input data. Case (2a): $\delta=10^{-5}$, $\Gamma=\{(x/2)^2+y^2=2.65^2\}$; (2b): $\delta=0.05$, $\Gamma=\{(x/2)^2+y^2=4^2\}$.

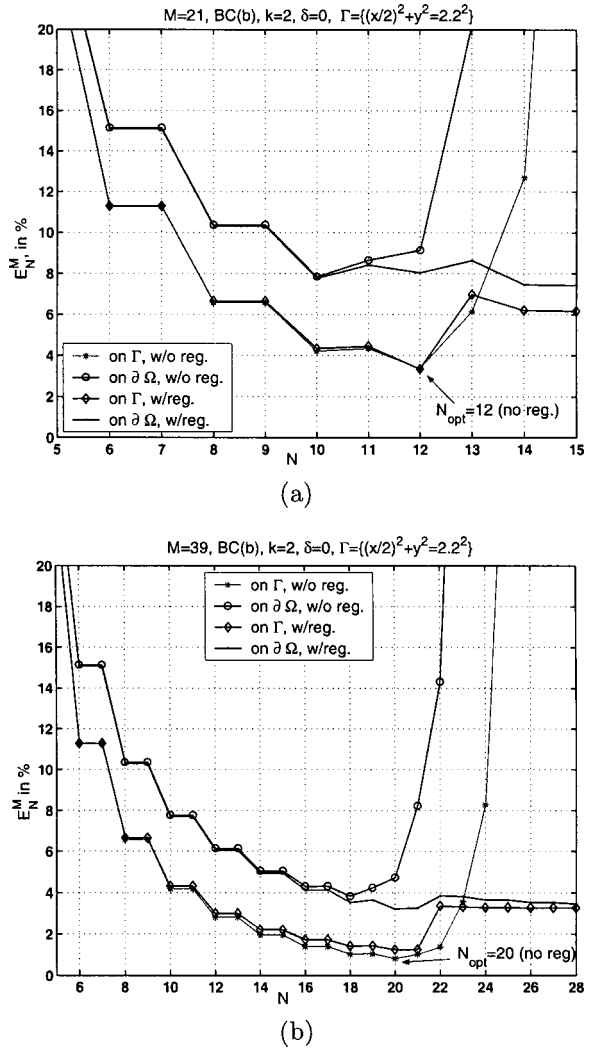


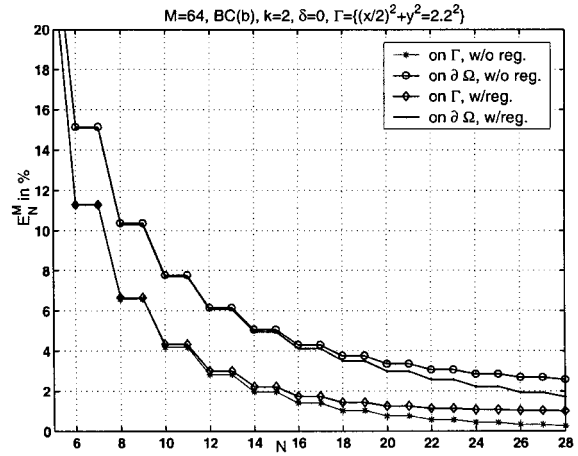
FIG. 3. Optimum expansion order N depending on M in HELS with and without regularization by quasisolutions when the Rayleigh hypothesis is not valid: the normalized integral error of approximation on the measurement contour $E_N^M(\Gamma)$ and the normalized integral error of reconstruction on the source boundary $E_N^M(\partial\Omega)$ versus the expansion order N, given exact input data. Case (3a): $N_{\max}=15$, $M=1.4N_{\max}$; case (3b): $N_{\max}=28$, $M=1.4N_{\max}$.

A slower rate of convergence calls for increasing the expansion order, however, this will usually yield an ill-conditioned matrix A due to the fact[26,40] that the expansion functions ϕ_s do not form a minimal system in $L^2(\Gamma)$ for most measurement contours Γ. In this case, regularization must be applied to overcome the ill-conditioning of matrix A, which is caused by the choice of the expansion functions and has nothing to do with backpropagation.

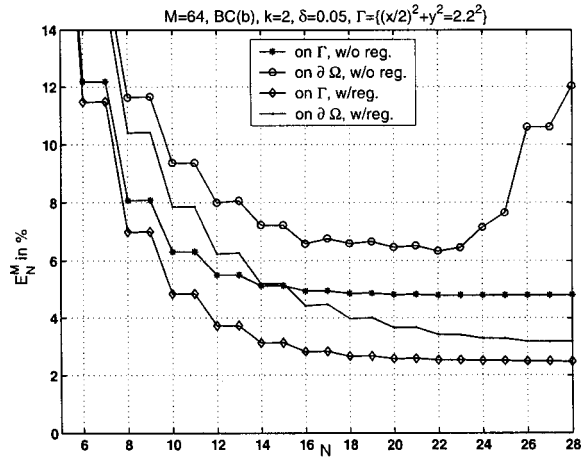
2. The choice of the number of measurements

When series (6) does not converge on the measurement contour Γ, the number of measurements $M=1.4N_{\max}$ may be insufficient in contrast to the case where the Rayleigh hypothesis is valid. Yasuura and Ikuno[15] showed that in scattering problems, taking $M>2N$ in LSAM discretized by a rectangular formula yields acceptable solutions for a sufficiently large expansion order N, regardless of the validity of the Rayleigh hypothesis. In contrast to other Rayleigh methods, the ratio M/N changes with N in HELS. Thus, if M is not large enough, this ratio M/N may be acceptable for a smaller N and insufficient for a larger N. For M to be sufficiently large for all $N=1,...,N_{\max}$, it should be at least greater than $2N_{\max}$.

If M is not large enough, the error of reconstruction $E_N^M(\Gamma)$ decreases for $N<N_{\text{opt}}$ and increases for $N>N_{\text{opt}}$ (see Fig. 3). The error of approximation behaves in a similar way instead of leveling off once it has reached the noise level. It is easily seen that the growth of $E_N^M(\Gamma)$ for $N>N_{\text{opt}}$ is not caused by the presence of noise in the input data as on that shown in Fig. 2, but by the fact that the chosen M is not large enough for $N>N_{\text{opt}}$. Indeed, the input data are exact in cases shown in Fig. 3 and an increase in M causes an increase in N_{opt}. Moreover, for the problem of reconstruction with boundary conditions (23) with $l=10$, in which case the convergence of the approximating sequence is much slower, the errors $E_N^M(\Gamma)$ and $E_N^M(\partial\Omega)$ take their minimums at the same values of N_{opt} as shown in Fig. 3 (with the same M and $N_{\max}$



(a)



(b)

FIG. 4. HELS with and without regularization by quasisolutions when the Rayleigh hypothesis is not valid: the normalized integral error of approximation on the measurement contour $E_N^M(\Gamma)$ and the normalized integral error of reconstruction on the source boundary $E_N^M(\partial\Omega)$ versus the expansion order N with $N_{\max}=15$, $M=2.3N_{\max}$. Case (4a): exact input data—both approximation and reconstruction errors decrease with N. Case (4b): noisy input data with $\delta=0.05$-reconstruction error increase with N if no regularization is applied.

as the ones used in Fig. 3), but at much higher levels. For example, in this case $E_N^M(\partial\Omega)=87\%$ at $N=10$, $E_N^M(\Gamma)=62\%$ at $N=12$ [cf. Fig. 3(a) with $N_{\max}=15$, $M=1.4N_{\max}$] and $E_N^M(\partial\Omega)=51\%$ at $N=18$, $E_N^M(\Gamma)=18\%$ at $N=20$ [cf. Fig. 3(b) with $N_{\max}=28$, $M=1.4N_{\max}$].

Figure 4(a) shows that taking the same value of $N_{\max}$ as that in Fig. 3(b) and increasing the value of M further yields

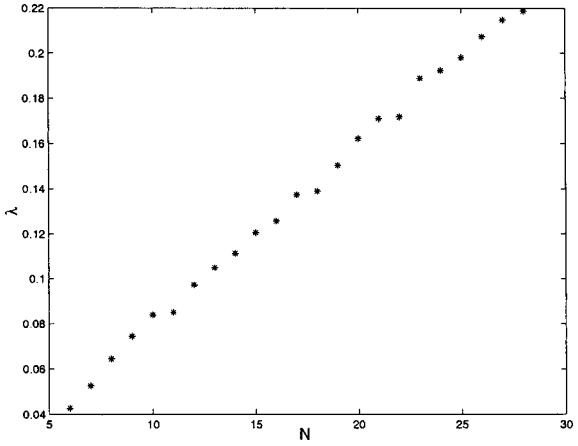


FIG. 5. Regularization parameter λ versus the expansion order N for BC (b), $M=64$, $k=2$, $\delta=0.05$, $\Gamma=(x/2)^2+y^2=2.2^2$.

monotonically decreasing normalized integral errors of approximation on Γ and reconstruction on $\partial\Omega$.

D. The choice of the optimum expansion order for noisy input data

Sometimes it is possible to obtain an optimum expansion order as described in Sec. III B. Such an example is shown in Fig. 2(b), where the normalized integral errors of approximation on Γ and the reconstruction on $\partial\Omega$ have a minimum at the same $N_{\text{opt}}=3$. For $N>N_{\text{opt}}$, the integral error of approximation levels off at the error level with fluctuations smaller than the noise level itself, while the integral error of reconstruction increases. Thus $N=3$ is indeed the optimum expansion order in this case.

However, the normalized integral errors $E_N^M(\Gamma)$ and $E_N^M(\partial\Omega)$ may behave differently in other cases, e.g., as shown in Figs. 2(a) and 4(b). While $E_N^M(\Gamma)$ is leveled off at the noise level [5% in Fig. 4(b)], the integral error of reconstruction $E_N^M(\partial\Omega)$ levels off [at 6.5% in Fig. 4(b)] first but then starts to grow beyond a certain N_1 due to the amplification of noise in the higher-order terms. This N_1 depends on the noise level and standoff distance, as well as on the rate of convergence of the sequence; it decreases if either the noise level or standoff distance is increased, and increases if the rate of convergence of the sequence is slower. Thus, it is impossible to predict when $E_N^M(\partial\Omega)$ reaches its minimum by examining the behavior of $E_N^M(\Gamma)$. Note that the regularization parameter changes with N as shown in Fig. 5.

Table I shows that a better accuracy can be achieved if

TABLE I. HELS with and without regularization by quasisolutions when the Rayleigh hypothesis is not valid: maximum pointwise relative error e_N^M of reconstruction of $\partial\Omega$, in %, for different values of N and M. The input data are noisy with $\delta=0.05$.

N		10[a]	10	12[b]	12	12	18[c]	18	21[d]	21	26	26
M/N		2.1	2.5	1.75	2.0	2.5	2.2	2.5	1.86	2.5	2.0	2.5
e_N^M in %	w/o reg.	2.8	2.4	4.7	2.7	2.4	2.4	2.0	5.1	4.3	4.9	4.3
	w/reg.	2.0	1.8	2.5	1.8	1.6	1.4	1.2	1.1	1.2	1.3	1.0

[a] $E_N^M(\partial\Omega)$ reaches a minimum when $N_{\max}=15$, $M=1.4N_{\max}$.
[b] $E_N^M\Gamma$ reaches a minimum when $N_{\max}=15$, $M=1.4N_{\max}$.
[c] $E_N^M(\partial\Omega)$ reaches a minimum when $N_{\max}=28$, $M=1.4N_{\max}$.
[d] $E_N^M\Gamma$ reaches a minimum when $N_{\max}=28$, $M=1.4N_{\max}$.

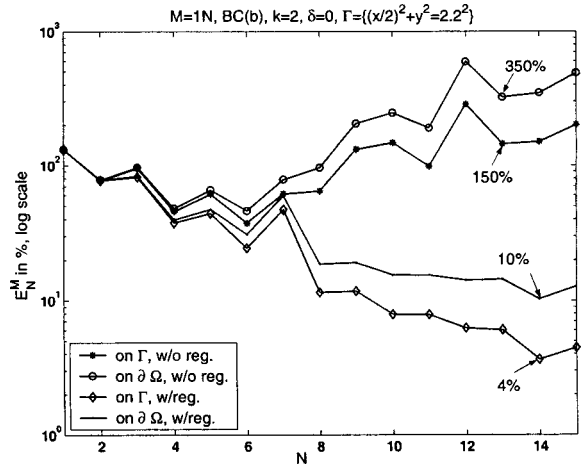


FIG. 6. PMM with and without regularization by quasi-solutions when the Rayleigh hypothesis is not valid: the normalized integral error of approximation on the measurement contour $E_N^M(\Gamma)$ and the normalized integral error of reconstruction on the source boundary $E_N^M(\partial\Omega)$ versus the expansion order N, given exact input data. Both reconstruction and approximation errors increase with N if no regularization is applied.

the expansion order N and ratio M/N are increased. Note that without Tikhonov regularization, HELS needs a much larger M to achieve a comparable accuracy.

Without Tikhonov regularization HELS still can be used in reconstruction when the standoff distance is sufficiently small and the expansion order is selected optimally for reconstruction. However, in most cases it is impossible to find an optimum expansion order without the knowledge of the exact solution. On the other hand, when regularization by quasisolutions is applied, there is no need to determine the optimum expansion order. Any sufficiently large N will produce a satisfactory reconstruction.

E. PMM: the effect of regularization

When the Rayleigh hypothesis is not valid and the measurement contour Γ lies partially inside the minimum circle, PMM does not converge. However, Fig. 6 shows that applying regularization by quasisolutions produces relatively good results for both approximation on Γ and reconstruction on $\partial\Omega$. For $N=14$ and 5% noise level, the pointwise error of approximation on Γ is 1% with regularization and 90% without it, the pointwise relative error of reconstruction on $\partial\Omega$ is 3% with regularization and 220% without it. Thus, if Tikhonov regularization is used, both PMM and HELS with small M will work, however, increasing M will increase the accuracy. These results are consistent with the remark made by Apel'tsin and Kyurkchan[41] that the number of measurements may be lowered to some extent if regularization by imposing an *a priori* restriction on the class of functions in which the solution is sought is applied to ''solving diffraction problems when the coefficients in Rayleigh's expansion are found from the requirement that the boundary condition be satisfied at points whose number exceeds the number of terms in the expansion.''

V. CONCLUSION

We have considered HELS in the framework of the Rayleigh methods and examined its interrelationship with the Rayleigh hypothesis. Although both HELS and PMM are special cases of LSAM, the HELS method differs from other Rayleigh methods in that it uses one set of measurements to calculate approximate solutions of different expansion orders.

When the Rayleigh hypothesis is valid, HELS with a small number of measurements can be used in reconstruction. When the Rayleigh hypothesis is not valid and the measurement contour lies inside the minimum circle, the number of measurements must be increased so that $M/N_{\max}>2$ holds.

We have used regularization by the method of quasisolutions. This regularization may be omitted if the measurements are taken at a sufficiently close range to the source and the right number of expansion terms is selected for a given standoff distance and noise level in the input data. However, the expansion order that minimizes the normalized integral error of approximation is not necessarily optimal for reconstruction. Thus, without regularization, it may be unclear how many expansion terms to take in order to prevent amplification of noise through the higher-order terms. On the other hand, if Tikhonov regularization is applied, any large expansion order will suffice. Moreover, even if the series does not converge on the measurement contour, the number of measurements may be reduced compared with that without regularization. Besides, a regularized solution may provide a better accuracy than a nonregularized solution. Similar conclusions hold for three-dimensional problems as well.

It must be pointed out that regularization by the method of quasisolutions may be impractical, since it is very sensitive to the choice of the constraint and requires some *a priori* information about the magnitude of the true solution that can only be obtained with a large number of measurements. We plan to consider other regularization methods in a separate paper.

For reconstruction, HELS seems to be the most preferable one among the Rayleigh methods for engineering applications. Its efficiency is high since it uses only one set of measurements. The adaptive PMM or LSAM with a quadrature formula other than the rectangular formula may provide a better accuracy, however, they require precise positioning of the microphones, which is not easy to implement in practice. On the other hand, HELS with equidistant or even arbitrary points can be implemented easily and cost effectively. A possible difficulty may lie in the rate of convergence of HELS, which depends on the shape of the object and boundary conditions. When the approximating sequence converges sufficiently fast, HELS can be an effective alternative to the integral methods in reconstruction.

ACKNOWLEDGMENTS

The authors wish to thank Dr. V. Isakov for a helpful discussion and Dr. E. G. Williams for his continued interest in this work. This work was supported by a grant from National Science Foundation Grant No. ITR/ACS 0081270.

[1] E. G. Williams and J. D. Maynard, ''Holographic imaging without the wavelength resolution limit,'' Phys. Rev. Lett. **45**, 554–557 (1980).
[2] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, United Kingdom, 1999).
[3] M. R. Bai, ''Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrary shaped geometries,'' J. Acoust. Soc. Am. **92**, 533–549 (1992).
[4] T. Delillo, V. Isakov, N. Valdivia, and L. Wang, ''The detection of the source of acoustical noise in two dimensions,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **61**, 2104–2121 (2001).
[5] Z. Wang and S. F. Wu, ''Helmholtz equation-least-squares method for reconstructing the acoustic pressure field,'' J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[6] S. F. Wu, ''On reconstruction of acoustic pressure fields using the Helmholtz equation least-squares method,'' J. Acoust. Soc. Am. **107**, 2511–2522 (2000).
[7] S. Wu, N. Rayess, and X. Zhao, ''Visualization of acoustic radiation from a vibrating bowling ball,'' J. Acoust. Soc. Am. **109**, 2771–2779 (2001).
[8] N. Rayess and S. F. Wu, ''Experimental validation of the HELS method for reconstructing acoustic radiation from a complex vibrating structure,'' J. Acoust. Soc. Am. **107**, 2955–2963 (2000).
[9] P. M. van den Berg, ''Reflection by a grating: Rayleigh methods,'' J. Opt. Soc. Am. **71**, 1224–1229 (1981).
[10] R. F. Millar, ''Rayleigh hypothesis in scattering problems. (With author's reply),'' Electron. Lett. **5**, 416–418 (1969).
[11] R. F. Millar, ''The Rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatterers,'' Radio Sci. **8**, 785–796 (1973).
[12] A. I. Kleev and A. B. Manenkov, ''The convergence of point-matching techniques,'' IEEE Trans. Antennas Propag. **37**, 50–54 (1989).
[13] H. Ikuno and K. Yasuura, ''Improved point-matching method with application to scattering from a periodic surface,'' IEEE Trans. Antennas Propag. **AP-21**, 657–662 (1973).
[14] J. P. Hugonin, R. Petit, and M. Cadilhac, ''Plane-wave expansions used to describe the field diffracted by a grating,'' J. Opt. Soc. Am. **71**, 593–598 (1981).
[15] K. Yasuura and H. Ikuno, ''On the modified Rayleigh hypothesis and the mode-matching method,'' in *International Symposium on Antennas and Propagation*, Sendai, Japan (Institute of Electronics and Communication Engineers of Japan, Tokyo, Japan, 1971), pp. 173–174.
[16] W. Williams, N. G. Parke, D. A. Moran, and C. H. Sherman, ''Acoustic radiation from a finite cylinder,'' J. Acoust. Soc. Am. **36**, 2316–2322 (1964).
[17] I. N. Vekua, ''On completeness of a system of methaharmonic functions,'' Dokl. Akad. Nauk SSSR **90**, 715–718 (1953).
[18] V. Isakov and S. F. Wu, ''On theory and application of the Helmholtz equation least squares method in inverse acoustics,'' Inverse Probl. **18**, 1147–1159 (2002).
[19] D. Colton and R. Kress, ''Inverse acoustic and electromagnetic scattering theory,'' *Applied Mathematical Sciences* (Springer-Verlag, Berlin, 1992), Vol. 93, Sec. 2.5.
[20] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Winston, Washington, DC, Wiley, New York, 1977), Sec. 1.2.
[21] V. Isakov, ''Inverse problems for partial differential equations,'' in Ref. 19 (Springer-Verlag, New York, 1998), Vol. 127.
[22] P. M. van den Berg and J. T. Fokkema, ''The Rayleigh hypothesis in the theory of diffraction by a cylindrical obstacle,'' IEEE Trans. Antennas Propag. **AP-27**, 577–583 (1979).
[23] R. F. Millar, ''Singularities and the Rayleigh hypothesis for solutions to the Helmholtz equation,'' IMA J. Appl. Math. **37**, 155–171 (1986).
[24] A. Doicu, Y. Eremin, and T. Wriedt, *Acoustic and Electromagnetic Scattering Analysis Using Discrete Sources* (Academic, New York, 2000).
[25] A. G. Ramm, ''Modified Rayleigh conjecture and applications,'' J. Phys. A **35**, L357–L361 (2002).
[26] A. G. Ramm, ''Scattering by obstacles,'' of *Mathematics and its Applications* (Reidel, Dordrecht, 1986), Vol. 21.
[27] L. Lewin, ''On the restricted validity of the point-matching techniques,'' IEEE Trans. Microwave Theory Tech. **MTT-18**, 1041–1047 (1970).
[28] R. G. Barantsev, V. V. Grudtsyn, F. T. Zabolotnyi, and V. V. Kozachek, ''On the Rayleigh hypothesis in the problem of scattering from finite arbitrary shaped bodies,'' Vestn. Leningr. Univ. **7**, 56–62 (1971) (in Russian).
[29] W. C. Meecham, ''Variational method for the calculation of the disctribution of energy reflected from a periodic surface. i.,'' J. Appl. Phys. **27**, 361–367 (1956).
[30] R. G. Barantsev and V. V. Kozachek, ''Analysis of the scattering matrix for the objects of complex shape,'' Vestn. Leningr. Univ. **7**, 69–77 (1968) (in Russian).
[31] S. Gutman and A. G. Ramm, ''Numerical implementation of the MRC method for obstacle scattering problems,'' J. Phys. A **35**, 8065–8074 (2002).
[32] A. I. Kleev and A. B. Manenkov, ''Adaptive point-matching method in two-dimensional problems,'' Radiophys. Quantum Electron. **29**, 414–420 (1986).
[33] S. Christiansen and R. E. Kleinman, ''On a misconception involving point collocation and the Rayleigh hypothesis,'' IEEE Trans. Antennas Propag. **44**, 1309–1316 (1996).
[34] S. F. Wu, ''Nearfield acoustic holography and its evolution,'' J. Acoust. Soc. Am. (in print).
[35] W. Gander, ''Least squares with a quadratic constraint,'' Numer. Math. **36**, 291–307 (1980/81).
[36] Å. Björck, *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics (SIAM, Philadelphia, PA, 1996), Sec. 5.3.
[37] R. G. Barantsev, V. V. Grudcyn, F. T. Zabolotnyi, and V. V. Kozaček, ''Rayleigh's hypothesis in the problem of scattering by finite bodies of arbitrary form,'' Vestn. Lening. Univ. **7**, 56–62 (1971).
[38] R. Kress, ''Linear integral equations,'' Ref. 19 (Springer-Verlag, Berlin, 1989), Vol. 82, Sec. 12.2.
[39] G. Weinreich and E. B. Arnold, ''Method for measuring acoustic radiation fields,'' J. Acoust. Soc. Am. **68**, 404–411 (1980).
[40] G. F. Crosta, ''Complete families and Rayleigh obstacles,'' J. Comput. Acoust. **9**, 611–622 (2001).
[41] B. V. Apel'tsin and A. G. Kyurkchan, ''Rayleigh's hypothesis and the analytic properties of wave fields,'' Sov. J. Commun. Technol. Electron. **30**, 97–111 (1985).

Partial sound field decomposition in multireference near-field acoustical holography by using optimally located virtual references

Yong-Joe Kim[a] and J. Stuart Bolton
Ray W. Herrick Laboratories, Purdue University, 140 South Intramural Drive, West Lafayette, Indiana 47907-2031

Hyu-Sang Kwon
Acoustics and Vibration Research Group, Korea Research Institute of Standards and Science, P.O. Box 102, Yusong, Taejon, 305-600, Korea

(Received 20 March 2003; revised 15 November 2003; accepted 24 November 2003)

It has been shown previously that the multiple reference and field signals recorded during a scanning acoustical holography measurement can be used to decompose the sound field radiated by a composite sound source into mutually incoherent partial fields. To obtain physically meaningful partial fields, i.e., fields closely related to particular component sources, the reference microphones should be positioned as close as possible to the component physical sources that together comprise the complete source. However, it is not always possible either to identify the optimal reference microphone locations prior to performing a holographic measurement, or to place reference microphones at those optimal locations, even if known, owing to physical constraints. Here, post-processing procedures are described that make it possible both to identify the optimal reference microphone locations and to place virtual references at those locations *after* performing a holographic measurement. The optimal reference microphone locations are defined to be those at which the MUSIC power is maximized in a three-dimensional space reconstructed by holographic projection. The acoustic pressure signals at the locations thus identified can then be used as optimal ''virtual'' reference signals. It is shown through an experiment and numerical simulation that the optimal virtual reference signals can be successfully used to identify physically meaningful partial sound fields, particularly when used in conjunction with partial coherence decomposition procedures. [DOI: 10.1121/1.1642627]

PACS numbers: 43.60.Sx [EGW] Pages: 1641–1652

I. INTRODUCTION

Near-field acoustical holography (NAH) is a useful tool for visualizing noise sources and their associated sound fields since it allows sound fields that are measured on a surface to be reconstructed throughout a three-dimensional space.[1,2] For example, after the sound pressure is measured on a surface of constant coordinate in plane, cylindrical, or spherical coordinates, the pressure, particle velocity, and sound intensity fields can be reconstructed on plane or concentric surfaces, as appropriate, either closer to or farther from the source.

For the NAH procedure to be successful in its simplest form, the measured sound field must be spatially coherent. Such a spatially coherent sound field can be obtained by capturing the sound pressure on the entire measurement surface simultaneously. However, the latter measurement requires the use of many field microphones, perhaps hundreds or thousands, which may not always be practical. When the sound field generated by a single coherent source is stationary, however, a relatively small number of scan microphones can be used in combination with a fixed-location reference microphone to sample data on portions of the hologram surface sequentially. The reference spectral data can then be combined with the transfer functions linking the references and the field points on the hologram surface to create a complex pressure distribution on the latter surface that may then be projected as desired.

A multireference procedure, referred to as the Spatial Transformation of Sound Fields (STSF), was introduced by Hald to accommodate situations in which the sound field is generated by a composite source comprising more than one mutually incoherent source.[3] The latter procedure, described in terms of a reference cross-spectral matrix and a cross-spectral matrix linking the reference and field signals, is based on using singular value decomposition (SVD) to separate the total sound field into a set of partial fields that are incoherent with each other. The partial fields can then be independently projected and added together to yield the quadratic properties of the sound field on the reconstruction surfaces of interest. Note that in this procedure the characteristics of the partial fields depend on the reference microphone locations with respect to the component sources, and thus the partial field decomposition is not unique. As a result, the partial fields determined in this way cannot usually be associated with particular component sources.

Since the introduction of STSF, a number of investigations have focused on separating the total sound field into

[a]Corresponding author. Phone: +1-765-496-6008; fax: +1-765-494-0787; electronic mail: kimyj1@purdue.edu

individual partial sound fields that *can* be associated with meaningful physical sources. As an alternative to the SVD procedure, Hallman and Bolton[4] suggested the application of partial coherence decomposition (PCD), based on Gauss elimination,[5] to separate the partial fields. Kwon and Bolton[6] compared the performance of the SVD and PCD methods; they also introduced the use of the multiple signal classification (MUSIC) algorithm[7] for the selection of the best references from among many candidates positioned around a composite sound source. It was shown that the reference microphones should be positioned close to the individual physical sources if physically meaningful partial sound fields were to be obtained. However, in practice, the locations of the most prominent sources are sometimes not known prior to performing a holographic measurement. There are also cases in which it is impossible to place references at the desired locations, even when those locations are known: for example, when a reference microphone placed close to a physical source might itself induce flow noise or when the desired locations are otherwise inaccessible for practical reasons.

Thus, to facilitate the decomposition of a sound field into physically meaningful components, it would be desirable to be able to place ''virtual'' references at optimal locations *after* performing a holographic measurement made using a set of sufficient, but nonoptimally located references.

Nam and Kim introduced a virtual reference procedure in which virtual references are located at positions on the source surface where the amplitude of an acoustical property of interest, such as the pressure, particle velocity, or active intensity, is a maximum.[8] The latter procedure was based on a vector representation[9] of the sound field combined with a knowledge of various point-to-point transfer functions, both measured and estimated by using holographic projections. Note that in their procedure it is required that the number of ''real'' and ''virtual'' references, first used to conduct the measurement and then to post-process the data, be the same. In general, the number of virtual references should equal the number of statistically independent sound sources. However, when a holographic measurement is performed, it may be desirable to use a number of real references larger than the number of independent sound sources both to ensure that the total sound field is captured completely and to allow the reduction of noise effects in subsequent processing. In the latter case, the virtual reference signals in excess of the number of independent sources may not be properly determined by using the procedure proposed by Nam and Kim, since those virtual reference signals would be primarily related to measurement noise. Through numerical simulations, Nam and Kim demonstrated that a sound field can be decomposed into physically meaningful partial fields when the virtual references are positioned appropriately on the source surface, and particularly when PCD is performed.

In the present article, the partial field decomposition procedure based on the use of virtual references introduced by Nam and Kim is recast in a form that is based on the use of signal cross-spectral matrices and transfer functions with the intent of developing a procedure that is compatible with conventional multireference NAH techniques.[3] The procedure also incorporates explicit relations between principal reference signals (i.e., noise-free singular values derived from the reference cross-spectral matrix) and partial sound field signals on the hologram and reconstruction surfaces. Thus, the procedure to be described here makes it possible to use a larger number of real than virtual references, as would usually be the case in practice. In addition, a signal processing algorithm that can be used to calculate the MUSIC power anywhere within a three-dimensional sound field reconstructed by NAH projection is introduced. An optimization procedure combined with the latter algorithm can be used to help identify the optimal reference locations: i.e., the optimal reference locations were identified by searching for the positions in three-dimensional space at which the MUSIC power was locally maximized. By using the proposed procedure, it is possible to identify the optimal reference positions in terms of both the projection surface location and the location on the projection surface; in contrast, by simply maximizing the amplitude of a particular acoustical property, e.g., pressure, only the virtual reference location on any particular projected surface can be identified. It is shown here, through both experiment and numerical simulation, that the optimal virtual reference procedure can yield good estimates of the physical partial sound fields. In particular, it is shown that the use of a set of optimal virtual references positioned at the maximum MUSIC power locations results in a more accurate estimation of the physically meaningful partial fields than when the decomposition is based on a set of virtual references located on the source plane.

II. THEORY

A. Multireference NAH

When the total sound field generated by a composite source comprising a finite number of uncorrelated physical sources is completely sensed by a set of reference transducers (i.e., the number of references is equal to or larger than the number of uncorrelated sources and each uncorrelated source is sensed by at least one reference transducer), the field signals on the hologram surface can be expressed in the frequency domain as a linear combination of reference signals multiplied by appropriate acoustical transfer functions:[6,10] i.e.,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1M} \\ h_{21} & h_{22} & \cdots & h_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \cdots & h_{NM} \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_M \end{bmatrix}, \tag{1a}$$

where y_i represents the ith field signal on the hologram surface ($i=1,2,...,N$, and N is the total number of measurement points on the hologram aperture), r_j denotes the jth reference signal ($j=1,2,...,M$, $M \geqslant 1$, and M is the total number of reference microphones), and h_{ij} denotes the transfer function between the jth reference and ith field signal. Equation (1a) can be expressed in compact vector-matrix form as[6,10]

$$\mathbf{y} = \mathbf{H_{yr}} \mathbf{r}, \tag{1b}$$

where $\mathbf{y}$ and $\mathbf{r}$ are the N by 1 field signal vector and the M by 1 reference signal vector, respectively, and $\mathbf{H_{yr}}$ is the N by M

transfer function matrix that relates the reference and field signals. When the holographic measurement is made using a scanning procedure in which data over portions of the hologram aperture are captured sequentially, it is convenient to describe the sound field statistically. In the latter case, it is convenient to multiply Eq. (1b) by the Hermitian transpose of the reference vector and then evaluate the expectation of the result to give[6,10]

$$\mathbf{S_{yr}} = \mathbf{S_{ry}^H} = \mathrm{E}\{\mathbf{y r^H}\} = \mathbf{H_{yr}}\mathrm{E}\{\mathbf{r r^H}\} = \mathbf{H_{yr}}\mathbf{S_{rr}}, \tag{2}$$

where E denotes the expectation operator, the superscript H denotes the Hermitian transpose, $\mathbf{S_{ry}}$ is the cross-spectral matrix between the real reference and field signals, and $\mathbf{S_{rr}}$ is the reference cross-spectral matrix. Here it is assumed that the number of incoherent sources is $K \leqslant M$: the number of incoherent sources in any particular case can, in principle, be determined by counting the number of significant singular values when the reference cross-spectral matrix is decomposed by using SVD.[11] In the latter case, the reference cross-spectral matrix, $\mathbf{S_{rr}}$, can then be represented as the product of a diagonal matrix whose elements are the singular values, and a unitary matrix whose column vectors are the eigenvectors of $\mathbf{S_{rr}}(\mathbf{S_{rr}})^{\mathrm{H}}$. When the singular values associated with the noise subspace are small enough to be ignored, the reference cross-spectral matrix can be approximated as comprising only the first K singular values and eigenvectors: i.e.,

$$\mathbf{S_{rr}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}}, \tag{3}$$

where $\mathbf{V}$ is an M by K matrix whose mth column is the column vector of the unitary matrix and $\mathbf{\Lambda}$ is the K by K diagonal matrix whose mth diagonal element is the mth singular value. By substituting Eq. (3) into Eq. (2), the transfer matrix relating the reference and field signals can be expressed as

$$\mathbf{H_{yr}} = \mathbf{S_{ry}^H}\mathbf{S_{rr}^{-1}} = \mathbf{S_{ry}^H}\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^{\mathrm{H}}. \tag{4}$$

From Eq. (1b), the cross-spectral matrix of the field signals on the hologram surface can be estimated by using the reference cross-spectral matrix in Eq. (3): i.e.,

$$\mathbf{S_{yy}} = \mathbf{H_{yr}}\mathbf{S_{rr}}\mathbf{H_{yr}^H} = \mathbf{H_{yr}}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}}\mathbf{H_{yr}^H}. \tag{5}$$

The cross-spectral matrix of field signals on the hologram surface can then be decomposed to represent a set of incoherent partial fields subject only to the condition that $\mathbf{S_{yy}} = \mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, where each column of the matrix, $\mathbf{Y}$, represents a partial field vector: i.e.,

$$\mathbf{Y} = \mathbf{H_{yr}}\mathbf{V}\mathbf{\Lambda}^{1/2} = \mathbf{H_{yl}}\mathbf{\Lambda}^{1/2}, \tag{6}$$

where $\mathbf{\Lambda}^{1/2}$ is the principal reference signal matrix and $\mathbf{H_{yl}}$ is the transfer matrix between the principal reference signal matrix and the partial field matrix on the hologram surface ($\mathbf{H_{yl}} = \mathbf{H_{yr}}\mathbf{V}$). The acoustical field on any reconstruction surface can then be expressed as

$$\mathbf{Y'} = \mathbf{H_{y'y}}\mathbf{Y} = \mathbf{H_{y'y}}\mathbf{H_{yr}}\mathbf{V}\mathbf{\Lambda}^{1/2} = \mathbf{H_{y'y}}\mathbf{H_{yl}}\mathbf{\Lambda}^{1/2} = \mathbf{H_{y'l}}\mathbf{\Lambda}^{1/2}, \tag{7}$$

where $\mathbf{Y'}$ is the partial field matrix on the reconstruction surface and $\mathbf{H_{y'l}}$ represents the transfer matrix between the principal reference signals and the partial field signals on the reconstruction surface. The matrix $\mathbf{H_{y'y}}$ is composed of the transfer functions that relate the field signals on the hologram and reconstruction surfaces, and it represents the NAH projection procedure, including the spatial Fourier transform (in the planar case) and the propagation operation: the latter procedure also incorporates the effects of discretization, windowing, spatial filtering, zero-padding, etc.[1–3] Note that, in practice, the matrix $\mathbf{H_{yr}}$ appearing in Eq. (7) should be calculated from the cross-spectral matrices, $\mathbf{S_{yr,(step)}}$ and $\mathbf{S_{rr,(step)}}$, measured during each individual scan, by using Eqs. (3) and (4), while $\mathbf{V}$ and $\mathbf{\Lambda}$ should be calculated from the reference cross-spectral matrix, $\mathbf{S_{rr,(avg)}}$, averaged over all the scans in order to suppress spatial noise caused by source nonstationarity (see Ref. 10 for details regarding source nonstationarity compensation). Finally note that each column of $\mathbf{H_{y'l}}$ corresponds to the NAH projection of the corresponding column of $\mathbf{H_{yl}}$. The cross-spectral matrix of the field signals on a reconstruction surface can then be calculated from Eq. (7) as

$$\mathbf{S_{y'y'}} = \mathbf{Y'}\mathbf{Y'}^{\mathrm{H}} = \mathbf{H_{y'l}}\mathbf{\Lambda}\mathbf{H_{y'l}^H}. \tag{8}$$

B. Virtual reference procedure

Virtual references can, in principle, be placed anywhere within the three-dimensional space covered by the NAH projection. By making use of the partial field signal matrices evaluated on the reconstruction surfaces, as in Eq. (7), the virtual reference signal matrix can be expressed as

$$\mathbf{X} = \begin{bmatrix} \mathbf{c}_1^{\mathrm{T}}\mathbf{Y}'_1 \\ \mathbf{c}_2^{\mathrm{T}}\mathbf{Y}'_2 \\ \vdots \\ \mathbf{c}_k^{\mathrm{T}}\mathbf{Y}'_k \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1^{\mathrm{T}}\mathbf{H}_{y'_1 l} \\ \mathbf{c}_2^{\mathrm{T}}\mathbf{H}_{y'_2 l} \\ \vdots \\ \mathbf{c}_k^{\mathrm{T}}\mathbf{H}_{y'_k l} \end{bmatrix} \mathbf{\Lambda}^{1/2} = \mathbf{H_{xl}}\mathbf{\Lambda}^{1/2}, \tag{9}$$

where $\mathbf{X}$ is the K by K virtual reference signal matrix and $\mathbf{H_{xl}}$ represents the transfer matrix between the principal reference signals and the virtual reference signals. In Eq. (9), $\mathbf{c}_m$ represents the N by 1 reference selection vector: when the mth virtual reference is positioned at the ith field position on the reconstruction surface, all elements of $\mathbf{c}_m$ are zeros, except for the element at the ith row, which is itself unity. Note that the matrix $\mathbf{Y}'_m$ in Eq. (9) represents the partial field signal matrix on the reconstruction surface at which the mth virtual reference is placed, and that the vector, $\mathbf{c}_m$, denotes the mth virtual reference location on the mth reconstruction surface. Thus, the location of the mth reconstruction surface in combination with the vector, $\mathbf{c}_m$, determines the location of the mth virtual reference in a three-dimensional space. The cross-spectral matrices between the virtual reference signals, and between the virtual reference and the field signals on the hologram surface, can then be obtained from Eqs. (6) and (9): i.e.,

$$\mathbf{S_{xx}} = \mathbf{X}\mathbf{X}^{\mathrm{H}} = \mathbf{H_{xl}}\mathbf{\Lambda}\mathbf{H_{xl}^H}, \tag{10}$$

and

$$\mathbf{S_{xy}} = \mathbf{X}\mathbf{Y}^{\mathrm{H}} = \mathbf{H_{xl}}\mathbf{\Lambda}\mathbf{H_{yl}^H}, \tag{11}$$

where $\mathbf{S_{xx}}$ is the virtual reference cross-spectral matrix and $\mathbf{S_{xy}}$ is the cross-spectral matrix between the virtual reference and field signals on the hologram surface.

Recall that under ideal circumstances, the physical reference microphones should be placed at the locations of physical, uncorrelated sources. As argued earlier, in practice, it may not be possible to position real reference microphones at those points.

It has, however, been shown previously that the MUSIC algorithm can be combined with the SVD procedure to guide the selection of the best ''real'' reference microphones from among a large number of real microphones positioned around a noise source.[6] Since the MUSIC algorithm can be used to find the locations of sources in a three-dimensional space,[7] the latter procedure has here been modified to identify the optimal virtual reference locations: the optimal locations are here identified as the points in a three-dimensional space at which the MUSIC power, defined below, is maximized.

When the cross-spectral matrix of the field signals on a reconstruction surface, i.e., $\mathbf{S_{y'y'}}$ in Eq. (8), is decomposed by using the SVD procedure ($\mathbf{S_{y'y'}}=\mathbf{W}\Sigma\mathbf{W}^{\mathrm{H}}$), the unitary matrix, $\mathbf{W}$, can be expressed in terms of the eigenvectors of $\mathbf{S_{y'y'}}(\mathbf{S_{y'y'}})^{\mathrm{H}}$: i.e., $\mathbf{W}=[\mathbf{w}_1\mathbf{w}_2\cdots\mathbf{w}_N]$, where $\mathbf{w}_n$ is the nth eigenvector associated with the nth singular value. Since the number of incoherent sources is K, the noise subspace, $\mathbf{R}_{\mathrm{noise}}$, can be defined in terms of the noise-related eigenvectors, $\mathbf{w}_n$ ($n=K+1$ to N), as

$$\mathbf{R}_{\mathrm{noise}}=\sum_{n=K+1}^{N}\mathbf{w}_n\mathbf{w}_n^{\mathrm{H}}. \tag{12}$$

The MUSIC power is then defined in terms of $\mathbf{R}_{\mathrm{noise}}$ as

$$P_{\mathrm{MUSIC}}=\frac{1}{\mathbf{u}^{\mathrm{H}}\mathbf{R}_{\mathrm{noise}}\mathbf{u}}, \tag{13}$$

where $\mathbf{u}$ is the trial vector. Since the signal subspace spanned by $\mathbf{w}_n$ ($n=1$ to K) is orthogonal to the noise subspace represented by $\mathbf{R}_{\mathrm{noise}}$, the MUSIC power should be infinite when $\mathbf{u}=\mathbf{w}_n$ ($n=1$ to K).

Assume, for example, that a vector representing a source at the nth discrete point of the reconstruction surface can be approximated as the trial vector,

$$\mathbf{u}_n=[0\quad\cdots\quad 0\quad 1\quad 0\quad\cdots\quad 0]^{\mathrm{T}}, \tag{14}$$

where the nth element of $\mathbf{u}_n$ is unity and the other $N-1$ elements are zeros. In that case, the MUSIC power associated with the trial vector, $\mathbf{u}_n$, would be very large.

After calculating the MUSIC powers associated with all possible trial vectors, $\mathbf{u}=\mathbf{u}_n$ ($n=1$ to N), a map of the MUSIC power on a two-dimensional reconstruction surface can be obtained by reshaping the MUSIC power vector into a matrix whose elements represent the MUSIC power at a point on a two-dimensional surface. A three-dimensional MUSIC power image can then be obtained by repeating the latter operations on other projection surfaces in sequence. Since the MUSIC power at source locations is large, the optimal virtual reference locations are those where the MUSIC power is maximized locally. The virtual reference signals at those optimal locations can then be obtained from Eq. (9) since the reference locations that are identified by performing the optimal search are expressed in terms of the reconstruction surface locations and the reference selection vectors that appear in the latter equation.

The trial vector defined in Eq. (14) closely represents the sound field generated by a monopole source. It is thus likely that the MUSIC power procedure as described here will accurately identify the locations of monopole-like sources. Other trial vectors may be more appropriate for identifying the locations of more complicated sources, e.g., panel modes, but here attention is limited to trial vectors of the type defined in Eq. (14).

Consider next two monopole sources that are spatially separated but that are coherent: the MUSIC power criterion as defined above would identify sources at both locations. However, in this case there is only one coherent source mechanism: i.e., the number of incoherent sources is smaller than number of local MUSIC power maxima. In the latter case, virtual references should first be placed at all local MUSIC power maxima. Based on the virtual reference cross-spectral matrix, Eq. (10), the coherence between virtual references i and j can be calculated: i.e.,

$$\gamma_{ij}^2=\frac{|S_{ij}|^2}{S_{ii}\cdot S_{ij}}, \tag{15}$$

where S_{ii} and S_{jj} are the autospectra of the virtual references i and j, respectively, and S_{ij} is the cross-spectrum between the virtual references i and j. If the coherence between the signals at two candidate source locations is nearly unity, the sources at those two locations are coherent. The virtual reference at the location at which the MUSIC power is smaller should then be removed.

C. Partial sound field decomposition

Once the optimal virtual reference locations are identified, the virtual reference signals at those positions can be obtained from Eq. (9). In addition, the cross-spectral matrices between the virtual reference signals, and between the virtual reference and field signals on the hologram surface, can then be calculated from Eqs. (10) and (11). The cross-spectral matrix of the field signals on the hologram surface can be expressed in terms of the latter cross-spectral matrices: i.e.,

$$\mathbf{S_{yy}}=\mathbf{S_{xy}^{H}}\mathbf{S_{xx}^{-1}}\mathbf{S_{xy}}. \tag{16}$$

The cross-spectral matrix of field signals in Eq. (16) can be decomposed into a partial field matrix, $\mathbf{Y_x}$, based on the ''virtual'' reference signals subject only to the condition that $\mathbf{S_{yy}}=\mathbf{Y_x}(\mathbf{Y_x})^{\mathrm{H}}$: i.e.,

$$\mathbf{Y_x}=\mathbf{S_{xy}^{H}}\mathbf{S_{xx}^{-1/2}}. \tag{17}$$

Since the decomposition in Eq. (17) is not unique (because the square root inverse of the reference cross-spectral matrix is not uniquely determined), it is next desirable to select ''good'' partial fields from among all the possible ones.

The SVD and PCD procedures have been widely used to perform the partial field decomposition expressed in Eq.

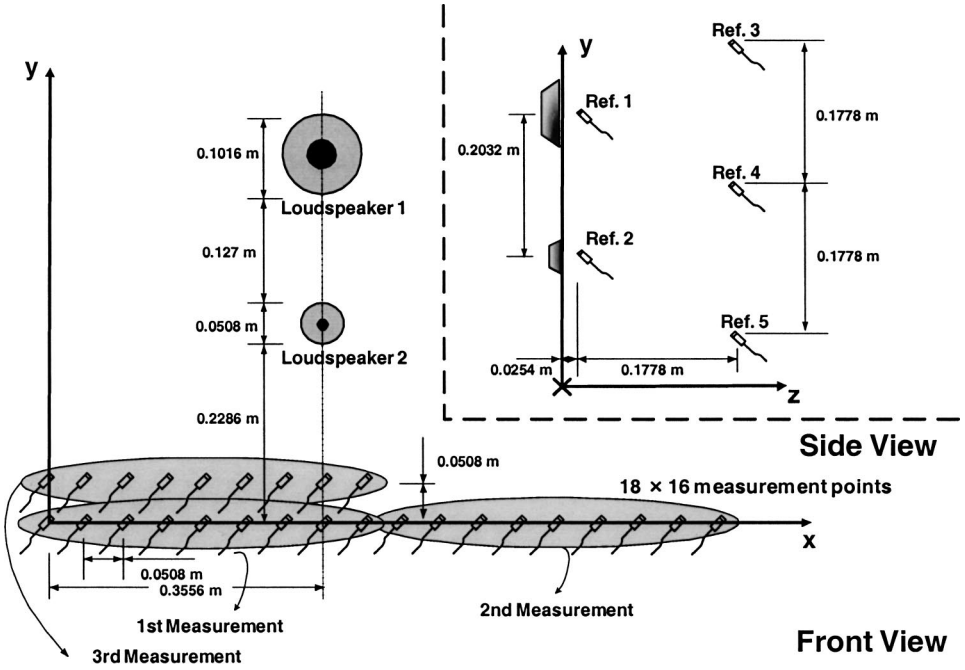


FIG. 1. Sketch of the experimental setup.

(17).[6] In the SVD method, the virtual reference cross-spectral matrix is represented as $\mathbf{S_{xx}}=\mathbf{USU}^{\mathrm{H}}$, where $\mathbf{U}$ is the unitary matrix and $\mathbf{S}$ is the diagonal matrix of singular values. The partial fields can thus be represented as

$$\mathbf{Y}_{\mathbf{x},\mathrm{SVD}}=\mathbf{S}_{\mathbf{xy}}^{\mathrm{H}}\mathbf{US}^{-1/2}. \tag{18}$$

The PCD procedure is based on the use of LU decomposition, obtained from the Gauss elimination process,[5] to separate the reference cross-spectral matrix into two matrices: i.e., $\mathbf{S_{xx}}=\mathbf{LDL}^{\mathrm{H}}$, where $\mathbf{L}$ is the lower triangular matrix whose diagonal elements are unity and $\mathbf{D}$ is the diagonal matrix. In the latter case the partial fields are written as

$$\mathbf{Y}_{\mathbf{x},\mathrm{PCD}}=\mathbf{S}_{\mathbf{xy}}^{\mathrm{H}}\mathbf{L}^{\mathrm{H}^{-1}}\mathbf{D}^{-1/2}. \tag{19}$$

If the partial decomposition is based instead on the use of real reference signals, the cross-spectral matrices in Eqs. (18) and (19) should be calculated using the real rather than the virtual reference signals: i.e., $\mathbf{S_{rr}}$ and $\mathbf{S_{ry}}$ instead of $\mathbf{S_{xx}}$ and $\mathbf{S_{xy}}$. Note that when performing a partial coherence decomposition, the number of references should be equal to the number of incoherent sources.

III. EXPERIMENT AND NUMERICAL SIMULATION

A planar acoustical holography experiment was performed in an anechoic chamber to illustrate the optimal virtual reference procedure described above. Two loudspeakers having different frequency characteristics were driven by independent white noise sources (see Fig. 1).[11] A horizontal line array of nine microphones spaced 0.051 m apart was used to scan the hologram following the procedure indicated in Fig. 1. The aperture thus comprised 32 subholograms and a total of 18 by 16 field points. The source surface (i.e., the $z=0$ m plane) coincided with the front surfaces of the loudspeakers. The field measurements were made on a hologram surface located at $z=0.05$ m. During the scanning, five reference microphones were fixed in front of the loudspeakers, as shown in Fig. 1. In each scan, the data record length was 512 points at a sampling rate of 4096 Hz and 20 linear averages were performed when estimating the various spectra; during the latter operations a 256 point overlap was used and a Hanning window was applied to each record.

A numerical simulation involving three monopole sources was also performed to verify the performance of the optimal virtual reference procedure described above. In the latter case, monopole sources were located at each of the loudspeaker locations in the x-y plane as in Fig. 1, and a third one was located at $(x,y)=(0.457\ \mathrm{m},0.356\ \mathrm{m})$: their z locations were chosen to be -0.005 m, -0.010 m, and -0.015 m, respectively: i.e., slightly behind the nominal source plane at $z=0$ m. References 1, 2 and 4 were placed at the monopole locations on the source plane (i.e., $z=0$ m) while Refs. 3 and 5 were placed at the locations of real Refs. 3 and 5, as shown in Fig. 1. The three monopole source strengths were represented by three independent random signals of approximately equal strength. The data obtained from the simulation was processed in the same way as the experimental data except that the number of virtual references was chosen to be three: the latter number could be determined by counting the significant singular values of the reference cross-spectral matrix.[11]

A source nonstationarity compensation was performed before applying the NAH projection procedure to the hologram data in both the experimental and simulation cases.[11] During the latter procedures, a spatial Tukey window with a 75% flat-top was applied to the hologram data to reduce truncation effects and the hologram aperture size was extended to 36 by 32 by zero padding. No spatial filtering was applied during the projections that were performed as part of the search for the optimal reference locations.

IV. RESULTS

The singular values of the measured reference cross-spectral matrix averaged over all scans are shown in Fig. 2. It

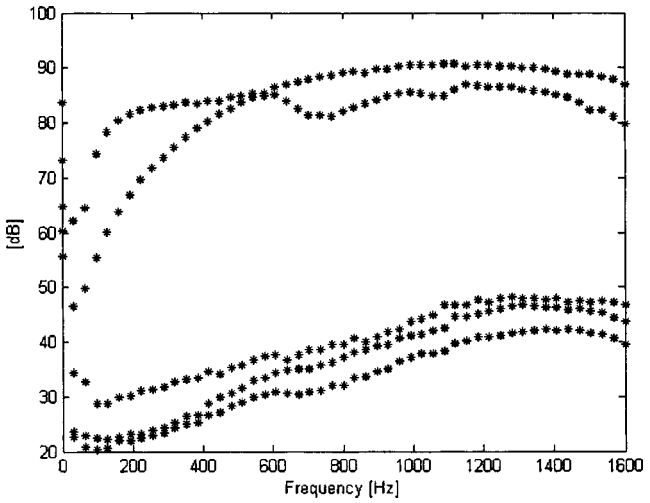


FIG. 2. Spectra of the singular values of the reference cross-spectral matrix (experiment).

is clear that there are two incoherent sources since there are only two significant singular values except at the lowest frequencies (below approximately 50 Hz) in which range the loudspeakers could not be driven effectively.

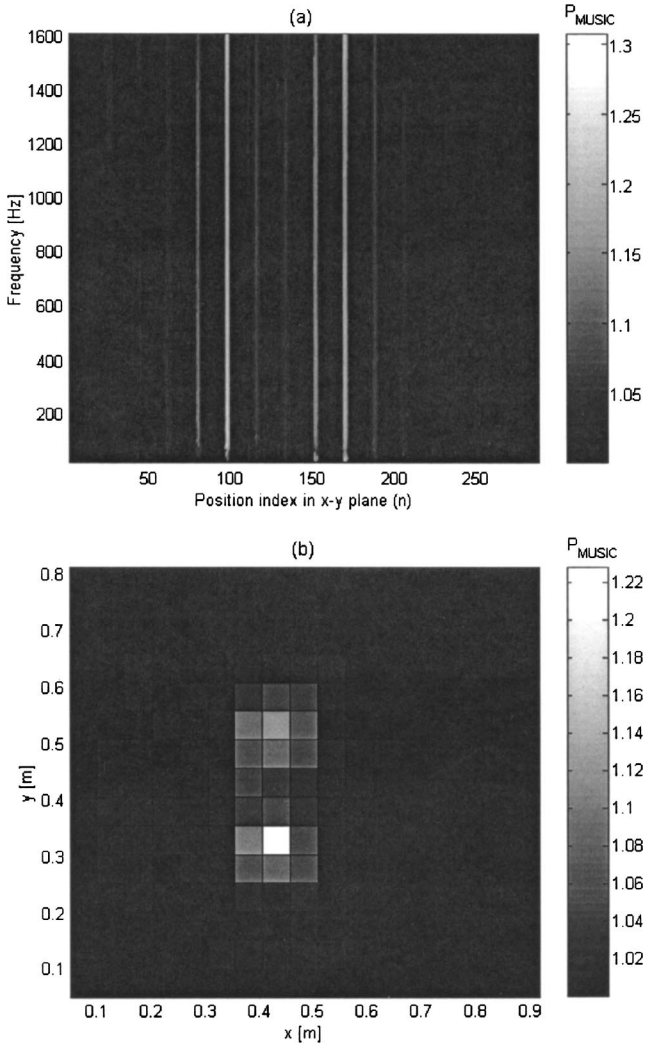


FIG. 3. Music power on the source plane ($z=0$ m) (experiment): (a) from 0 to 1600 Hz (maxima at $n=98$ and 170), and (b) mapped onto the x-y plane at 1200 Hz.

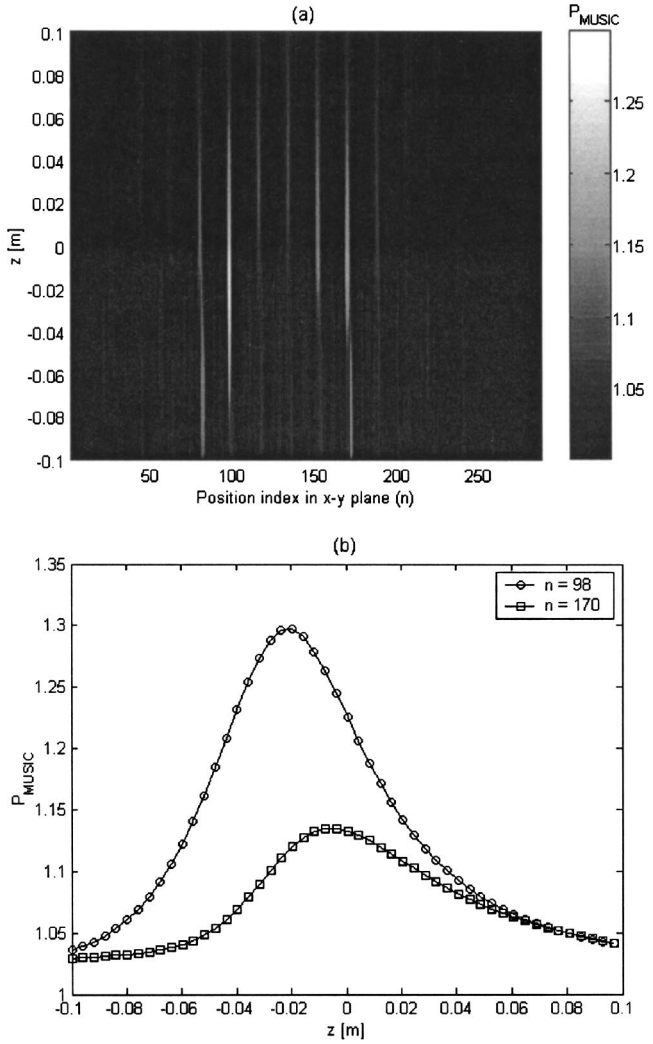


FIG. 4. Music power at 1200 Hz (experiment): (a) as a function of $n=1$ to 288 and $z=-0.1$ to 0.1 m, and (b) plotted from $z=-0.1$ to 0.1 m when $n=170$ [$\mathrm{MAX}(P_{\mathrm{MUSIC}})=1.2976$ at $z=-0.0212$ m] and when $n=170$ [$\mathrm{MAX}(P_{\mathrm{MUSIC}})=1.1351$ at $z=-0.0054$ m].

The MUSIC power on the source plane (i.e., $z=0$ m) is shown in Fig. 3(a) from 0 to 1600 Hz: a measurement point on the x-y plane is represented here by the vector index, n, on the x axis. In Fig. 3(b), the MUSIC power vector at 1200 Hz is plotted on the $z=0$ plane: the vector was reshaped into a matrix whose i-, jth element represents the i-, jth measurement point on the x-y plane. The two loudspeaker locations can easily be identified from the positions of the local maxima of the MUSIC power: those locations correspond to vector indices, $n=98$ and $n=170$, the positions of loudspeakers 1 and 2, respectively. It can also be seen in Fig. 3(a) that there were distinct maxima at $n=98$ and 170 at all frequencies except the very lowest.

The MUSIC power at 1200 Hz was then calculated on projection planes from $z=-0.1$ to 0.1 m, as shown in Fig. 4(a). The MUSIC powers at $n=98$ and $n=170$, i.e., at the two loudspeaker locations as identified from the results of Fig. 3, are plotted as a function of z in Fig. 4(b). Note that the two plots in Fig. 4(b) are cross-sectional views of Fig. 4(a) when $n=98$ and $n=170$. Local maxima of the MUSIC

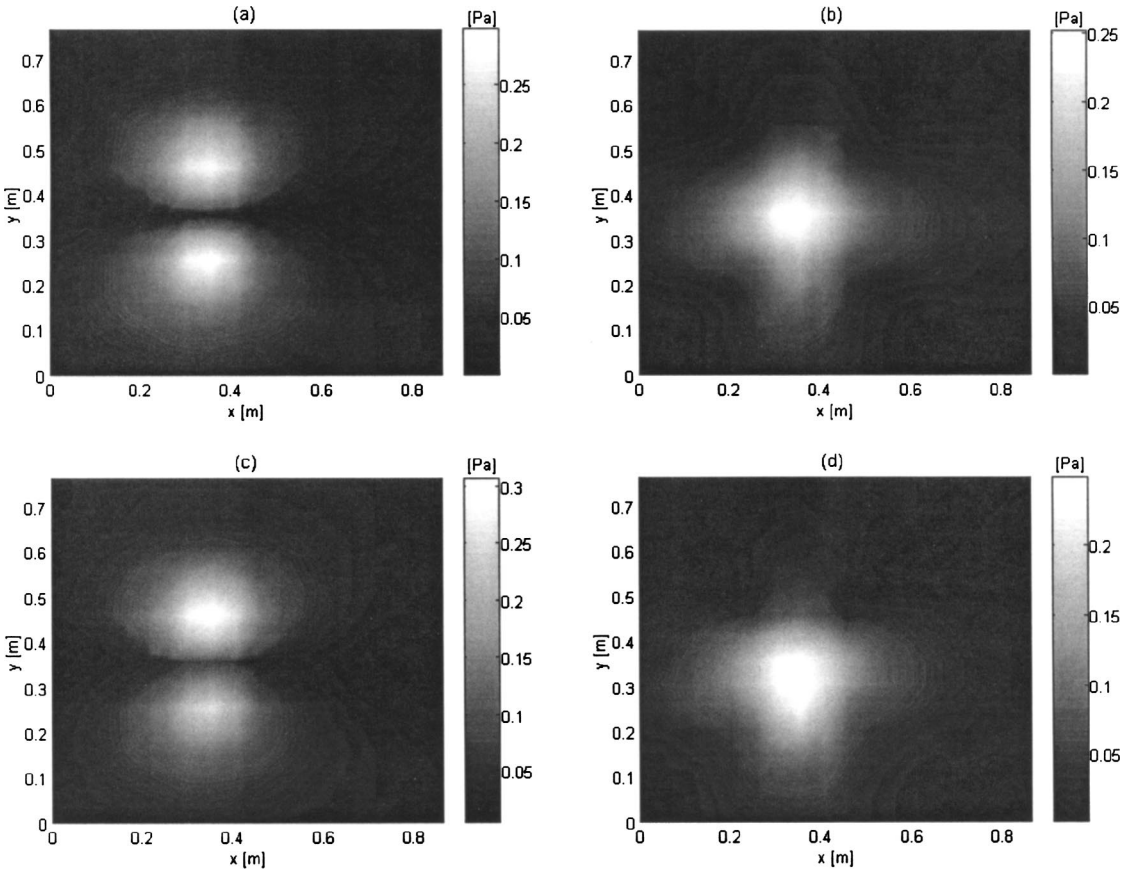


FIG. 5. Amplitudes of decomposed partial pressure fields on the hologram surface at 1200 Hz when the measurement was made using Refs. 3 and 5 (experiment): (a) first field (SVD); (b) second field (SVD); (c) first field (PCD); and (d) second field (PCD).

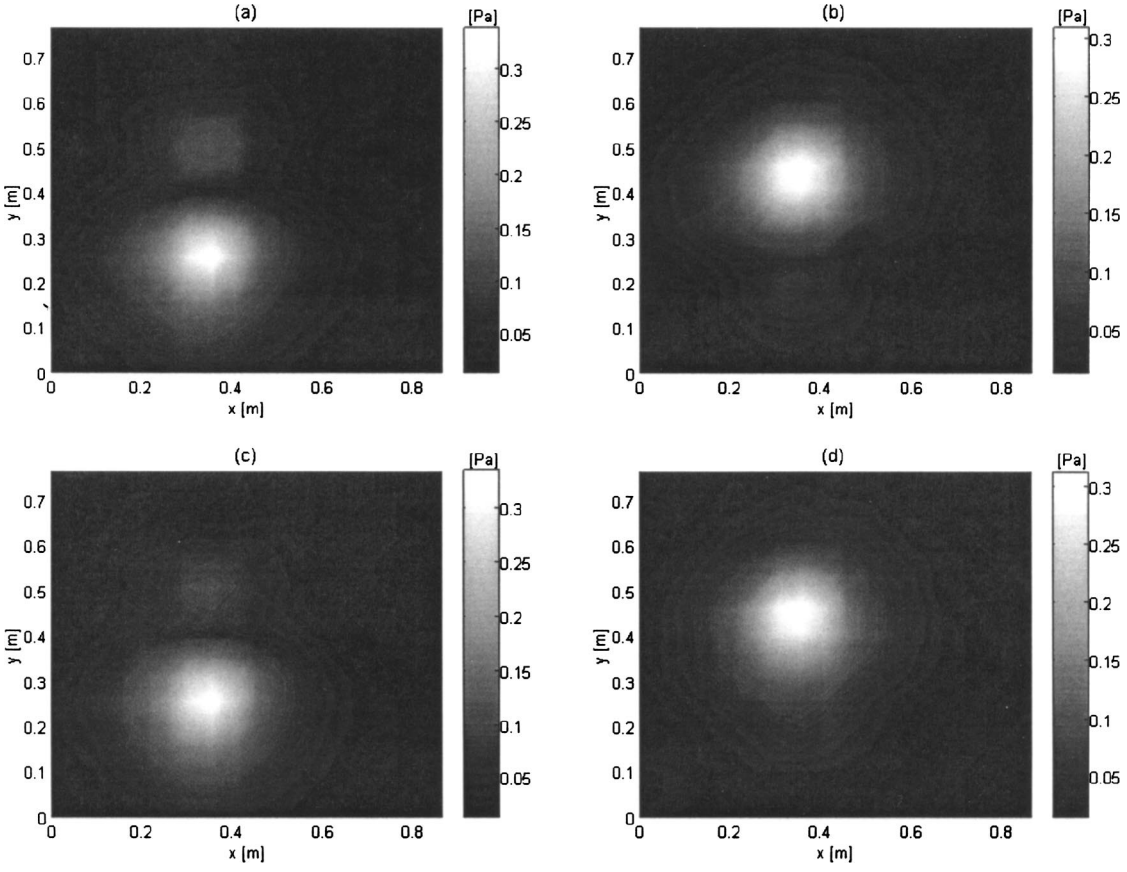


FIG. 6. Amplitudes of decomposed partial pressure fields on the hologram surface at 1200 Hz obtained using optimal virtual references when the measurement was made by using Refs. 3 and 5 (experiment): (a) first field (SVD); (b) second field (SVD); (c) first field (PCD); and (d) second field (PCD).

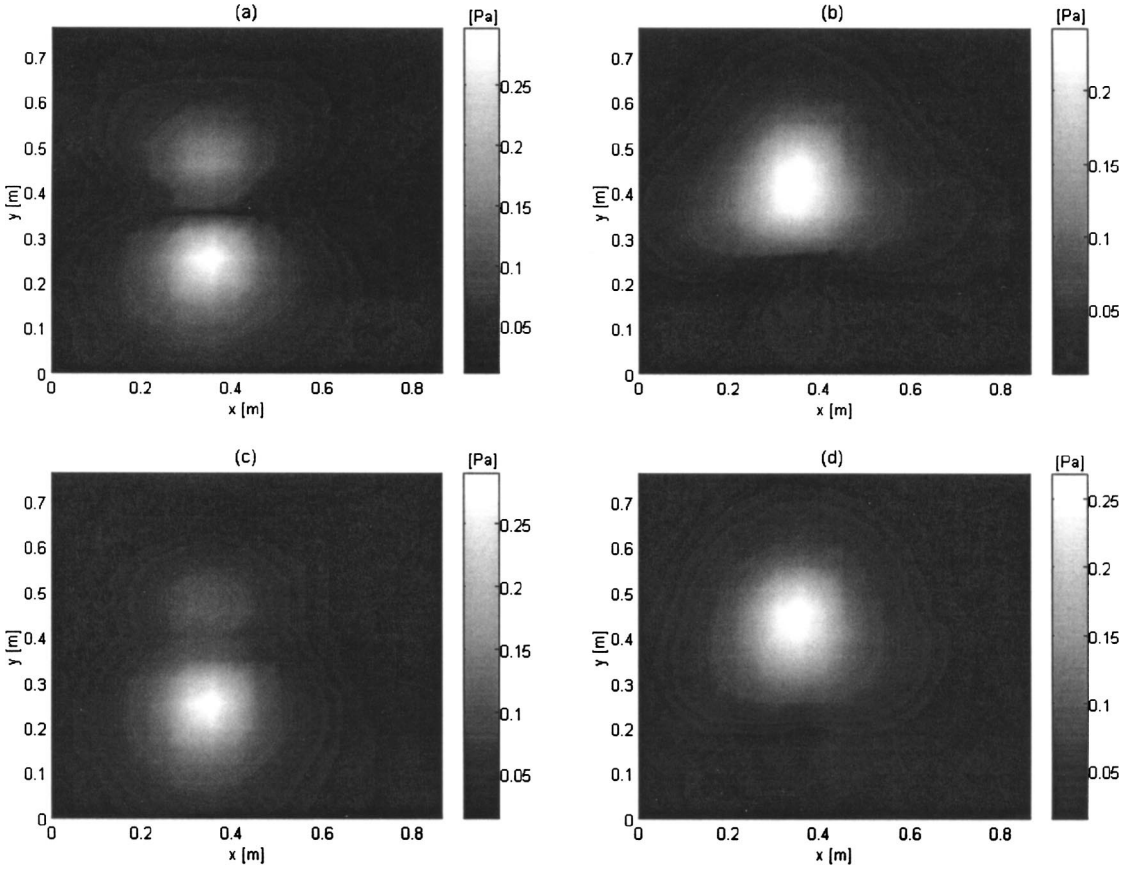


FIG. 7. Amplitudes of decomposed partial pressure fields on the hologram surface at 1200 Hz obtained using the real references when the measurement was made by using Refs. 1 and 2 (experiment): (a) first field (SVD); (b) second field (SVD); (c) first field (PCD); and (d) second field (PCD).

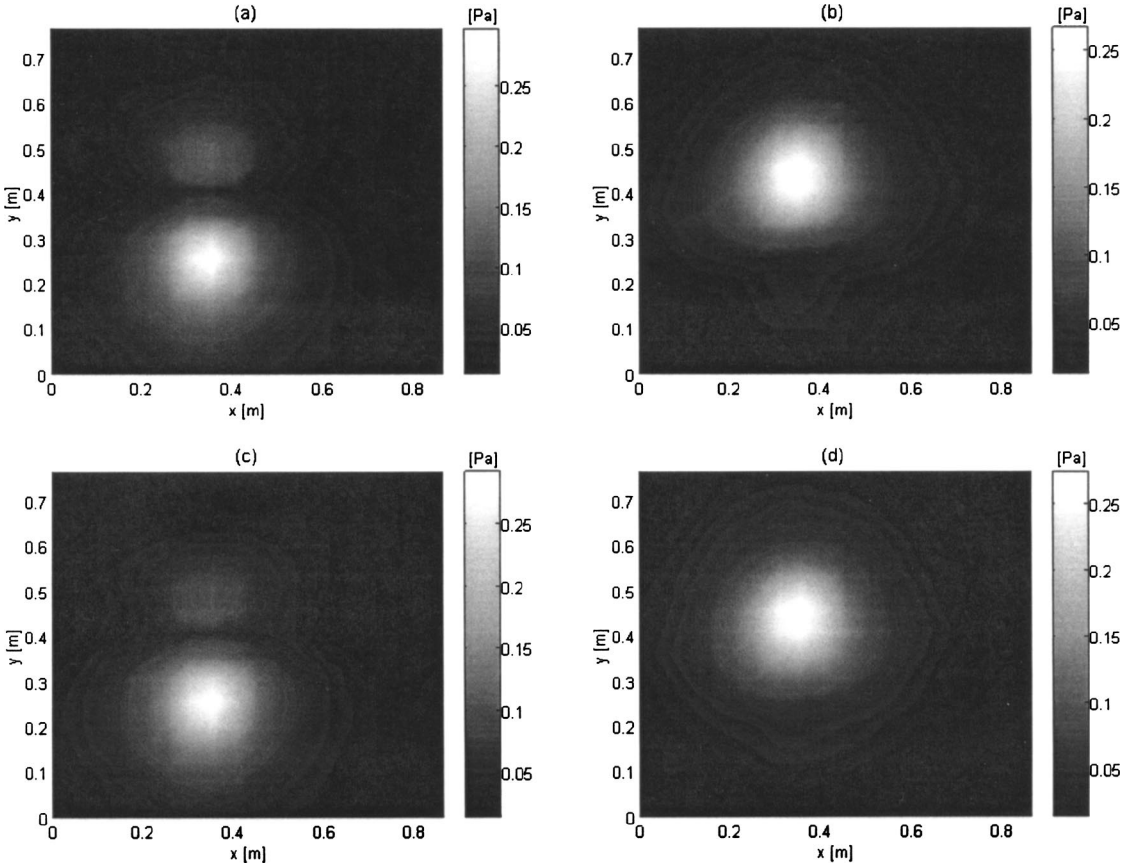


FIG. 8. Amplitudes of decomposed partial pressure fields on the hologram surface at 1200 Hz obtained using optimal virtual references when the measurement was made by using all references (experiment): (a) first field (SVD); (b) second field (SVD); (c) first field (PCD); and (d) second field (PCD).

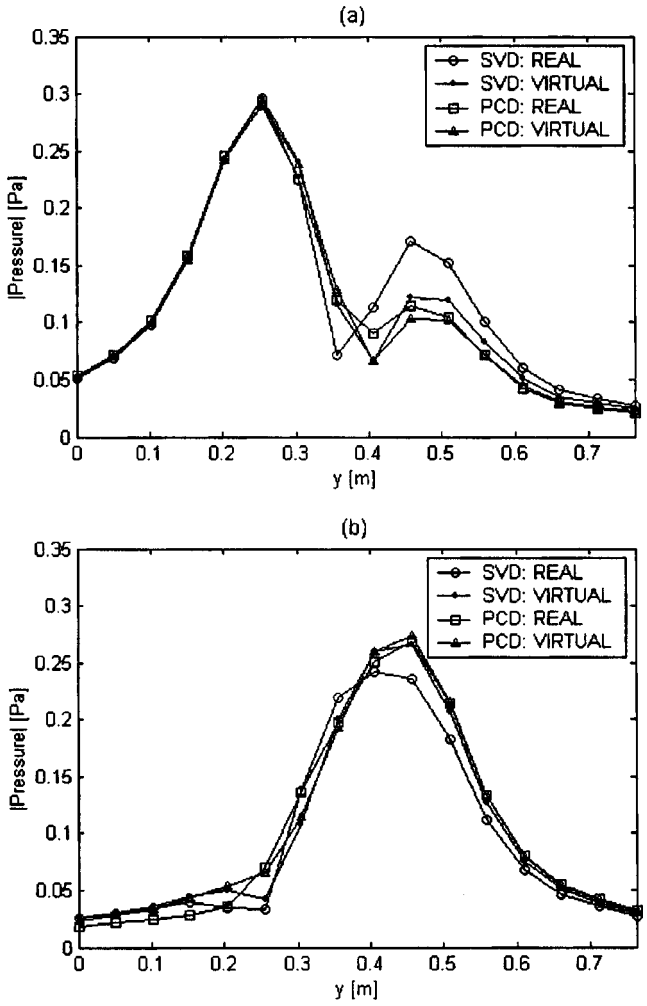


FIG. 9. Amplitudes of partial sound pressure fields at $x=0.356$ m (obtained from Figs. 7 and 8) as a function of y: (a) first partial pressure field, and (b) second partial pressure field.

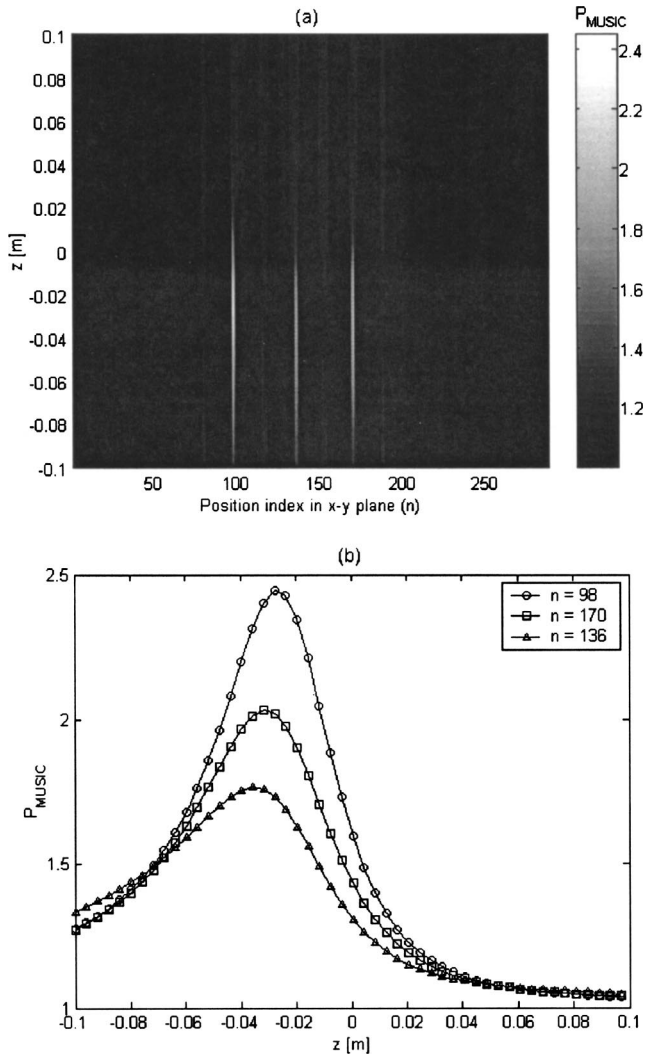


FIG. 10. Music power at 1200 Hz (simulation): (a) as a function of $n=1$ to 288 and $z=-0.1$ to 0.1 m, and (b) plotted from $z=-0.1$ to 0.1 m when $n=98$ [$\mathrm{MAX}(P_{\mathrm{MUSIC}})=2.4484$ at $z=-0.0268$ m], when $n=170$ [$\mathrm{MAX}(P_{\mathrm{MUSIC}})=2.0323$ at $z=-0.0311$ m], and when $n=136$ [$\mathrm{MAX}(P_{\mathrm{MUSIC}})=1.7636$ at $z=-0.0353$ m].

power were found at $z=-0.021$ m for $n=98$ and $z=-0.005$ m for $n=170$. These locations lie slightly behind the nominal source plane at $z=0$ m, and these perhaps coincide with the acoustical centers of the loudspeakers, which were cone shaped.

Since data on the x-y plane are available only at the discrete measurement points, an interpolation should, in principle, be performed to obtain continuous data on the reconstruction plane. To avoid the complexities introduced by an interpolation function, however, it is assumed here that a virtual reference may only be placed at a discrete measurement point in the x-y plane. In the latter case the virtual reference locations on the x-y plane did not change with small variations of the z location [e.g., see Fig. 4(a)], since the x-y resolution, as determined by the distance between the discrete measurement points, was relatively coarse compared with the resolution in the z direction. Thus, the virtual references were first located on the x-y plane at the positions of the maxima of the MUSIC power on the $z=0$ surface: e.g., see Fig. 3(b). The optimal reference locations were subsequently identified as those points at which the MUSIC power was maximized as a function of z at the corresponding location in the x-y plane: e.g., see Fig. 4(b).

The measured data at 1200 Hz were used for the purpose of comparing the characteristics of the partial fields decomposed by using either the SVD or PCD procedures combined with either the set of real or optimal virtual references: the resulting partial pressure fields on the hologram surface (i.e., at $z=0.05$ m) are shown in Figs. 5–8. The partial pressure fields decomposed by using the real and optimal virtual references are shown in Figs. 5 and 6, respectively, when the measurement was made by using real Refs. 3 and 5. Figure 7 shows the decomposed partial fields obtained by using the real references when the measurement was made by using Refs. 1 and 2. The decomposed partial fields obtained by using the optimal virtual references are shown in Fig. 8 when the measurement was made by using all of the real references.

When real Refs. 3 and 5 were used as the basis for the partial field decomposition, the decomposed partial fields do not represent the physical partial fields associated with the loudspeakers, as may be seen in Fig. 5: i.e., both partial fields contain contributions from both loudspeakers. How-

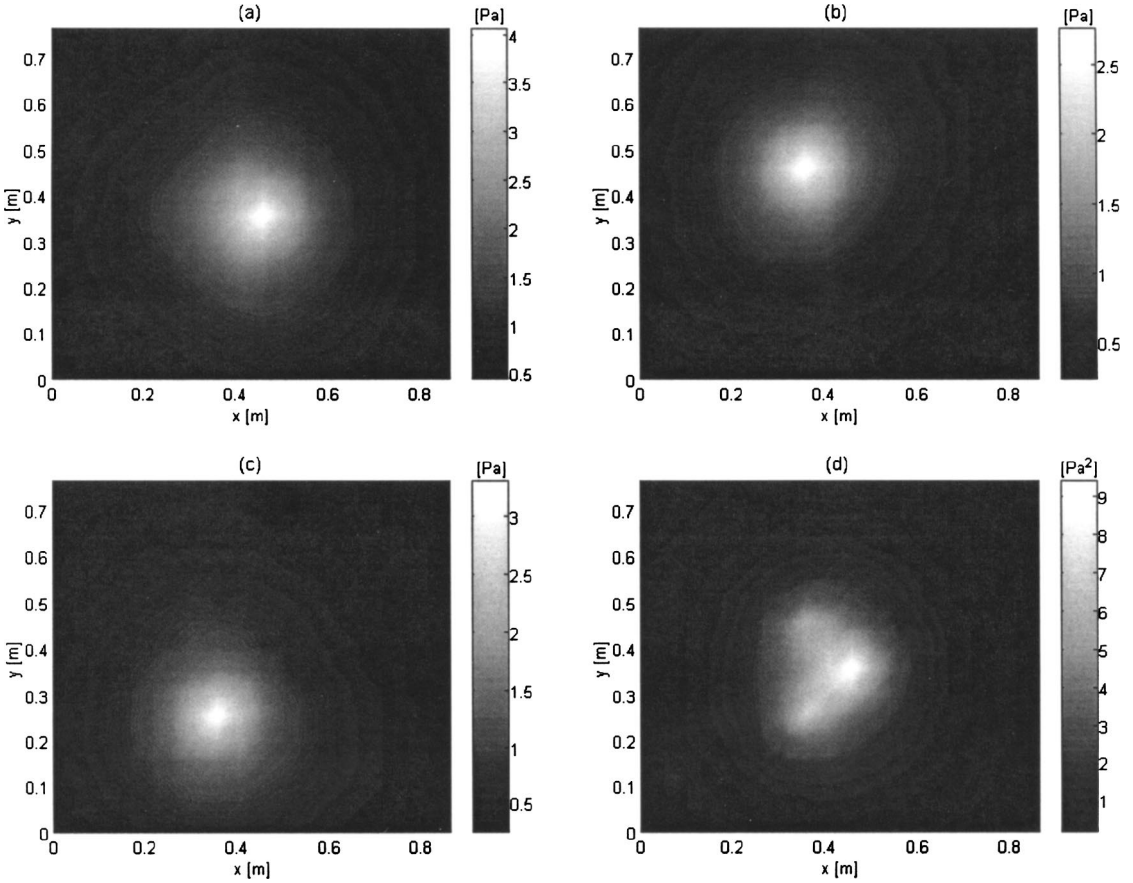


FIG. 11. Decomposed partial pressure fields on the hologram surface at 1200 Hz obtained by using the PCD procedure in combination with optimal virtual references (simulation): (a) first field; (b) second field; (c) third field; and (d) total field (i.e., summed mean square pressures).

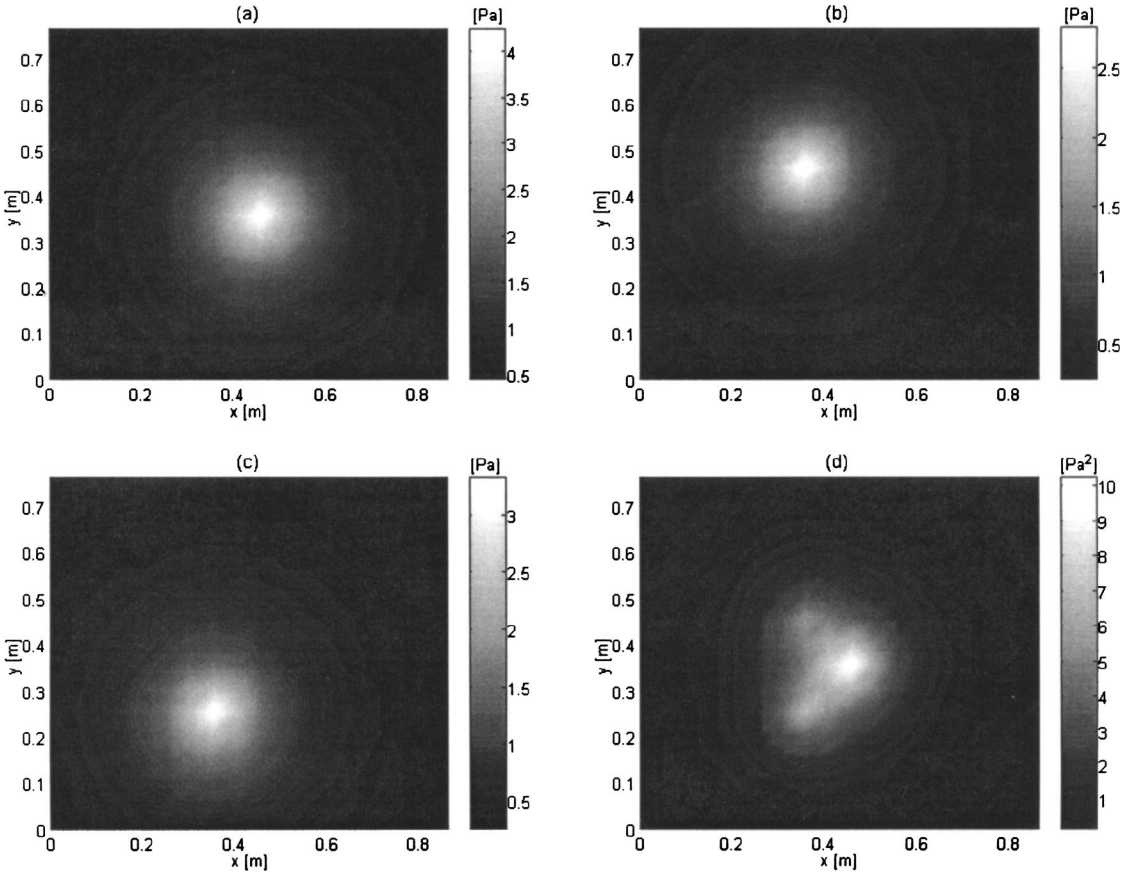


FIG. 12. ''Exact'' partial pressure fields on the hologram surface at 1200 Hz (simulation): (a) first field; (b) second field; (c) third field; and (d) total field (i.e., summed mean square pressures).

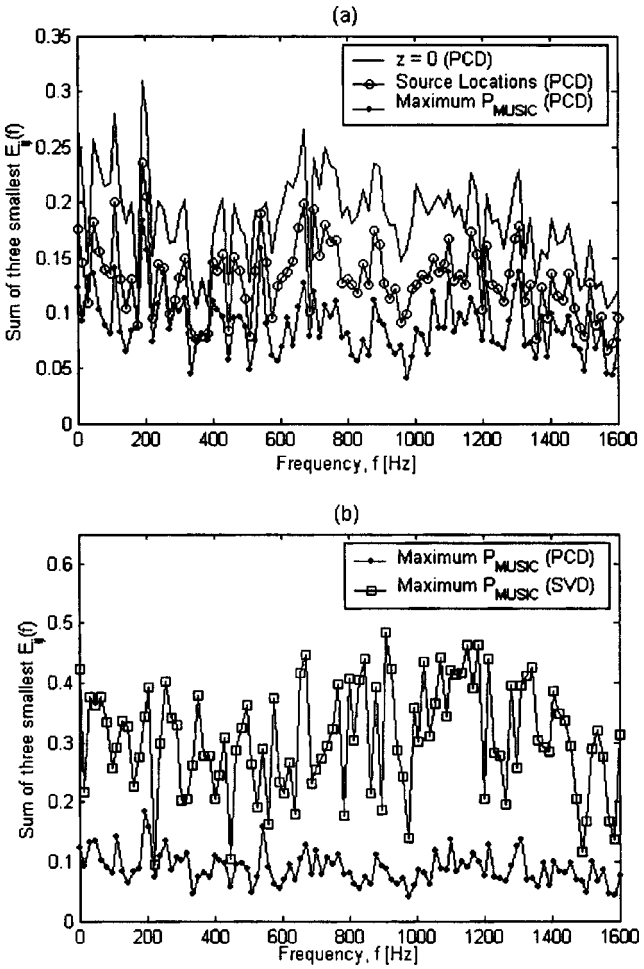


FIG. 13. Error between the exact partial pressure fields and the partial pressure fields decomposed by using virtual references (simulation): (a) when the virtual references were placed on the source plane (PCD), at the source locations (PCD), and at the maximum MUSIC power locations (PCD), and (b) when both PCD and SVD procedures were applied in combination with virtual references placed at the maximum MUSIC power locations.

ever, the partial fields decomposed on the basis of a set either of ''good'' real references positioned close to the loudspeakers (i.e., Refs. 1 and 2) or the optimal virtual references, are closely associated with the physical sources. Note, however, that even in these cases there was a small amount of ''leakage'' from one partial field to another: i.e., a faint trace of the second source is visible in the first partial field [see parts (a) and (c) of Figs. 6–8]. By comparing Figs. 5(a), 5(c), 7(a), and 7(c) for the first partial field or Figs. 5(b), 5(d), 7(b), and 7(d) for the second partial field, it can also be seen that the characteristics of the decomposed sound fields depend on both the reference locations and the decomposition procedure.[6]

In order to compare the results of Figs. 7–8 quantitatively, the amplitudes of the individual partial sound pressure fields at $x=0.356$ m are plotted as a function of the y position in Fig. 9. In the first partial field, it can be seen that there is a contribution from the second loudspeaker at $y=0.457$ m (i.e., the location of loudspeaker 2): the leakage was smaller when the sound fields were decomposed by using the PCD procedure rather than the SVD procedure, and when using the optimal virtual references rather than the real references [see Fig. 9(a)]. That is, the smallest leakage occurred when the PCD procedure was used in combination with the optimal virtual references. Note that the pressure amplitudes at $y=0.457$ m in the second partial field appear in the reverse order of the leakage amplitude in the first partial field [see Fig. 9(b)]. That is, the sound pressure radiated from the upper loudspeaker leaks into the first partial field which is primarily associated with the lower of the two loudspeakers, and is thus lost from the second partial field.

The results of the numerical simulation made using three monopole sources are shown in Figs. 10–13. Figure 10 shows the MUSIC power at 1200 Hz as the location of the projection plane is varied from $z=-0.1$ to 0.1 m. It can be seen in Fig. 10(a) that there are peaks of the MUSIC power at $n=98$, $n=136$, and $n=170$: i.e., at the x-y locations of the three sources. The MUSIC powers at those three locations are replotted in Fig. 10(b): the z locations of the local maxima were found to occur at $z=-0.0268$ m for $n=98$, $z=-0.0311$ m for $n=170$, and $z=-0.0353$ m for $n=136$. Note that the locations identified according to the MUSIC power criterion differ slightly from the real source locations: this discrepancy may result from small errors in estimating the transfer matrix, $\mathbf{H}_{\mathbf{y}'\mathbf{y}}$, representing the NAH projection procedure, owing to discretization, truncation, spatial windowing, etc.

The partial pressure fields decomposed by using the PCD procedure in combination with the optimal virtual references (located according to the results of Fig. 10) are shown in Fig. 11. The corresponding ''exact'' partial pressure fields obtained by turning on only one source at a time are shown in Fig. 12. It can be seen that each of the decomposed partial pressure fields closely resembles the exact partial pressure fields.

To compare the decomposed partial pressure fields and the exact partial pressure fields on a quantitative basis, the error, E_{ij}, between the ith decomposed partial field, p_i ($i=1$ to 3) and the jth exact partial field, P_j ($j=1$ to 3) was defined as

$$E_{ij}(f)=\frac{1}{18\cdot 16}\sum_{n=1}^{18}\sum_{m=1}^{16}\left|\frac{|p_i(f,x_n,y_m)|^2-|P_j(f,x_n,y_m)|^2}{2}\right|, \tag{19}$$

where f is frequency, x_n is the location of the nth measurement point along the x axis, and y_m is the location of the mth measurement point along the y axis. The total error at the frequency, f, is then defined as the sum of the three smallest $E_{ij}(f)$ among all possible combinations of i and j. In Fig. 13(a), the total errors in the cases when the set of virtual references was placed first on the source plane, then at the monopole locations, and finally at the maximum MUSIC power locations are shown: in all cases the PCD procedure was used. The calculation based on the set of virtual references placed at the maximum MUSIC power locations yielded the least error: the error in the latter case is even smaller than when the virtual references were placed at the known source locations. In Fig. 13(b), the partial fields decomposed by using the PCD procedure are compared with those obtained by using the SVD procedure. At all frequen-

cies, the error associated with the PCD procedure was smaller than that resulting from the SVD procedure. Since the source levels were chosen here to be nearly the same, the significant singular values of the optimal virtual reference cross-spectral matrix were close to each other, which may make it difficult for the SVD procedure to reproduce the physical partial fields: the latter effect does not affect the PCD procedure.[6]

V. CONCLUSIONS

In this article, a post-processing procedure has been described that makes it possible to identify virtual reference signals that can be used to identify physically meaningful partial fields *after* performing a holographic measurement based on a nonoptimal, but sufficient, reference set. The optimal virtual references were placed at the positions where the MUSIC power was locally maximized. It was shown through both experiment and numerical simulation that the optimal virtual reference procedure results in partial fields that are very similar to the ''real'' partial fields associated with individual physical sources regardless of the locations of the real references; however, the partial fields decomposed by using the real reference signals directly were found to be physically meaningful only when the references were located very close to the actual sources. It was also found that use of the set of virtual references located at the maximum MUSIC power locations resulted in more accurate estimates of the physically meaningful partial fields than were obtained using any other set of virtual references (e.g., on the source plane or at the actual source locations) or even a set of ''good'' real references. It was also shown that the partial fields obtained using the PCD procedure suffered less leakage than those obtained using the SVD procedure. Thus, it was found that the most meaningful partial fields were obtained when the PCD procedure was used in combination with optimally located virtual references.

[1] J. D. Maynard, E. G. Williams, and Y. Lee, ''Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1413 (1985).

[2] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).

[3] J. Hald, ''STSF—A unique technique for scan-based nearfield acoustical holography without restriction on coherence,'' B&K Technical Review, 1988, Vol. 1.

[4] D. L. Hallman and J. S. Bolton, ''A comparison of multi-reference nearfield acoustical holography procedures,'' Proc. NOISE-CON **94**, 929–934 (1994).

[5] Gilbert Strang, *Linear Algebra and Its Application*, 3rd Ed. (Harcourt, Inc., 1988).

[6] H.-S. Kwon and J. S. Bolton, ''Partial field decomposition in nearfield acoustical holography by the use of singular value decomposition and partial coherence procedures,'' Proc. NOISE-CON **98**, 649–654 (1998).

[7] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice-Hall 1993).

[8] Kyoung-Uk Nam and Yang-Hann Kim, ''Visualization of Multiple Incoherent Sources by the Backward Prediction of Near-Field Acoustic Holography,'' J. Acoust. Soc. Am. **109**, 1808–1816 (2001).

[9] J. M. Danthez and R. Aquilina, ''Separation of Broadband Sources Processing Concept of the Labrador Software,'' Mech. Syst. Signal Process. **11**, 91–106 (1997).

[10] Hyu-Sang Kwon, Yong-Joe Kim, and J. Stuart Bolton, ''Compensation for Source Non-Stationarity in Multi-Reference, Scan-Based Nearfield Acoustical Holography,'' J. Acoust. Soc. Am. **113**, 360–368 (2003).

[11] M. S. Kompella, P. Davies, R. J. Bernhard, and D. A. Ufford, ''A Technique to Determine the Number of Incoherent Sources Contributing to the Response of a System,'' Mech. Syst. Signal Process. **8**, 363–380 (1994).

Bandwidth of spectral resolution for two-formant synthetic vowels and two-tone complex signals[a)]

Qiang Xu,[b)] Ewa Jacewicz, and Lawrence L. Feth[c)]
Department of Speech and Hearing Science, Ohio State University, Columbus, Ohio 43210

Ashok K. Krishnamurthy
Department of Electrical Engineering, Ohio State University, Columbus, Ohio 43210

(Received 20 November 2002; accepted for publication 22 August 2003)

Spectral integration refers to the summation of activity beyond the bandwidth of the peripheral auditory filter. Several experimental lines have sought to determine the bandwidth of this ''supracritical'' band phenomenon. This paper reports on two experiments which tested the limit on spectral integration in the same listeners. Experiment 1 verified the critical separation of 3.5 bark in two-formant synthetic vowels as advocated by the center-of-gravity (COG) hypothesis. According to the COG effect, two formants are integrated into a single perceived peak if their separation does not exceed approximately 3.5 bark. With several modifications to the methods of a classic COG matching task, the present listeners responded to changes in pitch in two-formant synthetic vowels, not estimating their phonetic quality. By changing the amplitude ratio of the formants, the frequency of the perceived peak was closer to that of the stronger formant. This COG effect disappeared with larger formant separation. In a second experiment, auditory spectral resolution bandwidths were measured for the same listeners using common-envelope, two-tone complex signals. Results showed that the limits of spectral averaging in two-formant vowels and two-tone spectral resolution bandwidth were related for two of the three listeners. The third failed to perform the discrimination task. For the two subjects who completed both tasks, the results suggest that the critical region in vowel task and the complex-tone discriminability estimates are linked to a common mechanism, i.e., to an auditory spectral resolving power. A signal-processing model is proposed to predict the COG effect in two-formant synthetic vowels. The model introduces two modifications to Hermansky's [J. Acoust. Soc. Am. **87**, 1738–1752 (1990)] perceptual linear predictive (PLP) model. The model predictions are generally compatible with the present experimental results and with the predictions of several earlier models accounting for the COG effect.
[DOI: 10.1121/1.1624066]

PACS numbers: 43.66.Ba, 43.71.Es, 43.66.Lj [DOS] Pages: 1653–1664

I. INTRODUCTION

The spectral envelopes of spoken vowels normally contain multiple peaks called formants. The interest in approximating phonetic quality by reducing the number of formants in synthetic vowels dates from the early 1950s. Delattre *et al.* (1952) experimented with back vowels whose first two formants are close in frequency and observed that the quality of these vowels is still preserved when a listener is presented with a single intermediate formant. This formant, located somewhere between the two, comprised an overall quality of the vowel, preserving its unique ''color.'' When formant separation was large, such as in front vowels, no single formant could be found that successfully approximated their quality. Delattre *et al.* (1952) concluded that an auditory mechanism must effectively average two formants which are relatively close in frequency. This phenomenon, known today as formant averaging or spectral integration, suggests that the auditory system performs an additional filtering of vowels beyond the level of the cochlea.

A further exploration of the ''center of prominence'' (or ''center of gravity'') of formant cluster in approximating phonetic quality of vowels led to the concept of the perceptually grounded ''effective second formant'' ($F2'$). $F2'$ was to substitute for both $F2$ and higher formants of a natural vowel in their two-formant approximation represented by $F1$ and $F2'$. Empirical formulas for $F2'$ computations from formant frequency values were proposed (Fant, 1959; Carlson *et al.*, 1970, 1975; Bladon and Fant, 1978). Center of gravity (COG) was understood as a possible correlate of $F2'$ (Carlson *et al.*, 1970). Further work on COG showed that when two formant peaks are separated by less than 3.5 bark, they are integrated into a single perceived peak, called the ''perceptual formant'' (Chistovich *et al.*, 1979). The next step in the conceptual development was to link the 3.5-bark ''critical distance'' with predicting $F2'$ values (Bladon, 1983), which resulted in further computational models of $F2'$ (Escudier *et al.*, 1985; Mantakas *et al.*, 1988). The model by Mantakas *et al.* was later applied to predicting the

[a)]Portions of this work have appeared in the first author's Master's thesis and in poster presentations given at the 3rd joint meeting of the Acoustical Society of America and Acoustical Society of Japan, 1996, Honolulu, Hawaii, and at the 133rd meeting of the Acoustical Society of America, 1997, State College, Pennsylvania.

[b)]Current address: Nokia Product Creation Center, NMP Dallas, Irving, TX 75039.

[c)]Author to whom correspondence should be addressed. Electronic mail: feth.1@osu.edu

organizational trends in vowel systems of human languages (Schwartz *et al.*, 1997).

This line of research has established the validity of $F2'$ as a perceptual parameter in approximating vowel quality in two-formant models. It also brought to light the hypothesis that the auditory system performs a ''large-scale spectral integration'' over a limited frequency range of 3.5 bark (e.g., Schwartz and Escudier, 1989). This work focused necessarily on formant frequency values, assuming only implicit knowledge of the relation between the relative amplitudes of the formants. The role of formant amplitudes in spectral integration was explicitly addressed in a series of other studies (e.g., Bedrov *et al.*, 1978; Chistovich and Lublinskaja, 1979; Chistovich *et al.*, 1979; Chistovich, 1985). This line of research investigated the COG *effect*, i.e., the process restricted to a limited frequency range of 3.5 bark. Accordingly, when two existing vowel formants are perceptually integrated, the frequency of the perceived peak (''perceptual formant'') is closer to that of the stronger formant. The amplitudes of the two formants play an important role in that a change in their ratio is equivalent to a *frequency* change of a single-formant vowel which approximates their quality (Chistovich and Lublinskaja, 1979). The 3.5-bark critical distance indicates a possible limit on spectral integration in that the COG effect disappears with larger formant separation.

The present study is a continuation of the latter line of research. It first re-examines the COG effect in the traditional format of two-formant vowels to verify the role of the critical separation of 3.5 bark in spectral integration (experiment 1). It then presents experimental evidence from psychoacoustics, showing that spectral integration in two-tone complex signals occurs within the limit of 3.5-bark resolution bandwidth (experiment 2). It advocates the view that spectral integration is a fundamental property of the auditory system. Further, the 3.5-bark critical distance is used by the auditory system in general signal processing, and is not restricted to speech perception. The focus of this work is not on approximating the phonetic vowel quality but on the auditory processing of complex signals before higher-level decision processes apply, such as making phonetic decisions. The following sections sketch the background of the study.

A. The COG effect in two-formant vowels and the 3.5-bark critical distance

Chistovich and Lublinskaja (1979) reported that their two listeners matched a single-formant vowel of variable frequency (F_v) to a frequency between $F1$ and $F2$ of a two-formant reference signal, when the distance between the two formants was less than about 3.5 bark. The locus of F_v was called the ''perceptual formant.'' Depending on $A2/A1$ ratio (the ratio of the formant amplitudes), the matches fell midway between the frequencies of the two formants ($A2/A1 = 0$ dB) or were directed towards either one of the stronger formants ($A2/A1 = -20$ or 20 dB). When the separation between $F1$ and $F2$ exceeded 3.5 bark, there was no agreement as to the frequency of matching. One listener matched the single formant to one of the two actual formants and the second performed highly unreliable matches to intermediate frequencies between the two. Chistovich and Lublinskaja (1979) concluded that spectral integration fails for more than critical formant separation, i.e., for more than 3.5 bark.

Chistovich (1985) proposed a method for predicting listeners' matches to a two-formant reference vowel. The perceptual formant was defined as the centroid of the vowel spectrum weighted on the bark scale rather than absolute frequency (in hertz). However, this centroid measurement was only good for predicting results for signals with closely spaced formants. Chistovich and Lublinskaja's COG and critical distance effects were well manifested when listeners matched a single-formant signal of variable frequency to a two-formant reference. Results contrary to the COG effect were obtained, subsequently, in which multiformant vowels were used as both reference and variable signals (Beddor and Hawkins, 1990; Assmann, 1991).

Testing the interaction of frequency and amplitude in low-frequency formants predicted by the COG effect, Assmann (1991) carried out two experiments. Using additive harmonic synthesis to control for altering the levels of adjacent harmonics in the region of the formant, he manipulated the amplitude of the first two closely spaced formants in back vowels. The experimental tokens (both the matching and the reference stimuli) included six formants to achieve more natural-sounding vowels. In one experiment, listeners adjusted the frequency of both $F1$ and $F2$ in the target tokens to match the reference tokens with modified amplitudes of $F1$ and $F2$. In the second task, listeners identified the tokens with modified amplitude ratios as instances of five selected English vowels. The stimuli were presented in two conditions, with low f_0 (125 Hz) and high f_0 (250 Hz). Results from the first experiment failed to replicate the findings from the single-formant matching experiments as in Chistovich and Lublinskaja (1979). Support for the COG effect was also weak from the results of the second experiment as some shifts in vowel quality were only obtained in the high f_0 condition.

An important finding from Beddor and Hawkins' study (1990) was that the influence of the spectral envelope on perceived vowel quality was greater when spectral peaks in the low-frequency region were less pronounced. That is, no COG effect was observed for oral vowels with well-defined peaks, whose quality was perceptually determined by the frequency of $F1$. For nasal vowels with less-visible peaks and valleys, the matches were more consistent with the COG effect and listeners relied less on the frequency of $F1$. Their third experiment addressed the issue of the effect of well- versus poorly defined formants on the perceived vowel quality by manipulating formant bandwidths. The results indicated that, in perceiving vowel quality, listeners relied on formant frequency for vowels with well-defined peaks (the narrow-bandwidth condition, BW=45 Hz), whereas the combination of both frequency and bandwidth influenced the perception of vowels with poorly defined peaks (the wide-bandwidth condition, BW=150 Hz).

In summary, the results from experiments in which a single formant was adjusted to match reference signals supported the COG effect. However, in the experiment where two formants were adjusted simultaneously to match the reference signal, the results were incompatible with the pre-

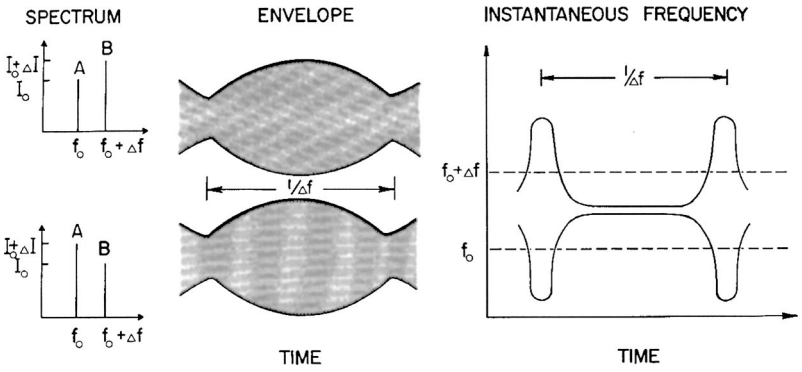


FIG. 1. Long-term spectrum, envelope, and instantaneous frequency functions for both members of the two-component complex tones that make up the simplest set of common envelope signals, as defined by Voelcker (1966a). From Feth and O'Malley (1977) with permission.

dicted pattern. This suggests that the COG effect may not directly contribute to the phonetic perception of a full vowel spectrum, playing a different role in speech signal processing. Consequently, listeners may use different strategies in responding to different types of signals in matching tasks.

B. Two-tone complex signals and the 3.5-bark resolution bandwidth

Independently of the development of both the COG effect and the spectral centroid model by the Leningrad group, Feth (1974) and Feth and O'Malley (1977) studied spectral integration using psychoacoustic signals. This research investigated the pitch of two-tone complexes, following Voelcker's unified theory of modulation (1966a, 1966b). Voelcker began with the envelope–fine-structure representation of a signal

$$S(t)=E(t)*\cos(\varphi(t)), \tag{1}$$

where $E(t)$ is the envelope of the signal, and $\varphi(t)$ is the angle. The envelope function represents slow variations in the amplitude of the signal and is often characterized as the result of amplitude modulation. The first time derivative of the angle function is often defined as the instantaneous frequency, $f(t)$. Voelcker defined sets of common envelope signals that have identical envelope functions but different fine structure. The simplest set of common envelope signals is shown in Fig. 1. For each signal, two sinusoids are separated by a small difference in frequency, ΔF, and their amplitudes differ by a small amount, ΔI. For one signal, the lower-frequency component is more intense; for the complementary signal the higher-frequency component has the higher intensity. It can be shown that these two-component signals have identical envelope functions, as shown in the middle panel of Fig. 1. The instantaneous frequency functions change in opposite directions. When the lower component is the more intense, $F(t)$ moves to lower frequencies as the envelope moves toward its minimum value. When the higher-frequency component is more intense, the frequency modulation moves toward higher frequencies. Listeners hear these fine-structure differences as small changes in the spectral pitch of the signal (Helmholtz, 1954; Jeffress, 1968).

Feth and O'Malley (1977) used the discriminability of complementary, two-component complex tones suggested by Voelcker to investigate the spectral resolving power of the auditory system. They reported that the percentage of correct responses, $P(C)$, in a 2IFC task increased from chance to approach 100% for moderate separations (1 to 3 bark) of the two-component frequencies. However, further separation of the components led to decreased discriminability when the components were apparently resolved by the peripheral auditory filter. The frequency separation for which the Voelcker signals become indiscriminable was suggested as a psychophysical estimate of auditory spectral resolving power. For each center frequency tested, this estimate is approximately 3.5 bark. Thus, in two different experimental paradigms, i.e., vowel matching and two-tone discrimination, the 3.5-bark critical separation plays a role. The critical distance can be viewed as a limit to the range of spectral integration that the auditory system can perform.

In an earlier study, Feth (1974) had proposed that the perceived pitch of a two-tone complex with components of unequal amplitude should correspond to the envelope-weighted average of the instantaneous frequency (EWAIF) of the waveform. According to the EWAIF model, the instantaneous frequency occurring where the envelope was large would contribute more to the perceived pitch of the two-tone complex. The EWAIF model accounted for experimental data using narrow-bandwidth complex tones. Its later version, the intensity-weighted average of instantaneous frequency model (IWAIF) (Anantharaman *et al.*, 1993), is a simple mathematical variation of the EWAIF that eliminates the computational difficulty of the EWAIF calculation. Anantharaman *et al.* (1993) showed that the IWAIF formulation can be transformed into the frequency domain where the IWAIF value corresponds to the COG of the signal spectrum. Given that speech signals are essentially dynamic and complex, changing in both frequency and amplitude over time, the IWAIF model can be viewed as a candidate for testing the COG effect in vowels. IWAIF model predictions were verified for the stationary two-formant speech signals used in Chistovich and Lublinskaja (1979) by Anantharaman (1998).

C. Modeling the COG effect

Hermansky (1990) proposed the perceptual linear prediction model (PLP) to account for spectral integration in vowel-like signals. The model is consistent with both the $F1-F2'$ concept and with the 3.5-bark auditory integration theory. A speech signal is first processed to produce the so-called auditory spectrum by using three processes derived from psychoacoustics: critical band frequency resolution, equal loudness compensation, and intensity-loudness com-

pression. This auditory spectrum is then approximated by an autoregressive all-pole filter, just as in traditional LP modeling, to produce a low-dimensional representation of the speech-signal spectrum. Hermansky suggested that a fifth-order PLP model could extract at most two major peaks from the auditory spectrum, removing the minor spectral peaks. Experimental results from natural and synthetic speech signals confirmed model predictions. When two vowel formants were far apart, the PLP estimated two distinct spectral peaks. With smaller separation of formants of about 3.5 bark, these peaks merged into one peak.

In this paper, we use a modified PLP model to predict and test the COG effect, taking into account the amplitude ratio of two closely spaced formants. In so doing, we investigate whether the COG effect disappears with the 3.5-bark separation of the formants, which would be predictable by the model. Since we manipulate the amplitude ratio in testing the COG effect, we introduce two more stages as our modification to Hermansky's original model, i.e., a peak detection (stage 6) and a decision maker (stage 7). An additional modification is the use of a fourth-order LP model spectrum instead of the fifth-order LP filter because our stimuli do not have more than two formants. Consequently, the number of detected peaks at stage 6 is likely to be either one or two depending on the separation of the two formants and their relative levels. We use the modified PLP model to make predictions of the perceptual formant (F_v) as a function of formant frequency separation and amplitude ratio of a two-formant synthetic signal. We hypothesize that if only one local peak is detected, listeners match it at the F_v. If two peaks are detected, the decision about F_v is based on the frequency-level vector of two local peaks returned from the stage of peak detection. A more detailed description of the model is presented in Xu (1997).

In testing the COG effect in experiment 1, the modified PLP model is used to predict listeners' performance in matching an adjustable, single-formant vowel signal to two-formant vowel signals. The model predicts that

(1) listeners will match single-formant variable signals to the COG of two-formant reference vowels (i.e., F_v changes systematically with respect to the relative level $A2/A1$) only when the distance between $F1$ and $F2$ is smaller than the critical distance of 3.5 bark; and
(2) when the distance between $F1$ and $F2$ exceeds 3.5 bark, listeners match F_v either closer to $F1$ or $F2$ or exhibit chance performance.

Experiment 1 also investigates the effect of spectral shape (poorly defined spectral prominence versus well-defined spectral prominence) on the matching values of F_v; and how the type of excitation source (noise versus pulse train) affects the matching values of F_v.

II. EXPERIMENT 1: COG EFFECT IN TWO-FORMANT SYNTHETIC VOWELS

A. Methods

1. Listeners

Three young adults, two female and one male, participated. Two listeners were volunteer graduate students in speech and hearing science and one was an undergraduate student paid for her participation. All listeners showed normal hearing at all audiometric frequencies on two screening tests: a tone-threshold test in quiet and a frequency difference limen (DLF) test. All listeners were given extensive practice before participating in the experiment. Each participant listened 2 h per day, 5 days per week, for a period of several weeks. All experimental data were collected after their response patterns in the adaptive-tracking procedure were convergent and stable.

TABLE I. Synthesis parameters for two-formant reference vowels used in experiment 1.

Vowel	$F1$ (Hz)	$F2$ (Hz)	$F2-F1$ (bark)	BW1 (Hz)	BW2 (Hz)	Source
1	800	1300	2.5	80	80	Impulse
2	800	1300	2.5	80	80	Noise
3	800	1300	2.5	45	45	Noise
4	800	1300	2.5	150	150	Noise
5	800	1400	3.0	80	80	Noise
6	700	1400	3.6	80	80	Noise
7	700	1500	4.0	80	80	Noise
8	600	1700	5.3	80	80	Noise

2. Stimuli

Two-formant vowel-like signals (reference vowels) were generated using a formant synthesizer in parallel configuration. The excitation source was either an impulse train ($f_0 = 100$ Hz) or broadband, Gaussian noise. The $F2-F1$ distance was varied in five conditions: 1300–800, 1400–800, 1400–700, 1500–700, and 1700–600 Hz, equivalent to 2.5, 3.0, 3.6, 4.0, and 5.3 bark, respectively. For the 2.5-bark separation, four reference vowels were created with either a different excitation source or different bandwidths. For the remaining four conditions, only the formant frequency separation varied and the other parameters remained constant. A summary of the parameters is given in Table I. The relative level $A2/A1$ of the two formants was manipulated from -20 dB to $+20$ dB in steps of 10 dB. This resulted in five relative amplitude conditions for each reference vowel.

The output sequence of the variable single-formant signals was obtained by passing a source through a second-order IIR filter with the required $F1$ and BW1. In order to generate a sequence with two formants, the source was passed through a second IIR filter ($F2$, BW2). The output of the second filter was then scaled and added to that from the first formant. The scale factor ($A2/A1$) denotes the level of the second formant peak relative to that of the first.

Both the reference (two-formant) and the variable single-formant target signals were generated by a TDT system II D/A board controlled by a laboratory PC (166-MHz Pentium) at a sampling frequency of 20 kHz. Signal duration was 300 ms. After being shaped by a 5-ms cosine rise–fall window, the signals were attenuated by a TDT programmable attenuator and filtered by a passive analog filter (TTE), with cutoff frequency at 8 kHz. Signals were played, monaurally, through the right channel of a Sennheiser HD-414-SL headset at 65 dB SL.

The use of noise-excited signals was motivated by the following two reasons. First, when synthetic vowels are impulse-train excited, the interaction between source harmonics and formant frequencies on the effective bandwidth may become significant. The target spectral shapes of both reference and variable signals may be distorted, especially when the frequencies of formant peaks are not in alignment with source harmonics and the bandwidths of those formants are relatively narrow. This suggests that noise-excited synthetic vowels might be better suited for this kind of vowel-matching task. A second advantage of using noise excitation is to encourage subjects to focus on the formant frequency comparison rather than on the fundamental frequency.

3. Model predictions

The modified PLP model was applied to each two-formant reference vowel. Selected results for vowel # 1 (2.5-bark separation, pulse-train-excited) are displayed in Figs. 2(a)–(c), which shows spectra for three relative formant levels $A2/A1 = -20$, 0, $+20$ dB, respectively). The acoustic spectrum (dashed line), auditory spectrum (dashed-dotted line), and PLP-model spectrum (solid line) are plotted together. The maximum level for each plot is aligned at 0 dB for comparison. Essentially, the auditory spectrum reduces the contrast between the peaks and valleys of the acoustic spectrum (note, however, that the two peaks are still visible). The model smoothes the auditory spectrum further so that in each condition, there is only one peak. In (a) ($A2/A1 = -20$ dB), the single peaks of the model spectra fall in the vicinity of $F1$ (800 Hz), while in (c) ($A2/A1 = +20$ dB), they are close to $F2$ (1300 Hz). In (b) ($A2/A1 = 0$ dB), the F_v prediction is between $F1$ and $F2$ but is skewed toward the lower frequency.

The simulation results for the noise-excited vowel # 2 are similar to those for vowel # 1. The only difference is that the model spectra have even broader peaks for the noise-excited source than for the pulse-excited source (see Xu, 1997, for more details). In simulations with supracritical separation of formants (e.g., vowel # 7 with 4.0-bark separation), two peaks are detected in the 0- and +10-dB conditions. The difference between the two peaks is greater than 1 dB. Thus, the predicted value of F_v is the higher peak near $F2$ in both conditions. In the remaining three $A2/A1$ conditions (i.e., -20, -10, and $+20$ dB), only one peak is detected, and the predicted F_v is the frequency at that peak for each condition.

Note that the model spectrum more likely results in two peaks whenever the two formants of the original signal are of equal level (i.e., $A2/A1 = 0$ dB). For a comparison across frequency separations, the results for vowels # 2, # 6, and # 8 each with $A2/A1 = 0$ dB are shown in Fig. 3. For the subcritical separation shown in (a) (2.5 bark, vowel # 2), only one peak it is detected in the model spectrum. For the two supracritical separations depicted in (b) (3.6 bark, vowel # 6), and (c) (5.3 bark, vowel # 8), two peaks are detected successfully, although the second peak in (b) is not separated by a valley as it is in (c). In the latter plot, the second formant peak is more than 1 dB higher in level than the first. The trend observed with the increasing distance between the

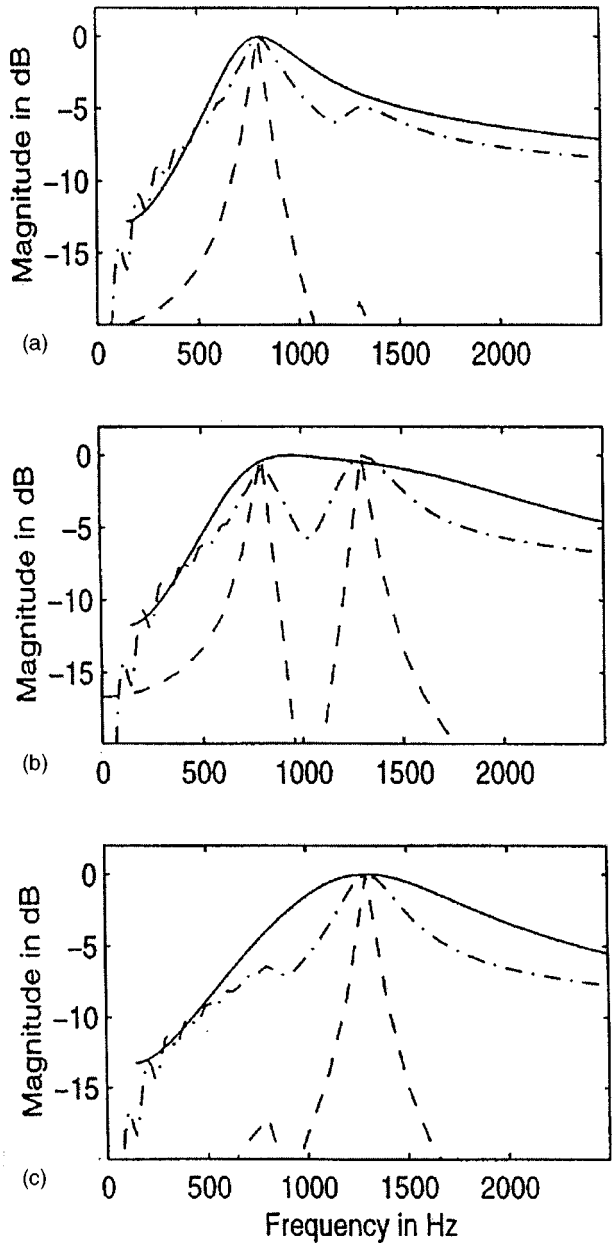


FIG. 2. Spectra derived from the proposed PLP model for a two-formant vowel # 1 (2.5-bark formant separation, pulse-train excited). Relative level ($A2/A1$) ranges from (a) -20 dB, (b) to 0 dB, (c) to $+20$ dB. Acoustic spectra (dashed line), auditory spectra (dashed-dotted line), and PLP model spectra (solid line) are plotted for each $A2/A1$ ratio.

formants reveals that (1) the single peak in the model spectrum becomes broader as separation increases, and (2) two peaks are detected when separation exceeds the critical distance. The two peaks become sharper as the formant distance increases.

These simulation results applied to a set of two-formant vowel-like signals indicate that the modified PLP model can account for the COG effect reasonably well.

4. Procedures

The listeners were asked to indicate whether the spectral pitch of the variable single-formant vowel was higher or lower than that of the two-formant reference vowel using a ''double-staircase'' adaptive tracking procedure (Jesteadt, 1980). The two-formant reference vowel was always pre-

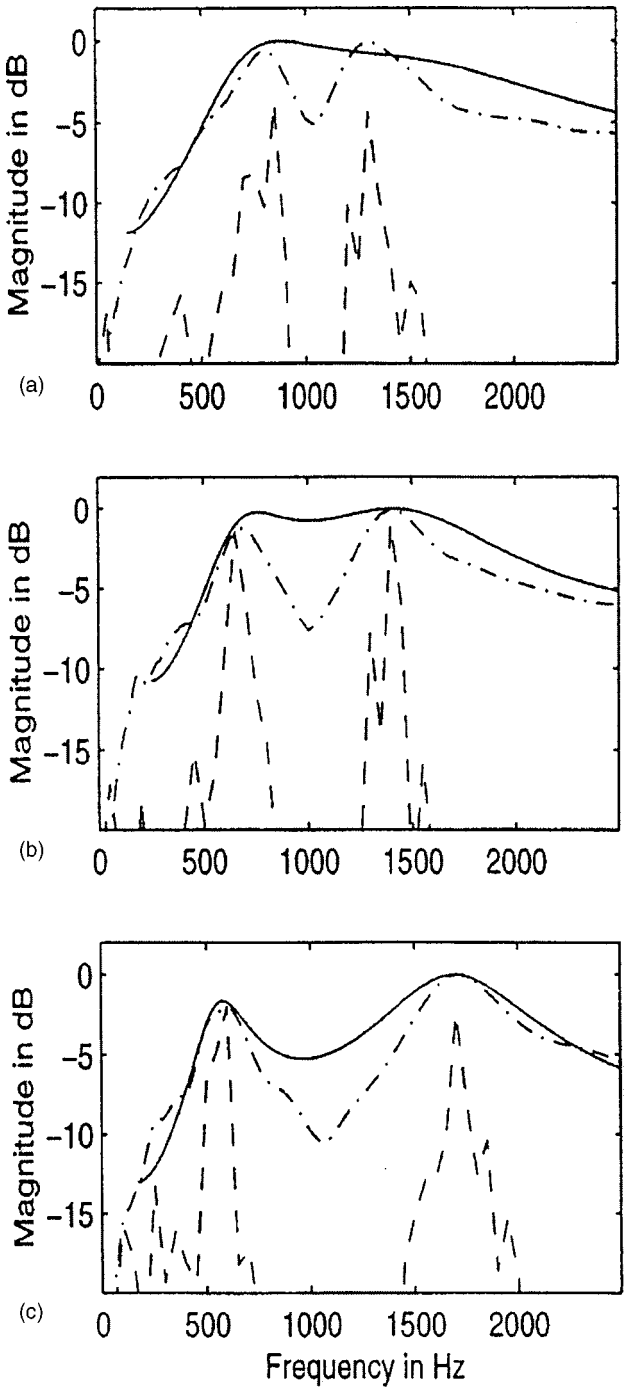


FIG. 3. Spectra derived from the proposed PLP model for two-formant vowels for three different $F2-F1$ separations and at a constant level $A2/A1=0$ dB. (a) vowel # 2 (2.5 bark); (b) vowel # 6 (3.6 bark); (c) vowel # 8 (5.3 bark). Acoustic spectra (dashed line), auditory spectra (dashed-dotted line), and model spectra (solid line) are plotted for each vowel.

sented first and the variable single-formant target vowel second, after a 600-ms silent interval. The listeners were instructed to push the right button of the mouse whenever they judged the variable signal to be higher in pitch than the reference signal. When they judged it to be lower in pitch than the reference, they pushed the left button. The double-staircase procedure maintains two adaptive tracking rules simultaneously. The ''two-up, one-down'' rule (Levitt, 1971) was applied to both the ascending and descending tracking sequences. Thus, the F_v estimate for the descending sequence was at a level for which the variable signal was judged higher than the reference signal on 71% of the trials. For the ascending sequence, F_v was judged lower on 29% of the trials.

In order to obtain the value of F_v at a given $A2/A1$ ratio, each of these estimates was based on eight consecutive 70-trial blocks of the double-staircase procedure. For each block, the F_v estimates for the descending and ascending sequences were first obtained by averaging the reversal points within each sequence. Then, the point-of-subjective-equality, PSE, was estimated by simply averaging the F_v estimates from the two sequences.

The use of the double-staircase adaptive procedure in lieu of the traditional matching tasks, as in early experiments on COG, was dictated by preliminary results from Lester (1996), who tested the COG hypothesis in a direct matching experiment. The listeners heard one of the two signals, alternatively. They were asked to adjust the second variable vowel to match the first vowel for quality. The results revealed considerable variability in matching data. The matches were tightly grouped around the lower formant frequency when $A2/A1=-20$ dB, but more scattered when $A2/A1=+20$ dB. Lester noted that direct matching to the ''perceptual formant'' was difficult even for well-trained subjects of the study. In related work (Feth *et al.*, 1996), the double-staircase adaptive procedure was used. The results showed that the dispersion of matching values at the higher frequency formant disappeared. This indicates that improvement in performance is related to eliminating inherently subjective judgments and including an objective criterion to measure the correctness of responses. Thus, the decision rules no longer continually select signals near the PSE but focus instead on points above and below it. In this case, the selection of signals by the adaptive procedure is controlled only by the observer's use of the two subjective response categories and not by how the responses relate to the objective properties of the stimuli. Consequently, the difficulty of the subjective tasks is greatly reduced.

III. RESULTS AND DISCUSSION

A. Overall results

The overall results of the experiment are shown in Figs. 4 and 5. All three listeners showed similar F_v matching values for all conditions for vowels # 1 through # 5 (2.5- and 3.0-bark separations), and # 8 (5.3 bark). The data for these conditions are therefore collapsed across the listeners. However, the results showed individual differences for vowels # 6 and # 7 (3.6 and 4.0 bark, respectively), which have their frequency separations near the 3.5-bark critical distance. These data were plotted for each subject individually in Figs. 5(b) and (c). Overall, the results indicate that frequency separation between the two formant peaks of the reference signals along with their relative levels have an effect on F_v matching values. When the frequency separation between the two formants does not exceed 3.5 bark, listeners show similar performance. The same is true for the supracritical separation of 5.3 bark. A substantial difference is observed in

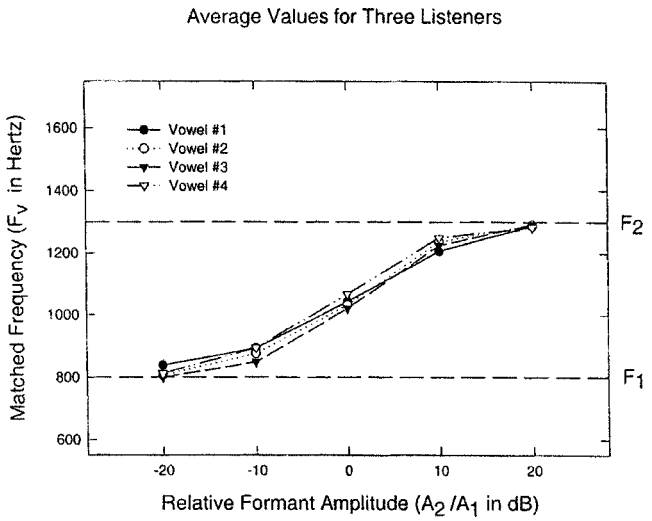


FIG. 4. Mean F_v matching values for vowels # 1 through # 4 (2.5-bark separation, variable source, variable bandwidth) with changing amplitude ratio ($A2/A1$). Dashed lines mark the locations of $F1$ and $F2$.

listeners' performance when the frequency separation approaches the "critical region" near 3.5 bark. For both 3.6- and 4.0-bark separations, variable patterns of responses were obtained.

The data were subjected to a two-way analysis of variance (ANOVA) with subject and relative level $A2/A1$ as factors. The relative level was significant for all vowel conditions ($p<0.001$). With $df=(4, 105)$, the largest F-ratio of 11 952.49 was obtained for vowel # 8 (5.3-bark condition) and the smallest F ratio of $F=492.68$ for vowel # 1 (pulse-train, 2.5-bark condition). The results of this ANOVA confirmed that, for each vowel condition, the level $A2/A1$ had a significant effect on F_v. Listener differences were not significant for vowels # 1 through # 5, and # 8, but they were significant for vowels # 6 and # 7 [$F(2,105)=31.78$, $p<0.001$ and $F(2,105)=22.51$, $p<0.001$, respectively]. This confirms the observed variability in listeners' performance for both 3.6- and 4.0-bark separations.

B. The COG effect

When the distance between the two formants of a reference vowel is smaller than the critical distance (3.5 bark), F_v values are similar to those reported by Chistovich and Lublinskaja (1979). For example, in Fig. 4, the data show a continuous relationship between F_v and $A2/A1$ for vowels # 1 through # 4 ($F2-F1=2.5$ bark). Consequently, all match-

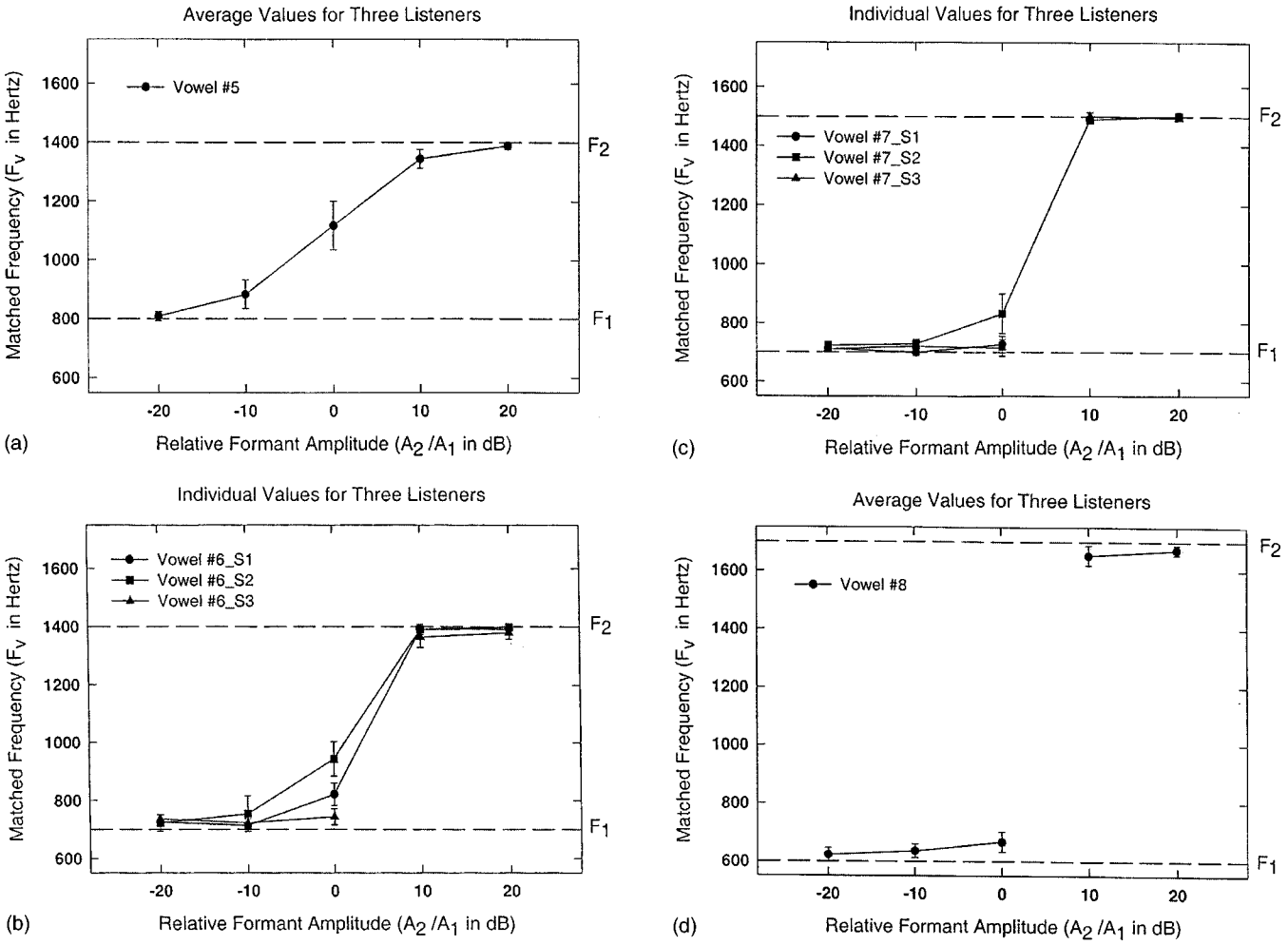


FIG. 5. F_v matching values for vowels # 5 through # 8 with changing $A2/A1$ ratio. (a) mean F_v values for three listeners for vowel # 5 (3.0 bark); (b) individual F_v values for each listener for vowel # 6 (3.6 bark); (c) individual F_v values for each listener for vowel # 7 (4.0 bark); (d) mean F_v values for three listeners for vowel # 8 (5.3 bark). Error bars at each data point represent 2 standard deviations. Horizontal dashed lines mark the locations of $F1$ and $F2$.

ing F_v points fall between $F1$ and $F2$. However, the data show a different pattern of variance with relative level. That is, the variance in the middle area ($A2/A1 = 0$ dB) is much greater than toward the ends ($A2/A1 = \pm 20$ dB), regardless of the excitation source or formant bandwidth. Standard deviation for the −20-dB condition ranges from 5.01–19.53 and from 13.80–24.46 for the 20-dB condition, whereas the range for 0 dB is from 48.03–76.37. This might indicate that listeners' uncertainty was greater when the auditory system integrated the spectra of two equal-level formants than when one of the two formants dominated. Another observation is that the range of $A2/A1$ over which F_v changed as a function of $A2/A1$ is about 40 dB (−20 to +20 dB). Figure 5(a) shows a similar result for 3.0-bark separation. These results confirm the existence of spectral COG effect for two-formant vowels when both formants are spaced closely enough to display the predicted variation in the matching frequency of the single-formant vowel.

C. The critical distance

Formant separations in the reference signals were set to 2.5, 3.0, 3.6, 4.0, and 5.3 bark to verify the existence of a critical distance. For all three listeners, a continuous relationship between the F_v values and $A2/A1$ was observed for frequency separations of up to 3 bark. However, when the separation between the formants of the reference vowel increased, this continuous relationship gradually disappeared. It is interesting to note that each listener showed a different breakdown point. For 3.6-bark separation [Fig. 5(b)], listener # 3 first shows a discontinuous relationship between F_v and $A2/A1$, while the other two maintain the continuous relationship. For 4.0-bark separation [Fig. 5(c)], listener # 1 demonstrates the breakdown, and only results from listener # 2 remain continuous. Finally, for 5.3-bark separation [Fig. 5(d)], the continuity disappeared in the performance of all three listeners. This indicates that the three subjects had different estimates of critical distance. This distance may be somewhere less than 3.6 bark for listener # 3, between 3.6 and 4.0 bark for listener # 1, and somewhat larger than 4.0 bark for listener # 2. A comparable pattern of responses was obtained by Chistovich and Lublinskaja (1979) for the supracritical separation. Performance of both their listeners differed considerably in that one listener showed a clear discontinuous relationship between F_v and $A2/A1$ and the continuity did not disappear in the performance of the second listener despite the large variation of the values of F_v. For the latter listener, no clear breakdown point was obtained.

Because the experimental procedure was already very time consuming, we did not explore additional values of $A2/A1$ to determine where exactly the F_v values shifted abruptly from $F1$ to $F2$. We can only determine that all fell within $A2/A1$ values ranging from 0 to +10 dB. This differs from results of Chistovich and Lublinskaja, where the shift from $F1$ to $F2$ occurred when $A2/A1$ was about −10 dB. This discrepancy in results for one experimental condition is most likely due to differences in the procedures used. Increased variability in the data collected in a direct matching task in Chistovich and Lublinskaja's study may have contributed to this result as well.

The data in Fig. 5 show that, for each listener, the variance of F_v is smaller within the supracritical separation than within the subcritical separation. This is true especially for the two formants with equal levels. This is similar to Chistovich and Lublinskaja's results, in which most values for the supracritical distance showed less variation than for subcritical separation. One explanation is that when the two formants of the reference signal are resolved, listeners simply pick either $F1$ or $F2$ for F_v match. However, the degree of uncertainty in making judgments may be considerably reduced in the double-staircase procedure than in the traditional adjustment task.

D. Formant bandwidths

For the subcritical distance of 2.5 bark, four reference vowels were generated with variable bandwidths. The results for the noise-excited vowels # 2, # 3, and # 4 (see Fig. 4) show that F_v values for vowel # 3 (the narrowest formant bandwidth) approximated closely the frequency values of either $F1$ or $F2$, whereas those for vowel # 4 (the widest bandwidth) were closer to the COG point. The only exception occurred when $A2/A1 = 10$ dB. To investigate the effect of formant bandwidth, a mixed-design ANOVA was performed on F_v data for vowel # 2 (BW=80 Hz), vowel # 3 (BW=45 Hz), and vowel # 4 (BW=150 Hz) with subject and formant level $A2/A1$ as two between-subject factors, and formant bandwidth as the within-subject factor. Formant bandwidth was significant [$F(2,210) = 17.172$, $p < 0.001$] and so were the interactions between bandwidth and $A2/A1$ [$F(8,210) = 3.621$, $p < 0.001$], and bandwidth and subject [$F(4,210) = 4.074$, $p < 0.004$]. The effect of subject was not significant [$F(2,105) = 0.323$, $p < 0.725$] and the interaction bandwidth × subject × $A2/A1$ was not significant [$F(16,210) = 1.651$, $p < 0.059$].

These results are in accord with Beddor and Hawkins' (1990) findings for their three bandwidth conditions: narrow (BW=45 Hz), medium (BW=75 Hz), and wide (BW=150 Hz), thus supporting the hypothesis that the relative contribution of formant frequency and amplitude might depend on the spectral characteristics of signals. Beddor and Hawkins propose that, in the matching task, formant frequency may be more important than spectral shape for signals with well-defined spectral peaks. Formant amplitudes and spectral shape might be more important for signals with poorly defined spectral peaks. In the former case, the F_v might lie closer to the formant frequency values. In our study, the vowel with narrower-than-normal formant bandwidth (45 Hz) had a well-defined spectral peak, while the vowel with wider-than-normal bandwidth (150 Hz) had a poorly defined spectral peak. Our results agree with Beddor and Hawkins' hypothesis except when $A2/A1 = 10$ dB.

E. Excitation source

Figure 4 shows the matching F_v values for the pulse-train-excited vowel # 1 and the noise-excited vowel # 2 with the same frequency separation (2.5 bark) and formant band-

width (80 Hz). Generally, listeners showed similar response patterns for the two reference signals. F_v values for vowel # 1 were lower than those for vowel # 2 when $A2/A1$ ranged from -20 to 0 dB, and higher when $A2/A1$ was 10 or 20 dB. The variances of the matching values were also comparable at each $A2/A1$ level for the pulse-train and noise-excited signals. For the two vowels, a mixed-design ANOVA was performed on the F_v data with subject and relative level $A2/A1$ as two between-subject factors, and excitation source type as a within-subject factor. Source type was not significant [$F(2,105)=1.281$, $p=0.260$] and subject effect was not significant [$F(2,105)=4.554$, $p=0.013$]; however, the interaction between source type and relative level $A2/A1$ was significant [$F(8,105)=5.352$, $p=0.001$]. This justifies the use of Gaussian noise excitation in generating all reference vowels to exclude potential effects of a harmonic complex tone source. The interaction between source type and subject was not significant [$F(4,105)=1.731$, $p=0.182$].

Overall, these data suggest that listeners are able to perform a spectral integration which depends on frequency separation between two formants. The results confirm the existence of a critical frequency region, in which the listeners' behavior changes rather drastically. This region, identified as ''critical distance'' of about 3.5 bark in early matching experiments by Chistovich and Lublinskaja (1979) and verified with various degrees of success by subsequent research, may indicate a limit to the range of spectral integration that the auditory system can perform. Thus, the observed COG effect may reflect a more broadly defined auditory behavior which is not peculiar to the perception of speech sounds. The second experiment was conducted to verify this possibility. The striking similarity of the two-tone resolution results of Feth and O'Malley (1977) to the critical distance observed in vowel matching tasks substantiates the claim that both the critical region and the complex-tone discriminability estimates may be related to a common mechanism, i.e., to an auditory spectral resolving power.

IV. EXPERIMENT 2: TWO-TONE SPECTRAL RESOLUTION BANDWIDTH

A. Methods

The same three trained listeners from experiment 1 participated. As in the synthetic vowel task, all listeners were given an extensive practice on complex signals before participating in the second experiment. Data were collected after there were no further improvements in their performance.

The stimuli used in experiment 2 were two-tone complex signals, which were geometrically centered around a frequency, f_c(1000 Hz) so that $f_c=(f_2f_1)^{1/2}$; $\Delta f=f_2-f_1$. In the target signal, the intensity difference ($\Delta I=L_2-L_1$) between the two components was $+1$ dB, while in the reference standard signal ΔI was -1 dB. The level of a given component at f_i is designated by L_i. All signals were generated by the laboratory PC using the TDT system II with D/A board. Each signal had duration of 300 ms and was shaped by a 5-ms cosine rise–fall window. The silent interval between signals was 300 ms. The signals were played at a 20-kHz sampling rate. The smoothing filter was the same analog low-pass filter with a cutoff frequency at 8 kHz, as used in experiment 1. Sound level was controlled by a programmable attenuator. All signals were presented monaurally through the HD-414-SL earphones at 50 dB SL.

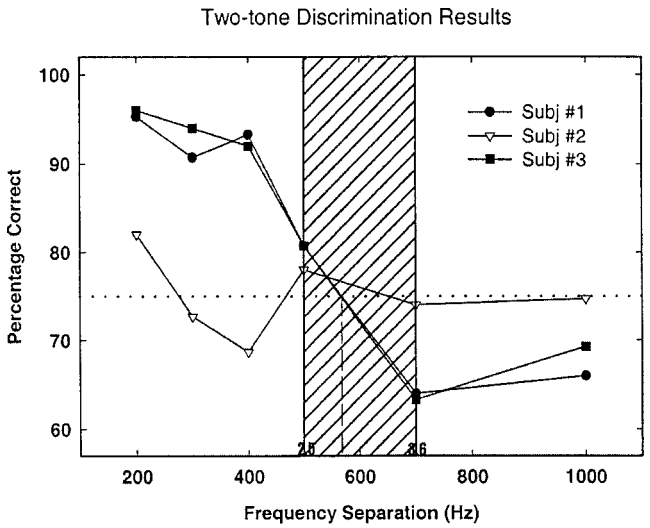


FIG. 6. Percentage of correct discrimination in a 2Q-2AFC task as a function of frequency separation, Δf, of the two components in each complex-tone pair. Each data point is the average of 150 trials. Individual psychometric functions are shown for each of the three listeners. The horizontal dotted line indicates the jnd at 75%. The shaded region delimits the frequency separation from 2.5 to 3.6 bark.

The two-cue, two-alternative, four-interval forced-choice paradigm was used. The target signal was presented with equal probability in either the second or the third interval. The reference signal was presented in the remaining three intervals. For the target signal, the higher-frequency component had the higher intensity, i.e., $\Delta I=+1$ dB. As in experiment 1, the listeners were asked to indicate whether the pitch of the variable signal was higher or lower than that of the reference signal.

At first, a two-up, one-down adaptive procedure was used with the initial value of the frequency separation, Δf, set at 100 Hz around the center frequency of 1000 Hz. However, the spectral pitch of both target and standard signals changed whenever Δf was changed from one value to another, and all three subjects reported that the task was too difficult to perform in this adaptive procedure. Therefore, a fixed standard procedure as in the Feth and O'Malley study (1977), was used instead. The Δf was fixed for a block of 50 trials and the percentage of correct discrimination, $P(C)$, was recorded. Each data point at a given Δf value was obtained from 150 trials. The psychometric function for each subject was generated by plotting $P(C)$ for Δf values of 200, 300, 400, 500, 700, and 1000 Hz.

B. Results and discussion

Figure 6 displays psychometric functions for each of the three participants. Percentage of correct responses, $P(C)$, is plotted as a function of the frequency difference (Δf) between the two tones. Listeners # 1 (●) and # 3 (■) show performance similar to that in Feth and O'Malley (1977), while the performance of listener # 2 (△) is clearly different.

For listeners # 1 and # 3, the $P(C)$ values are high for moderate Δf values near 200 Hz. Discrimination performance drops when Δf exceeds 400 Hz. The Δf at which the listener's $P(C)$ falls to 75% is used to estimate the spectral resolution bandwidth. For these two listeners, the estimate is 580 Hz.

Recall the results from experiment 1. We estimated that the critical distance for listener # 1 was somewhere between 3.6 and 4.0 bark, while it was less than 3.6 bark for listener # 3. The same tendency in performance is observed in experiment 2. The shaded area in Fig. 6 extends from a frequency separation equal to 2.5 bark to one equal to 3.6 bark. The ability of these two listeners to distinguish between the two-tone complexes drops below 75% within that same critical region. In the vowel task, when the formant separation was greater than the critical distance, we concluded that the auditory system failed to integrate across the two formants (i.e., the two formants were resolved). In the two-tone discrimination task, we assume that the discriminability of complementary two-tone pairs dropped below 75% when the two components were resolved by the auditory system. Thus, we conclude that both the critical distance between formants and the complex-tone discriminability limit are dependent upon the same auditory spectral resolving power.

The performance of listener # 2 in the discrimination task is very different from that of the other two listeners. The $P(C)$ values fluctuate across the whole Δf range as seen in Fig. 6. Even for moderate Δf values, $P(C)$ does not approach 100%. Furthermore, discriminability does not drop with larger Δf values. We have no explanation for why this listener performed so differently. Although this listener was given much more practice than the other two, she was unable to do the task and her performance is clearly different. Thus, we cannot determine the spectral resolution bandwidth for listener # 2. We infer from experiment 1 that her estimated critical distance is greater than 4.0 bark. Based on this value alone, we might conclude that her spectral resolution bandwidth is larger than 4.0 bark. This would lead us to predict that her "-75%" estimate is greater than 800 Hz in the two-tone task. However, her performance in experiment 2 was never regular enough to give us a solid basis to reach this conclusion.

V. GENERAL DISCUSSION AND CONCLUSIONS

As outlined in the Introduction, the phenomenon of spectral integration has been studied from two perspectives, i.e., as a perceptual averaging of a formant cluster in approximating phonetic quality of vowels ($F2'$ or center of gravity of the cluster), and as a frequency-displacement effect resulting from the interaction of frequency and relative amplitude ratio of two closely spaced formants (the COG effect). A common underlying approach for the two lines of research was to verify the early observations that spacing between two formants plays an important role in making perceptual decisions about vowel quality.

Although linking the 3.5-bark limit of spectral integration with $F2'$ brought consistent experimental results to confirm the importance of COG in estimating vowel quality in two-formant models, formalizing the COG effect encountered difficulties. The question of adequate methodology to study the COG effect arose as a consequence of attempts to formally account for its manifestation. Assmann's study (1991) showed no support for a COG effect using the first two formants of multiformant back vowels. His results from the vowel identification paradigm were also inconclusive. The use of matching tasks advocated by the Leningrad group has also been questioned as the only, and perhaps not very reliable procedure, to measure the COG effect. Considerable variability in the data from Chistovich and Lublinskaja (1979) and Lester (1996) provide additional evidence that a refinement in the experimental procedure to study the COG effect was necessary. It is worthwhile to note that both the "double-staircase" adaptive procedure used in Feth *et al.* (1996) and the two-cue, two-alternative forced-choice paradigm, as in Feth and O'Malley (1977), reduced the variability in the data.

A more reliable methodology gave rise to the question of the role of the COG effect and the formant amplitude ratio in estimating vowel quality. That is, for two closely spaced formants, listeners performance is predictable. When the frequency separation between the formants exceeds the 3.5-bark distance, relative levels of the formants do not contribute to otherwise predictable listening behavior: decisions about vowel identity are based entirely on the frequency of the formants. The experimental data on Russian vowels show that, for a subcritical separation of 3.5 bark, a change in $A2/A1$ ratio in two-formant vowels is perceptually equivalent to the frequency change which determines vowel quality.

In this study, we examined whether the COG effect plays role other than approximating phonetic vowel quality. Consequently, our listeners responded to the changes in vowel pitch, not vowel quality. In this respect, the task in experiment 1 was similar to that in experiment 2, in which the same listeners responded to differences in pitch in two-tone signals. In testing the COG effect in experiment 1, we sought to verify the role of the critical separation of 3.5 bark in spectral integration, using both a more reliable experimental procedure and model predictions incorporating the interaction between formant frequencies and their relative amplitudes. In this endeavor, we re-examined some aspects of Chistovich and Lublinskaja's results (1979) and introduced further modifications to the methods such as source type (Gaussian noise) and bandwidth manipulations, as in Beddor and Hawkins' study (1990). We used well-trained listeners as subjects of our study because we were interested in the optimal performance of the human auditory system in testing the COG effect.

The results of experiment 1 confirm that the COG effect occurs within the limit of spectral integration of 3.5 bark. Listeners' responses to changes in vowel pitch indicate a match in frequency according to the COG mechanism: when $F2$ is much weaker than $F1$ ($A2/A1 = -20$ dB), the matches fall in the vicinity of the stronger $F1$ and, conversely, listeners match the frequency closer to $F2$ when $F1$ is weaker ($A2/A1 = +20$ dB). However, when both formants are of equal strength ($A2/A1 = 0$ dB), the matches fall somewhere between the two formants, and an increased variability is observed in the data. Crucially, this tendency is maintained

for all conditions within the subcritical formant separation for vowels # 2 through # 5.

In designing experiment 1, we hypothesized that spectral shape (poorly defined spectral prominence versus well-defined spectral prominence) affects the matching values of F_v. Accordingly, three two-formant vowels with variable bandwidth were generated with subcritical separation of 2.5 bark (vowels # 2 through # 4). The vowel with narrower-than-normal formant bandwidth of 45 Hz had a well-defined spectral peak (vowel # 3), and the vowel with wider-than-normal bandwidth (150 Hz) had a poorly defined spectral peak (vowel # 4). Model predictions were verified in the experimental data. Listeners approximated the frequency of either $F1$ or $F2$ for the narrow-bandwidth vowel # 3, thus selecting one of the ''stronger'' peaks. For vowel # 4, the matches fell closer to a frequency between the formants, which is in accord with the COG effect. Our results and model predictions support Beddor and Hawkins' hypothesis with one exception, i.e., when $A2/A1 = 10$ dB.

Experiment 2 verified earlier findings by Feth and O'Malley (1977) that spectral integration in two-tone complex signals occurs within the limit of the 3.5-bark resolution bandwidth. Given that the responses in experiment 2 came from the same participants as in experiment 1, we conclude that the complex-tone discriminability and the spectral integration limits reflect the same auditory spectral resolving power. This further suggests that the auditory processing of complex signals at the intermediate stage, i.e., before higher-level decision processes apply, is the same for speech and nonspeech signals.

As discussed in the Introduction, several models were previously proposed for predicting the performance of listeners asked to match an adjustable one-formant to a fixed two-formant signal. Early work suggested that a simple calculation of the spectral centroid of the two-formant signal would serve this purpose (Chistovich *et al.*, 1979). We verified the predictions of the spectral centroid model and also used the IWAIF model for a comparison with the present results. Figure 7 displays a summary of predictions of our modified PLP model, the IWAIF model, and the spectral centroid model, along with the group data for vowel # 2 (2.5-bark separation, noise-excited) averaged across three listeners. To implement the IWAIF and spectral centroid models, we assumed that signal energy outside the critical distance was not available for the calculation.

Clearly, the spectral centroid model performs most poorly in this comparison. The predicted relationship between F_v and relative formant levels does not fit the experimental data. The endpoints are overestimated for $F1$ peak and underestimated for $F2$ peak, and only the value at the center of the $A2/A1$ range falls within the error bars. For both the PLP model and the IWAIF model, there is much better agreement with the data. The PLP model predicts well the performance at the endpoints (i.e., both extreme $A2/A1$ values) but underestimates listeners' matches by more than 2 standard deviations at $A2/A1 = 0$ dB. The IWAIF model, on the other hand, predicts listeners' matches that fall just above those obtained in the experiment. An exception to this tendency is the $F2$ region where the predicted values are lower

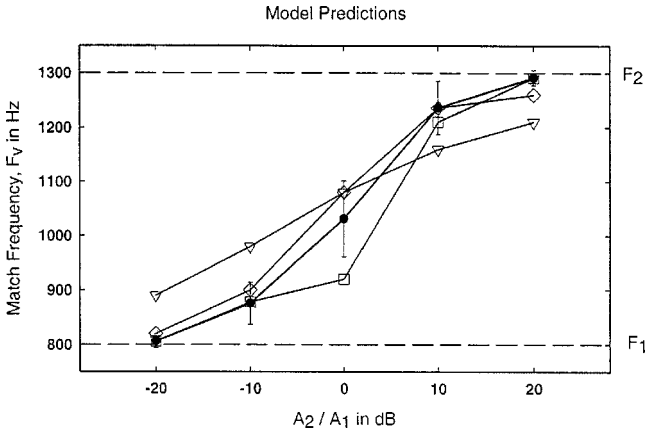


FIG. 7. Model predictions and experimental data for vowel # 2 (2.5 bark): (□) predictions by the proposed PLP model; (◇) by the IWAIF model, (▽) by the spectral centroid model with linear amplitude weighting, (●) the experimental data for vowel # 2; each error bar represents 2 standard deviations. The 3.5-bark rectangular window is used to select the spectra for the IWAIF and the centroid models. Horizontal dashed lines mark the locations of $F1$ and $F2$.

than those in the experimental data. At this stage, it is still premature to determine which of the two models, the PLP model or the IWAIF model, is better suited to predicting COG effects with greater accuracy. Further refining of the IWAIF model is probably necessary, since we had to assume an arbitrary limitation on signal bandwidth to produce a reasonable prediction of performance.

In conclusion, the two experiments reported here have shown that the 3.5-bark limit on spectral integration appears to reflect a property of auditory processing rather than some special processing unique to speech sounds. This study, and that of Chistovich and Lublinskaja (1979), used a small number of very well-trained listeners, but some evidence for individual differences in the size of the critical separation is apparent in both. The effects of formant bandwidth and excitation source were predictable. Finally, a preliminary comparison of model predictions of COG performance did not favor one model strongly; however, the simple spectral centroid model was clearly the poorest at predicting listener matches.

The signals used in the current study were spectrally static. That is, the parameters of both the adjustable and the reference signals remained constant for the entire duration of the sound. Real speech sounds are typically dynamic; that is their formant frequencies and amplitudes change, sometimes rapidly, over time. Lublinskaja (1996) has recently reported that the COG effect can be demonstrated dynamically by amplitude modulation of formant amplitudes. Further work on modeling of the COG effects therefore should take into account these dynamic effects.

ACKNOWLEDGMENTS

This work was supported by a grant from AFOSR. The authors wish to thank Douglas O'Shaughnessy and several anonymous reviewers of the manuscript for many helpful comments.

Anantharaman, J. N., Krishnamurthy, A. K., and Feth, L. L. (**1993**). ''Intensity weighted average of instantaneous frequency as a model for frequency discrimination,'' J. Acoust. Soc. Am. **94**, 723–729.
Anantharaman, J. N. (**1998**). ''A perceptual auditory spectral centroid model,'' Ph.D. thesis, The Ohio State University, Columbus, OH.
Assmann, P. F. (**1991**). ''The perception of back vowels: Centre of gravity hypothesis,'' Q. J. Exp. Psychol. **43A**, 423–448.
Beddor, P. S., and Hawkins, S. (**1990**). ''The influence of spectral prominence on perceived vowel quality,'' J. Acoust. Soc. Am. **87**, 2684–2704.
Bedrov, Y. A., Chistovich, L. A., and Sheikin, R. L. (**1978**). ''Frequency location of the 'center of gravity' of the formants as a useful parameter in vowel perception,'' Akust. Zh. **24**, 480–486 (Sov. Phys. Acoust. **24**, 275–282).
Bladon, A. (**1983**). ''Two-formant models of vowel perception: Shortcomings and enhancements,'' Speech Commun. **2**, 305–313.
Bladon, A., and Fant, G. (**1978**). ''A two-formant model and the cardinal vowels,'' Royal Inst. Tech., Stockholm, STL-QPRS **1**, 1–8.
Carlson, R., Granström, B., and Fant, G. (**1970**). ''Some studies concerning perception of isolated vowels,'' Royal Inst. Tech., Stockholm, STL-QPRS **2–3**, 19–35.
Carlson, R., Fant, G., and Granström, B. (**1975**). ''Two-formant models, pitch and vowel perception,'' in *Auditory Analysis and Perception of Speech*, edited by C. G. M. Fant and M. A. Tatham (Academic, New York and London), pp. 58–82.
Chistovich, L. A. (**1985**). ''Central auditory processing of peripheral vowel spectra,'' J. Acoust. Soc. Am. **77**, 789–804.
Chistovich, L. A., and Lublinskaja, V. V. (**1979**). ''The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli,'' Hear. Res. **1**, 185–195.
Chistovich, L. A., Sheikin, R. L., and Lublinskaja, V. V. (**1979**). '''Centres of gravity' and spectral peaks as the determinants of vowel quality,'' in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 55–82.
Delattre, P., Liberman, A., Cooper, F., and Gerstman, L. (**1952**). ''An experimental study of the acoustic determinants of vowel color,'' Word **8**, 195–210.
Escudier, P., Schwartz, J.-L., and Boulogne, M. (**1985**). ''Perception of stationary vowels: Internal representation of the formants in the auditory system and two-formant models,'' Franco–Swedish Seminar, SFA, Grenoble, 143–174.
Fant, G. (**1959**). ''Acoustic analysis and synthesis of speech with applications to Swedish,'' Ericsson Tech. **1**, 3–108.
Feth, L. L. (**1974**). ''Frequency discrimination of complex periodic tones,'' Percept. Psychophys. **15**, 375–378.
Feth, L. L., Lester, J., Xu, Q., and Ross, J. (**1996**). ''Testing the 'center of gravity' effect for vowel-like complex sounds,'' J. Acoust. Soc. Am. **100**, 2626(A).
Feth, L. L., and O'Malley, H. (**1977**). ''Two-tone auditory spectral resolution,'' J. Acoust. Soc. Am. **62**, 940–947.
Helmholtz, H. L. F. (**1954**). *On the Perception of Tone*, 2nd English ed. (Dover, New York).
Hermansky, H. (**1990**). ''Perceptual linear predictive (PLP) analysis of speech,'' J. Acoust. Soc. Am. **87**, 1738–1752.
Jeffress, L. A. (**1968**). ''Beating sinusoids and pitch changes,'' J. Acoust. Soc. Am. **43**, 1964.
Jesteadt, W. (**1980**). ''An adaptive procedure for subjective judgments,'' Percept. Psychophys. **28**, 85–88.
Lester, J. (**1996**). ''Spectrographic analysis of sound vs. auditory perception,'' Senior Honors thesis, The Ohio State University, Columbus, OH.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lublinskaja, V. V. (**1996**). ''The 'center of gravity' effect in dynamics,'' in *Proceedings of the Workshop on the Auditory Basis of Speech Production*, edited by W. Ainsworth and S. Greenberg, pp. 102–105, ESCA.
Mantakas, M., Schwartz, J.-L., and Escudier, P. (**1988**). ''Vowel spectrum processing and the large-scale integration concept,'' 7th FASE Congress, Edinburgh.
Schwartz, J.-L., and Escudier, P. (**1989**). ''A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception,'' Speech Commun. **8**, 235–259.
Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (**1997**). ''The dispersion-focalization theory of vowel systems,'' J. Phonetics **25**, 255–286.
Voelcker, H. B. (**1966a**). ''Toward a unified theory of modulation I. Phase-envelope relationship,'' Proc. IEEE **54**, 340–353.
Voelcker, H. B. (**1966b**). ''Toward a unified theory of modulation II. Zero manipulation,'' Proc. IEEE **54**, 735–755.
Xu, Q. (**1997**). ''A signal processing model for spectral integration of vowel sounds,'' M.A. thesis, The Ohio State University, Columbus, OH.

Sequential streaming and effective level differences due to phase-spectrum manipulations

Thomas H. Stainsby,[a] Brian C. J. Moore, Peter J. Medland, and Brian R. Glasberg
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 24 September 2003; revised 17 December 2003; accepted 7 January 2004)

Roberts *et al.* [J. Acoust. Soc. Am. **112**, 2074–2085 (2002)] demonstrated that sequential stream segregation occurs with stimuli that differ only in phase spectrum. We investigated if this was partly due to differences in effective excitation level. Stimuli were harmonic complexes with a 100 Hz fundamental, 1250–2500 Hz passband, and cosine, alternating, or random component phase. In experiment 1, the complex tones were used as forward maskers of 20-ms probe tones at 1000, 1250, 1650, 2050, 2500, and 3000 Hz. While there was no significant difference in the masking produced by the cosine- and alternating-phase stimuli, the random-phase stimulus produced significantly greater masking, equivalent to a difference in overall effective excitation level of 12.6 dB. Experiments 2 and 3 used the asynchrony detection and subjective streaming tasks of Roberts *et al.* Successive stimuli had identical phase, but differed in level by 0, 1, 3, 5, 10, or 15 dB. Stream segregation increased once the level difference reached 5 dB. While some of the stream segregation observed by Roberts *et al.* may have been due to a difference in effective excitation level, this does not account for the stream segregation between cosine- and alternating-phase stimuli. [DOI: 10.1121/1.1650288]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk, 43.66.Nm [GK] Pages: 1665–1673

I. INTRODUCTION

Streaming is the process of assigning successive sound events to perceived sound sources or streams. When events are assigned to the same stream, this is called fusion or integration, while when they are assigned to different streams, this is called fission or stream segregation (van Noorden, 1975; Bregman, 1990). Hartmann and Johnson (1991) studied streaming by measuring the ability to recognize a melody in the presence of interleaved tones; the tones forming the melody differed from the interleaved tones along a variety of dimensions. They concluded that the dominant factor determining whether stream segregation occurred was the degree to which successive tones excited different auditory channels. This is called the ''peripheral channeling'' theory, and it provides the basis for the computational models of streaming described by Beauvois and Meddis (1996), and McCabe and Denham (1997). However, several recent studies have shown that stream segregation can occur for stimuli whose excitation patterns overlap considerably (Vliegen *et al.* 1999; Vliegen and Oxenham, 1999; Grimault *et al.* 2000; 2001; 2002); for a review, see Moore and Gockel (2002).

Roberts *et al.* (2002) demonstrated that stream segregation could occur when successive complex tones had identical power spectra and differed only in their component phase relationships, and hence their temporal structure. They interpreted their results as indicating that differences in temporal structure alone were sufficient to produce stream segregation. In this paper, we assess an alternative explanation of their results. To understand the background to our experiments, it is necessary to consider the stimuli of Roberts *et al.* (2002) in some detail. Their tones had a fundamental frequency (F0) of 100 Hz and had three different starting phase relationships: cosine (C)—all components starting with 90° phase; alternating (A)—successive components starting alternately with 0° or 90° phase; and random (R)—each component having a different, randomly specified, starting phase. Stimuli were filtered so as to contain only high, unresolved harmonics. ''C'' phase leads to a waveform with a high peak factor; this evokes a distinct low pitch and a buzzy timbre (Schouten, 1940). ''A'' phase leads to a waveform with two major peaks per period; this evokes a pitch that is one-octave higher than for cosine phase (Shackleton and Carlyon, 1994) and a timbre that is less buzzy. ''R'' phase leads to a waveform with a relatively low peak factor, with several waveform peaks per period; this evokes a less distinct pitch than for cosine phase and leads to a noise-like timbre (Bilsen, 1973). In two different tasks (described later), Roberts *et al.* (2002) found stream segregation when the phase of successive tones alternated between C and R or between C and A.

As noted earlier, Roberts *et al.* (2002) assumed that the stream segregation found for their stimuli was based on differences in temporal structure. However, even for stimuli with the same power spectra but different phase relationships, the excitation patterns may not be identical. The fast-acting compression that occurs on the basilar membrane (Sellick *et al.*, 1982; Robles *et al.*, 1986; Recio *et al.*, 1998) means that the average excitation evoked at a given place depends on the peak factor of the waveform at that place; high peak factors tend to lead to lower average excitation. Hence, complex tones that evoke waveforms with high peak factors (Schroeder positive phase, sine phase or cosine phase) may evoke less excitation than complex tones that evoke waveforms with low peak factors (Schroeder negative

[a]Electronic mail: ths22@cam.ac.uk

phase or random phase). This has been observed physiologically (Recio and Rhode, 2000; Rhode and Recio, 2001a; 2001b) and inferred from changes in the amount of forward masking produced by manipulating the phase of complex tone maskers (Carlyon and Datta, 1997; Gockel *et al.*, 2003; Moore *et al.*, 2004).

It is possible that some of the effects of component phase on streaming found by Roberts *et al.* (2002) were due to differences in excitation level between successive stimuli. It is not clear that such differences in excitation level correspond directly to differences in loudness. Indeed, Gockel *et al.* (2003) reported that complex tones with components added in C phase sounded *louder* than tones with the same power spectra but with components added in R phase, even though the latter produced more forward masking than the former. Nevertheless, differences in excitation level and/or loudness may have influenced the results of Roberts *et al.* (2002). Previous research has shown that subjective stream segregation can occur based on differences in level (van Noorden, 1975) and that differences in level can lead to improved performance in a task requiring recognition of interleaved melodies (Dowling, 1973), in which the subject tries to hear segregation. However, to our knowledge, it has not been demonstrated that differences in level can induce obligatory stream segregation, i.e., segregation that occurs even when the best performance in a task would be achieved by hearing successive tones as integrated.

The current study had two main goals: first to quantify the differences in the effective excitation level produced by the stimuli of Roberts *et al.* (2002); second to measure both subjective and obligatory stream segregation as a function of the difference in level between successive tones. In experiment 1, the C-, A-, and R-phase stimuli were used as forward maskers, to quantify the differences in effective excitation level produced by the phase manipulations. The data from experiment 1 were used to select the level differences used in experiments 2 and 3; in the latter, the effects of differences in level on streaming were explored for the same rhythm discrimination and subjective streaming tasks as used by Roberts *et al.* (2002). In these two experiments, the A and B tones in a given condition were identical in passband and component phase relationship, but differed in level. Experiments 2 and 3 were intended to allow us to assess whether differences in effective excitation level of the stimuli used by Roberts *et al.* were sufficient to account for the phase-based streaming effects observed by them.

II. EXPERIMENT 1: ESTIMATION OF EXCITATION LEVELS USING FORWARD MASKING

In experiment 1, differences in the effective excitation levels of the complex tones were estimated by determining the forward-masked thresholds of pure-tone signals at six different frequencies. The forward-masked threshold of a signal is usually assumed to be monotonically related to the amount of excitation produced by the masker at the signal frequency (Plomp, 1964a; 1964b; Houtgast, 1972; Moore and Glasberg, 1983b). The masked threshold as a function of the signal frequency provides a measure of the profile of excitation produced by the masker across frequency. Several different masker levels were used so that differences in forward masking across masker types could be expressed as differences in effective excitation level.

A. Subjects

The same six subjects were tested in all experiments described in this paper. All had normal hearing, with thresholds at the standard audiometric frequencies that were within 20 dB of the ANSI standard (ANSI, 1969). Three were authors B.G., P.M., and T.S., the first and third of whom had previously participated in psychoacoustic experiments. The other subjects were paid for their services. Subject ages were 20, 20, 21, 23, 31, and 56. Details of training are given in the following.

B. Stimuli and apparatus

The stimuli were a subset of those used by Roberts *et al.* (2002). The maskers were harmonic complex tones with F0 $=100$ Hz. These complexes were bandpass filtered between 1250 and 2500 Hz, so that all the components within the passband were unresolved (Plomp, 1964a; Moore and Ohgushi, 1993). Components within the passband all had equal amplitude, and those outside were attenuated with slopes of 80 dB/octave. Components whose level would have been more than 60 dB below the level of the components within the passband were omitted. All maskers had a total duration of 200 ms, including 10-ms onset and offset ramps. Maskers were presented at three different levels, 48, 58, and 68 dB SPL per component, which is equivalent to overall masker levels of 60, 70, and 80 dB SPL. Components were added in one of the three starting phase relationships described in Sec. I: cosine (C), alternating (A), and random (R). For the R stimuli, the component starting phases were fixed within a run, but component phases were chosen randomly at the start of each run, and differed across runs.

Previous research has shown that the forward masking produced by periodic complex tones depends little, if at all, on the point during the masker cycle at which the masker is turned off (Carlyon and Datta, 1997; Gockel *et al.*, 2003; Moore *et al.*, 2004). To check that this was the case for our stimuli, this ''offset point'' was systematically varied. For each masker phase, level and signal frequency, three different offset points were used; the masker offset ramp started 0, 3.33, or 6.67 ms relative to ''zero'' time in a given period of the masker. For the C and A maskers, this zero time corresponded to a peak in the masker wave form. Similar to the approach taken by Moore *et al.* (2004), we intended to treat the data obtained with the different offset points as repetitions if no significant effect of masker offset was found, and to use the mean of such data in further analyses.

The signals were pure tones with 10-ms raised-cosine onset and offset ramps and no steady-state portion, making a total duration of 20 ms. The signal frequencies were 1000, 1250, 1650, 2050, 2500, and 3000 Hz. The delay between the offset of the masker and the onset of the probe was 0 ms. There were in total 162 conditions (3 masker phases, 3 levels, 3 offset times, and 6 signal frequencies).

All stimuli were generated using a Tucker-Davis Technologies System II. The frequency-shaped complex stimuli were calculated on-line using a TDT AP-2 processor card. Stimuli were played out at a sampling rate of 50 kHz with 16 bit resolution using a TDT DD1 DAC. Stimuli were then passed through a TDT PA4 programmable attenuator, a TDT SM3 summer, a TDT HB6 headphone buffer, a Hatfield 2125 manual attenuator, and a single earpiece of a pair of Sennheiser HD580 headphones. All listening was performed in an Industrial Acoustics double-walled sound attenuating chamber.

C. Procedure

Forward-masked thresholds were obtained using a three-interval forced-choice adaptive tracking procedure. A trial consisted of three bursts of the masker, separated by 500 ms, one of which, selected at random, was followed by the signal. Each interval was marked by a numbered LED on a response box, and subjects responded by pressing the corresponding numbered button. The 79.4% point on the psychometric function was estimated using a three-down one-up rule (Levitt, 1971), whereby the signal level was decreased by one step after three consecutive correct responses, and increased by one step after each incorrect response. Each run was continued until twelve turnpoints had been obtained. The step size was 5 dB until four turnpoints occurred, and 2 dB thereafter. The threshold was taken as the arithmetic mean of the levels at the last eight turnpoints. The level of the signal at the start of a run was 45 dB SPL when the masker level was 60 dB SPL and 55 dB SPL when the masker level was 70 or 80 dB SPL. Absolute thresholds for the signals were measured using the same procedure.

The order of testing of the 162 conditions was chosen to reduce practice and order effects. The signal frequencies were presented in ascending order for one half of the subjects, and in descending order for the other half. The masker phase was changed every six runs, in the order C, A, R for two subjects, A, R, C for another two, and R, C, A for the final two. For all subjects, the masker level was changed every 18 runs in the order of 60, 70, and 80 dB SPL. The masker starting-phase offset was changed every 54 runs in the following order: 6.67, 0, and 3.33 ms for two subjects, 0, 3.33, and 6.67 ms for two more subjects, and 3.33, 6.67, and 0 ms for the remaining two. Subjects typically required around ten 1.5-h sessions to complete all the conditions.

Prior to the test runs, subjects were given a block of 12 practice runs. These included examples of all signal frequencies, masker levels, and phase relationships, in order to familiarize subjects with the range of stimuli. The last four runs in the practice block were all identical, to check whether the performance of the subject had stabilized. If the threshold estimates for these four runs covered a range of more than 10 dB, the subject was given extra practice.

D. Results

The absolute threshold for each probe frequency was subtracted from each masked threshold, so all results are expressed as amount of masking. The pattern of results was similar across subjects and so the results were averaged across subjects. A four-way within-subjects ANOVA was performed, the factors being probe frequency (1000, 1250, 1650, 2050, 2500, and 3000 Hz), masker phase (C, A, or R), masker level (60, 70, or 80 dB) and masker-phase offset (0, 3.33, or 6.67 ms). No significant effect of masker-phase offset was found, nor any interaction of phase offset with other factors. Hence, the arithmetic mean across offsets is used in all subsequent discussion and figures. The mean amounts of masking for each masker level are shown in the three panels of Fig. 1.

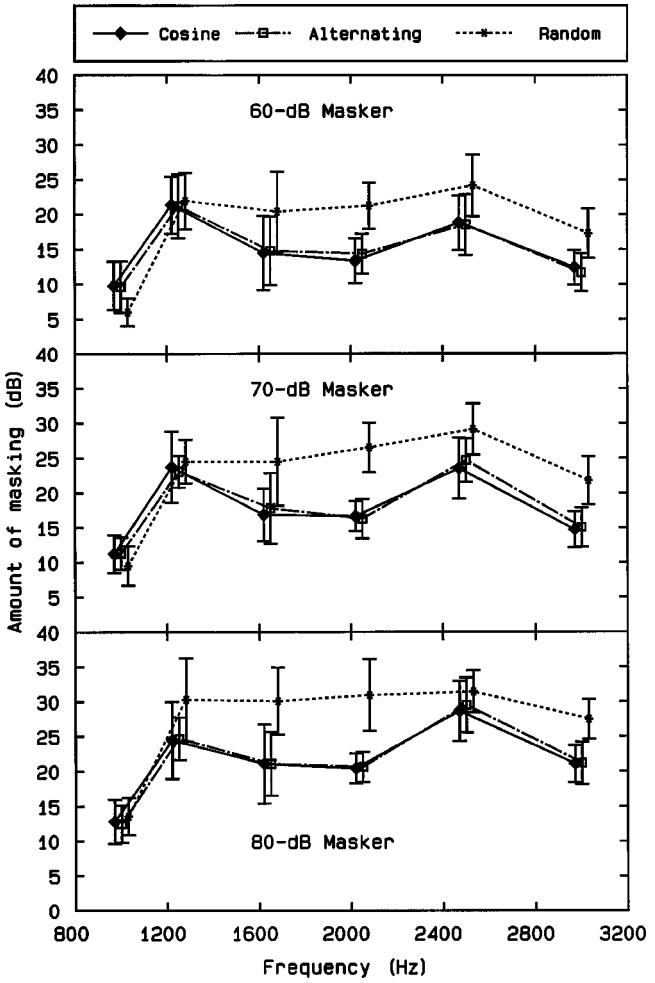


FIG. 1. Masking patterns (amount of masking as a function of signal frequency) for maskers with C, A, and R phases. The three panels show, from the top down, the mean thresholds obtained at the three different overall masker levels of 60, 70, and 80 dB SPL. The symbols for the different types of maskers are offset slightly along the *x* axis for clarity. Error bars show standard deviations across subjects.

There was a significant increase in the amount of masking with increasing masker level; $F(2,10)=220.1$, $p<0.001$. Mean amounts of masking were 16.2, 19.5, and 23.3 dB for the 60, 70, and 80 dB maskers, respectively. The main effect of phase was also highly significant; $F(2,10)=65.0$, $p<0.001$. Mean amounts of masking were 18.1, 18.2, and 22.6 dB for C, A, and R phase, respectively. The C- and A-phase stimuli produced masking patterns that were very similar to each other, and the difference in amount of masking was not significant according to the least-significant differences (LSD) test; $t(10)=0.16$. In contrast, the R-phase

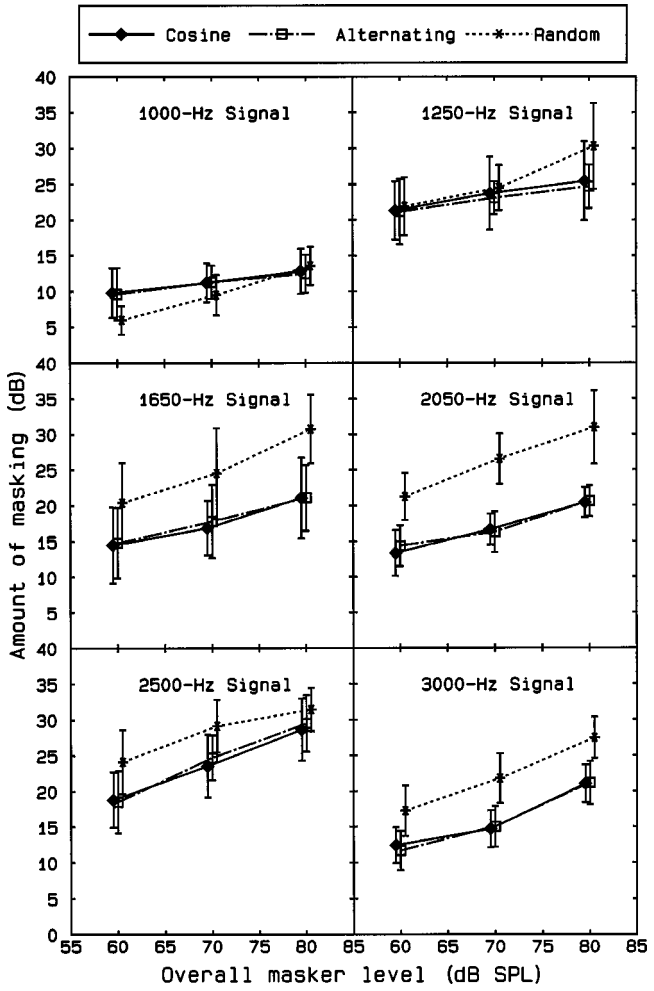


FIG. 2. Growth-of-masking functions for each signal frequency. Each panel shows the function obtained for each masker phase at a given signal frequency. Error bars show standard deviations across subjects.

masker produced more masking than the C- or A-phase maskers, and the difference was significant in both cases: R versus C, $t(10)=9.95$, $p<0.001$; R versus A, $t(10)=9.79$, $p<0.001$. The effect of frequency was highly significant; $F(5,25)=16.09$, $p<0.001$. There were significant interactions between phase and frequency ($F(10,50)=20.23$, $p<0.001$), masker level and frequency ($F(10,50)=6.13$, $p<0.001$) and between phase, masker level and frequency ($F(20,100)=2.43$, $p=0.002$).

Growth-of-masking functions, plots of amount of masking as a function of masker level, are shown in Fig. 2. The functions are roughly straight lines. The R-phase maskers produced more masking than the C- and A-phase maskers for all frequencies except 1000 Hz. Figure 3 shows the growth-of-masking functions averaged across the four frequencies within the passband of the masker. Linear regression fits were calculated for these growth-of-masking functions for the R- and C-phase maskers (the fitted lines were almost identical for the C- and A-phase maskers). The former had a slope of 0.44 and an intercept of −4.62, while the latter had a slope of 0.34 and an intercept of −3.08. From these equations, it can be calculated that a C-phase masker at 70 dB SPL would produce the same amount of masking as an R-phase masker at 57.4 dB—a difference of 12.6 dB. Thus,

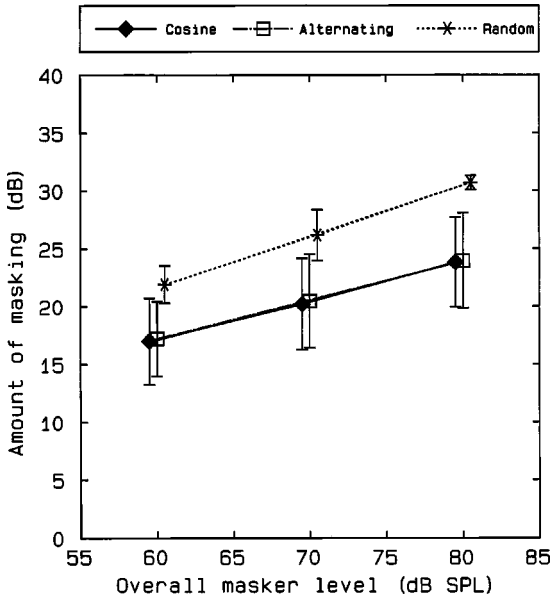


FIG. 3. Growth-of-masking functions for each masker phase, averaged across the four signal frequencies within the passband of the masker. Error bars show standard deviations across subjects.

the "effective excitation" of the C-phase masker is 12.6 dB less than that of the R-phase masker.

E. Discussion

Our results confirm the expectation that the maskers evoking waveforms with high peak factors on the basilar membrane (C and A phase) would produce less forward masking than the masker evoking waveforms with low peak factors (R phase). The shapes of the masking patterns resemble those presented by Gockel *et al.* (2003), in showing peaks for signal frequencies close to the edges of the masker passband. Also, the dip for frequencies well within the masker passband was smaller for R phase than for C or A phase; this is supported by the significant interaction of phase and frequency found in the ANOVA. This again resembles the findings of Gockel *et al.* for C and R phase (they did not use A phase). The peaks at the edges of the masker passband probably arise partly from the influence of lateral suppression. Suppression within a broadband sound occurs both from low frequencies toward higher ones and vice versa (Houtgast, 1974; Javel, 1981; Ruggero *et al.*, 1992). Suppression is strongest in a given frequency region when there is suppression from both lower and higher regions, i.e., when the frequency is well within the spectral range of the sound. Toward the spectral edge of the sound, there is effectively a release from suppression, which results in enhancement of the edge (Houtgast, 1972; 1974).

For the C- and A-phase stimuli, a second factor may have played a role in producing the peaks close to the edges of the masker passband. The data are consistent with the idea, discussed in Sec. I, that maskers evoking waveforms with high peak factors on the basilar membrane tend to produce less forward masking. At the places on the basilar membrane tuned to the edges of the masker passband, the waveforms evoked by the C- and A-phase maskers would have had lower peak factors than for places tuned well within the

masker passband. This would happen because, for a given place on the basilar membrane, masker components falling well outside the passband for that place are required to produce waveforms with high peak factors; see Alcántara *et al.* (2003).

As can be seen in Fig. 1, for the lowest signal frequency within the masker passband (1250 Hz), there was no phase effect for the two lower masker levels. However, for the highest signal frequency within the masker passband (2500 Hz), a phase effect occurred for all three masker levels. This may be related to the greater bandwidth of the auditory filter at the higher of these two frequencies, which would allow greater interaction of masker components.

The major conclusion to be drawn from this experiment is that there were clear differences in the forward masking produced by the R-phase masker on the one hand and the C- and A-phase maskers on the other hand. This implies that the R-phase masker differed from the other maskers in effective excitation level. This difference might have contributed to the streaming effects found by Roberts *et al.* (2002). However, the fact that no differences in forward masking were found for the C- and A-phase maskers implies that these two maskers evoked very similar excitation patterns. Therefore, the streaming found by Roberts *et al.* for the C and A stimuli cannot be explained in terms of differences in effective excitation level.

III. EXPERIMENT 2: MEASUREMENT OF LEVEL-BASED STREAMING USING RHYTHM DISCRIMINATION

Experiment 2 investigated the effects that differences in effective excitation level might have on the process of obligatory stream segregation (Bregman, 1990), using the rhythm discrimination paradigm of Roberts *et al.* (2002). If level differences between successive tones lead to increases in asynchrony detection thresholds, this would suggest that obligatory stream segregation can be produced by level differences alone. Comparison of the magnitude of any such effects with the effects observed by Roberts *et al.*, together with the results of experiment 1, would allow us to determine whether any of their results could be explained purely in terms of differences in effective excitation level.

A. Stimuli

The stimuli were very similar to the C- and R-phase complex-tone maskers used in experiment 1. They differed only in that they were shorter—60 ms including onset and offset ramps of 10 ms—and they were presented over a range of different levels. Tones in one set (A tones) were always presented at an overall level of 70 dB SPL, and tones in the second set (B tones) were presented at levels that were 0, 1, 3, 5, 10, or 15 dB lower. Pairs of tones always had the same phase i.e., C with C and R with R. Continuous pink noise was added to mask possible combination tones (Smoorenburg, 1972; Pressnitzer and Patterson, 2001). The noise was produced using an Ivie model IE-32 DB noise generator. The noise had a spectrum level of 27 dB at 100 Hz, and it was low-pass filtered at 800 Hz, using a Kemo VBF8/04 filter (slope 96 dB/oct). The rest of the equipment was identical to that used in experiment 1, and the pink noise was added to the other stimuli using the same TDT SM3 summer that was used to mix the A and B tones.

B. Procedure

Obligatory streaming was examined using the rhythm discrimination procedure of Roberts *et al.* (2002) (experiment 1), which was itself based on the paradigm of Cusack and Roberts (2000) (experiment 2). An alternating sequence of A and B tones was presented. Subjects were required to detect the change in the relative timing of the two sets of tones. In a control sequence, the timing of the A and B tones stayed perfectly regular: there was 40 ms of silence following each 60 ms tone, resulting in one onset every 100 ms. Twelve tones of each type were presented, making the whole stimulus sequence 2400 ms in duration. In a test sequence, a delay (ΔT) was introduced cumulatively between the B tones. A schematic diagram of the stimulus sequence and possible percepts is shown in Fig. 1 of Roberts *et al.* (2002). The first portion of the sequence (the first six pairs of tones) maintained isochrony, being identical to the control sequence during this portion. This was to allow for the build-up of stream segregation, which can take a little time; the build-up can be as long as 3–4 s, but for the rapid sequences of brief tones used here, it tends to occur much more rapidly than this (Bregman, 1978; Bregman *et al.*, 2000). During the second portion of the test sequence, over the next four pairs of tones, a cumulative delay was introduced. A total delay of $4\Delta T$ was reached and maintained for the final two pairs of tones, which were maximally anisochronous. The delay was introduced progressively in this manner to minimize the usefulness of attending to the timing of B tones alone, in preference to attending to the overall temporal pattern. An adaptive two-interval forced-choice procedure was used, where subjects were required to select the randomly assigned interval that contained the test sequence. The delay was adjusted logarithmically, using a three-down one-up rule (Levitt, 1971). A run terminated when six turnpoints had been obtained. The initial delay was 10 ms. This was the maximum possible, because the total cumulative delay would be 40 ms, which is the longest delay that could fit in the initial 40 ms offset–onset interval between the B and A tones. For the first two turnpoints, delays were adjusted by a factor of 1.414 (square root of 2). For the final four turnpoints, delays were adjusted by a factor of 1.189 (fourth root of 2). The recorded threshold was taken as the geometric mean of the delays at the last four turnpoints. If the standard deviation of the logarithmic values of the last four turnpoints exceeded 0.2, that run was rejected and a replacement run was conducted. There were no occurrences of this beyond the initial training period. The final threshold recorded for each condition was the geometric mean of the values from three such runs.

C. Results and discussion

The pattern of results was broadly similar across subjects, and Fig. 4 shows the group averages (geometric means) across all six subjects. The asynchrony detection threshold, expressed as cumulative delay ($4\Delta T$), is plotted

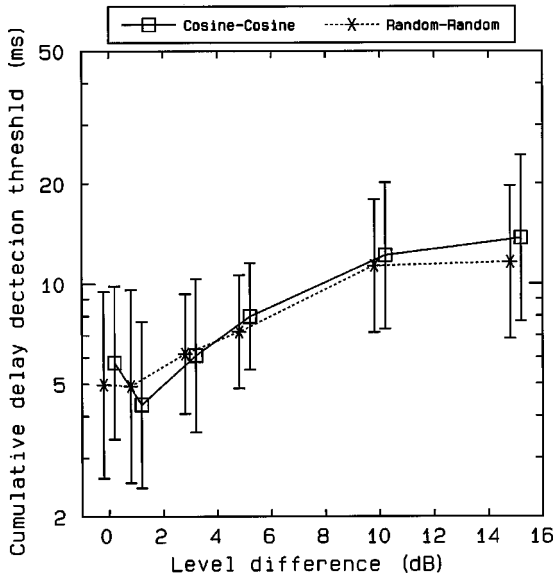


FIG. 4. The delay detection thresholds from experiment 2. The ordinate shows the total cumulative delay (i.e., $4\Delta T$) on a logarithmic scale, while the abscissa shows the level difference (in dB) between the A and B tones. The values plotted are the geometric means of the data for all subjects. Error bars show (geometric) standard deviations across subjects.

as a function of the level difference between the A and B tones. The thresholds increase with increasing level difference once the difference exceeds 3 dB. A two-way within-subjects ANOVA was performed with factors level difference (0, 1, 3, 5, 10, or 15 dB) and phase (C–C, or R–R). The effect of level difference was highly significant; $F(5,25) = 28.47$; $p < 0.001$. The mean thresholds were: 0 dB–5.84, 1 dB–5.03, 3 dB-6.44, 5 dB-8.05, 10 dB-12.35, 15 dB-13.57 ms. The effect of phase was not significant; $F(1,5) = 2.96$; $p = 0.146$. The mean thresholds were: R–R—8.16, C–C—8.92 ms. The interaction between phase and level difference was not significant; $F(5,25) = 1.06$, $p = 0.404$.

A one-tailed LSD test was used to assess the significance of the differences between mean thresholds for the level difference factor. The use of a one-tailed test is justified by our hypothesis that an increase in level difference would increase the delay detection threshold. Pairwise comparisons showed that there was a significant difference between the 0 and 5 dB conditions [$t(25) = 2.33$, $p < 0.05$] and between the 5 and 10 dB conditions [$t(25) = 4.54$, $p < 0.001$]. However, the difference between the 10 and 15 dB conditions was not significant [$t(25) = 1.29$].

It seems likely that the effects of level differences on rhythm discrimination found here were due mainly to stream segregation based on the level differences. However, it is important to consider alternative possibilities. One alternative explanation is that the level differences *per se* led to poorer discrimination. The rhythm discrimination task may be considered as comparable to the task of detecting changes in the silent interval between successive tones. Interval discrimination is impaired by large differences in level between successive stimuli (Divenyi and Danner, 1977). However, the data of Divenyi and Danner show that the threshold for discrimination of a silent interval worsens by only a factor of about 1.5 when the level difference of the ''markers'' bounding the interval increases from 0 to 25 dB (data were not presented for smaller level differences). Our data show that the delay detection threshold increased by a factor of about 2.3 when the level difference between successive tones was increased from 0 to 10 dB. Therefore, it seems unlikely that the effects of level *per se* on interval discrimination can account for all of the effect found in our experiments.

Another possibility is that there was forward masking of the lower-level (B) tones by the higher-level (A) tones. However, the silent interval between the A tones and the B tones was always at least 40 ms. Given that the level of the A tones was 70 dB SPL, the masked threshold of a *brief* tone pip presented 40 ms after the A tones would be below about 40 dB SPL (Jesteadt *et al.*, 1982; Moore and Glasberg, 1983a). Since the level of the B tones was always 55 dB or more, we can conclude that even the onsets of the B tones would not have been masked by the A tones.

In summary, the results suggest that level differences can produce obligatory stream segregation. The effect was significant with a level difference of only 5 dB. This is smaller than the difference in effective excitation level of 12.6 dB between the C- and R-phase tones, as estimated from experiment 1. Therefore, the results suggest that at least part of the obligatory streaming found by Roberts *et al.* (2002) for alternating C- and R-phase tones with the same physical level might have been the result of differences in effective excitation level.

IV. EXPERIMENT III: SUBJECTIVE MEASUREMENT OF STREAMING BASED ON LEVEL DIFFERENCES

Experiment 2 provided an indirect yet objective measure of stream segregation based on level differences alone. The third experiment also investigated stream segregation produced by level differences, but this time using subjective reporting of the percept—whether a pattern of alternating tones appeared to be integrated or segregated. As well as providing a direct measure of stream segregation, the use of this commonly employed paradigm allowed direct comparison with other studies using level differences (van Noorden, 1975).

A. Stimuli

The stimuli were the same as those used in experiment 2, having either C or R component phase. The complex tones had a total duration of 60 ms, including 10-ms raised-cosine onset and offset ramps. There were two alternating tones, A and B, in each stimulus sequence; A and B had the same phase characteristic (both C or both R). The particular set of random phases for the R–R stimuli was chosen independently at the start of each trial, and was different for the A and B tones. Possible consequences of this are discussed later. The A tones had a level of 70 dB SPL, while the B tones had a level 0, 1, 3, 5, 10, or 15 dB lower. To mask possible combination tones, pink noise with a spectrum level of 27 dB at 100 Hz, low-pass filtered at 800 Hz, was added. The equipment used was identical to that used in experiment 2.

B. Procedure

The procedure was essentially the same as that used by Roberts (2002) (experiment 2), which was itself based on the stimulus sequence of van Noorden (1975) and the method of Anstis and Saida (1985). The tones were presented in an ABA_ABA_... pattern. There was 40 ms of silence between each tone, except for the 140 ms silences denoted by _. Thus, one ABA_ cycle lasted 400 ms. This pattern was presented 75 times, making the total duration of each sequence 30 s. The sequence could be perceived as integrated, with a characteristic ''galloping'' rhythm, or as segregated, in which case two streams were heard, each of which had a regular isochronous rhythm, with the A stream twice as fast as the B stream. During an extended sequence, the percept could ''flip'' backwards and forwards between integration and segregation. Thus, a measure of the time that each was perceived was used. Subjects pressed a single button on the response box whenever the percept changed between integration and segregation. The current status was shown on the response box using two LEDs. Subjects were told they were free to press the button as many times as they liked during a trial. If the percept was ambiguous, subjects were asked to select whichever was dominant. The measure recorded was the percentage of time that segregation was perceived.

Twelve combinations of stimuli were used, with six level differences (where tone B was 0, 1, 3, 5, 10, or 15 dB lower in level than tone A), and two phase combinations (C–C and R–R). Each stimulus combination was presented six times, making a total of 72 runs, which were all completed within a single session. Subjects were given practice with a few example trials, including ones with 0 and 15 dB level difference, to familiarize them with the sounds of the combinations least and most likely to segregate, respectively. The final segregation percentages used were the arithmetic mean across the six trials.

C. Results

The general pattern of the results was similar across subjects, except that two subjects indicated very little stream segregation for the C–C stimuli, whatever the level difference between them. Also, one (different) subject indicated some stream segregation for the C–C stimuli even when there was no level difference. Figure 5 shows the arithmetic mean scores across subjects. The percentage of time that the subjects reported a segregated percept is plotted as a function of the level difference between the A and B tones. The pattern of results is not as clear as for the objective measurement task of experiment 2. The subjective nature of experiment 3 apparently led to greater variability in the results. However, there is a general trend for segregation to increase with increasing level difference. This was supported by the results of a within-subjects ANOVA, with factors level difference (0, 1, 3, 5, 10, 15 dB between A and B) and phase (C–C, or R–R). The main effect of level difference was highly significant; $F(5,25)=8.20$; $p=0.001$. The mean scores were: 0 dB—10.4, 1 dB—12.8, 3 dB—12.5, 5 dB—23.0, 10 dB—32.0, 15 dB—45.2%. Pairwise comparisons within the level difference factor were conducted using an LSD test.

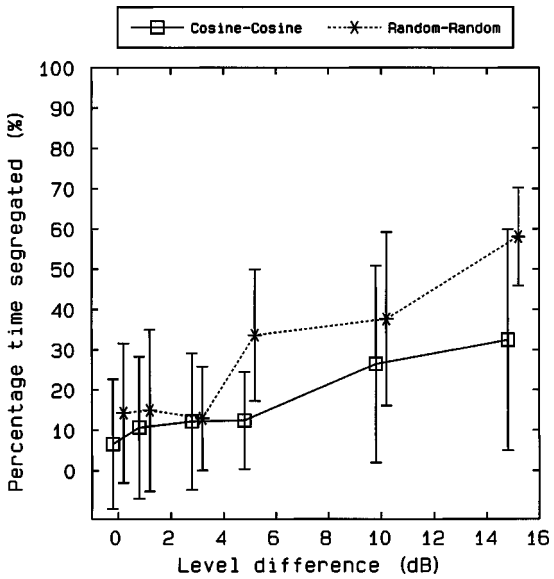


FIG. 5. Subjective streaming results for experiment 3. The percentage of time that a 30 s sequence of ABA_ABA_... tones was perceived as segregated is plotted as a function of the level difference (in dB) between the A and B tones. The arithmetic means across subjects are shown. Error bars show standard deviations across subjects.

Again, a one-tailed t-test was used (critical value=1.708; $p<0.05$), given the assumption that segregation would increase as level difference increased. There was a significant increase in segregation between the 0 and 5 dB conditions [$t(25)=1.85$], and also between the 5 and 15 dB conditions [$t(25)=3.26$].

There was also a significant main effect of phase; $F(1,5)=7.46$; $p=0.041$. Mean scores were: R–R—28.6, C–C—16.8%. There was no significant interaction between phase and level difference.

D. Discussion

The smaller amount of stream segregation for the C-phase stimuli may be related to the finding that, in experiment 1, the growth-of-masking functions were less steep for C-phase than for R-phase maskers (see Figs. 2 and 3). Similar effects were reported by Gockel *et al.* (2003) and by Moore *et al.* (2004). This may indicate that the rate of change of excitation level with physical level is lower for C-phase than for R-phase stimuli. Hence, in our experiment 3, the physical difference in level between the A and B tones may have led to smaller differences in effective excitation level for the C-phase tones than for the R-phase tones, and this may, in turn, have led to reduced subjective segregation for the C-phase than for the R-phase tones. However, it is curious that no comparable effect was found in experiment 2.

Another possible factor contributing to the phase effect found in experiment 3 is connected with the fact that, in the R–R phase condition, different phase randomizations were used for the A and B tones. When there was 0 dB level difference, only one subject reported any stream segregation for the C–C stimuli, yet three subjects reported some degree of segregation for the R–R stimuli. This could have been due to the different phases for the A and B tones in the R–R stimuli, which may have led to small timbre differences.

These timbre differences may also have contributed to stream segregation when there was a difference in level of the A and B tones.

The results found in experiment 3 are broadly consistent with the results of van Noorden (1975) for sequences of pure tones that alternated in level. He found that the fission boundary, the smallest difference in level required to hear two streams, was about 4 dB when the time between the onsets of successive tones was 100 ms (the same as in our experiment).

V. GENERAL DISCUSSION

The experiments reported here were motivated by the study of Roberts *et al.* (2002). They set out to test the peripheral channeling hypothesis of Hartmann and Johnson (1991), by seeing if stream segregation could be produced using stimuli that excited identical auditory channels but differed in temporal structure. Their results showed that robust stream segregation occurred between pairs of stimuli that differed only in the relative phases of their components. This was interpreted as indicating that stream segregation can be induced by differences in temporal structure alone. The results presented here suggest that, for some, but not all, of the stimulus pairs used by Roberts *et al.* (2002), the stream segregation might have been partly caused by differences in effective excitation level.

In experiment 1, we measured the forward masking produced by the stimuli of Roberts *et al.* (2002), using a range of levels around the level used by them. It was found that the C- and A-phase stimuli produced almost identical forward masking, for all masker levels. Thus, the stream segregation found by Roberts *et al.* (2002) for alternating C- and A-phase stimuli was probably based entirely on differences in temporal structure between the C- and A-phase stimuli, and not on differences in effective excitation level.

Experiment 1 also showed that the R-phase stimuli produced significantly more forward masking than the C- and A-phase stimuli. From the growth-of-masking functions, we estimated the difference in effective excitation level between the C-phase and R-phase maskers to be 12.6 dB. Thus the stream segregation observed previously for alternating C- and R-phase stimuli (Roberts *et al.*, 2002) might have been partly based on differences in effective excitation level.

Experiment 2 demonstrated that obligatory stream segregation, as measured with a rhythm discrimination task, occurs with complex stimuli that differ only in level. Experiment 3 showed similar effects for the subjective reporting of stream segregation. Both experiments showed that a level difference of 5 dB is sufficient to induce some stream segregation. Given that the effective excitation level of the C-phase and R-phase stimuli was estimated in experiment 1 to differ by over 10 dB, these results suggest that some of the stream segregation observed previously for alternating C- and R-phase stimuli (Roberts *et al.*, 2002) could have been caused by differences in effective excitation level.

We consider next whether the differences in effective excitation level estimated in experiment 1 are of sufficient magnitude to account completely for the degree of stream segregation observed for alternating C- and R-phase stimuli. The results of experiment 2 showed that a level difference of 10–15 dB (which spans the estimated difference in effective excitation level between the C- and R-phase stimuli) gives a delay detection threshold of about 12 ms. However, the results of Roberts *et al.* (2002) show a delay detection threshold of 26.5 ms for alternating C- and R-phase stimuli with identical levels. This implies that, while a difference in effective excitation level may have made a contribution to the streaming observed, it is not sufficient to wholly account for it. Hence differences in temporal structure must have made a contribution to the obligatory stream segregation observed for alternating C- and R-phase stimuli.

Comparison of the subjective ratings across studies yields a somewhat different conclusion. In experiment 3, the percentage of time that subjects reported the sounds as segregated was 26%–38% (mean 32%) for a level difference between the A and B tones of 10 dB, which corresponds roughly to the difference in effective excitation level between the C- and R-phase stimuli. For alternating C- and R-phase stimuli with the same passband as ours, Roberts *et al.* (2002) found a percentage of segregation of only about 15%. Thus, differences in effective excitation level appear to be sufficient to account entirely for the effects of phase on subjective streaming found by Roberts *et al.* for alternating C- and R-phase stimuli. The difference across studies (32% here versus 15% in Roberts *et al.*) may be due to individual differences, which can be substantial in tasks involving subjective ratings of streaming (see the large error bars in Fig. 5).

The results of Stainsby *et al.* (2004) support the idea that stream segregation can be produced by differences in temporal structure alone. They conducted experiments using stimuli and methods very similar to those of Roberts *et al.* (2002), and to experiments 2 and 3 of the present paper, but they tested subjects with cochlear hearing loss. For such subjects, the difference in forward masking produced by C-phase and R-phase maskers is much smaller than normal, and in some cases there is no difference at all (Moore *et al.*, 2004). Stainsby *et al.* (2003) found that stream segregation due to differences in phase spectrum was robust in subjects with cochlear hearing loss, and comparable to that found previously in young normally hearing subjects. This was the case both for alternating C- and A-phase stimuli and for alternating C- and R-phase stimuli. The finding that both obligatory and subjective stream segregation were similar for hearing-impaired and normally hearing subjects for the alternating C- and R-phase stimuli, suggests that differences in effective excitation level between the C- and R-phase stimuli do not play a major role in producing the stream segregation.

Overall, the present results are consistent with a growing body of evidence indicating that stream segregation does not require nonoverlapping excitation patterns of successive sounds in the peripheral auditory system. Stream segregation can be produced by differences between successive sounds in: (1) the F0 of complex tones containing only unresolved harmonics (Vliegen *et al.*, 1999; Vliegen and Oxenham, 1999; Grimault *et al.*, 2000); (2) the modulation rate of amplitude-modulated white noise (Grimault *et al.*, 2002); (3) temporal structure produced by manipulation of component

phase, at least for C- versus A-phase stimuli (Roberts *et al.*, 2002; Stainsby *et al.*, 2004); and (4) level (Dowling, 1973; van Noorden, 1975; and experiments 2 and 3 of the present study). Generally, the results are consistent with the proposal of Moore and Gockel (2002) that *any* sufficiently salient perceptual difference between successive tones can induce stream segregation.

ACKNOWLEDGMENTS

This research was supported by the Medical research Council (UK).

Alcántara, J. I., Moore, B. C. J., Glasberg, B. R., Wilkinson, A. J. K., and Jorasz, U. (**2003**). ''Phase effects in masking: Within- versus across-channel processes,'' J. Acoust. Soc. Am. **114**, 2158–2166.

ANSI (**1969**). *S3.6. Specifications for Audiometers* (American National Standards Institute, New York).

Anstis, S., and Saida, S. (**1985**). ''Adaptation to auditory streaming of frequency-modulated tones,'' J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

Beauvois, M. W., and Meddis, R. (**1996**). ''Computer simulation of auditory stream segregation in alternating-tone sequences,'' J. Acoust. Soc. Am. **99**, 2270–2280.

Bilsen, F. A. (**1973**). ''On the influence of the number and phase of harmonics on the perceptibility of the pitch of complex signals,'' Acustica **28**, 60–65.

Bregman, A. S. (**1978**). ''Auditory streaming is cumulative,'' J. Exp. Psychol. Hum. Percept. Perform. **4**, 380–387.

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

Bregman, A. S., Ahad, P. A., Crum, P. A., and O'Reilly, J. (**2000**). ''Effects of time intervals and tone durations on auditory stream segregation,'' Percept. Psychophys. **62**, 626–636.

Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Cusack, R., and Roberts, B. (**2000**). ''Effects of differences in timbre on sequential grouping,'' Percept. Psychophys. **62**, 1112–1120.

Divenyi, P. L., and Danner, W. F. (**1977**). ''Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra,'' Percept. Psychophys. **21**, 125–142.

Dowling, W. J. (**1973**). ''The perception of interleaved melodies,'' Cogn. Psychol. **5**, 322–337.

Gockel, H., Moore, B. C. J., and Patterson, R. D. (**2003**). ''Louder sounds can produce less forward masking: Effects of component phase in complex tones,'' J. Acoust. Soc. Am. **114**, 978–990.

Grimault, N., Bacon, S. P., and Micheyl, C. (**2002**). ''Auditory stream segregation on the basis of amplitude-modulation rate,'' J. Acoust. Soc. Am. **111**, 1340–1348.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2000**). ''Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency,'' J. Acoust. Soc. Am. **108**, 263–271.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2001**). ''Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing,'' Br. J. Audiol. **35**, 173–182.

Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Music Percept. **9**, 155–184.

Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.

Houtgast, T. (**1974**). ''Lateral suppression in hearing,'' Ph.D. thesis, Free University of Amsterdam.

Javel, E. (**1981**). ''Suppression of auditory nerve responses. I. Temporal analysis, intensity effects and suppression contours,'' J. Acoust. Soc. Am. **69**, 1735–1745.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). ''Forward masking as a function of frequency, masker level, and signal delay,'' J. Acoust. Soc. Am. **71**, 950–962.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

McCabe, S. L., and Denham, M. J. (**1997**). ''A model of auditory streaming,'' J. Acoust. Soc. Am. **101**, 1611–1621.

Moore, B. C. J., and Glasberg, B. R. (**1983a**). ''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., and Glasberg, B. R. (**1983b**). ''Masking patterns of synthetic vowels in simultaneous and forward masking,'' J. Acoust. Soc. Am. **73**, 906–917.

Moore, B. C. J., and Gockel, H. (**2002**). ''Factors influencing sequential stream segregation,'' Acta Acustica-Acustica **88**, 320–332.

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., Stainsby, T. H., and Tarasewicz, E. (**2004**). ''Effects of masker component phase on the forward masking produced by complex tones in normally hearing and hearing-impaired subjects'' (in press).

Plomp, R. (**1964a**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R. (**1964b**). ''The rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282.

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).

Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.

Rhode, W. S., and Recio, A. (**2001a**). ''Basilar-membrane response to multicomponent stimuli in chinchilla,'' J. Acoust. Soc. Am. **110**, 981–994.

Rhode, W. S., and Recio, A. (**2001b**). ''Multicomponent stimulus interactions observed in basilar-membrane vibration in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **110**, 3140–3154.

Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). ''Primitive stream segregation of tone sequences without differences in F0 or passband,'' J. Acoust. Soc. Am. **112**, 2074–2085.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.

Ruggero, M. A., 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.

Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.

Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.

Smoorenburg, G. F. (**1972**). ''Audibility region of combination tones,'' J. Acoust. Soc. Am. **52**, 603–614.

Stainsby, T. H., Moore, B. C. J., and Glasberg, B. R. (**2004**). ''Auditory streaming based on temporal structure in hearing-impaired listeners'' (submitted).

van Noorden, L. P. A. S. (**1975**). ''Temporal coherence in the perception of tone sequences,'' Ph.D. thesis, Eindhoven University of Technology.

Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). ''The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task,'' J. Acoust. Soc. Am. **106**, 938–945.

Vliegen, J., and Oxenham, A. J. (**1999**). ''Sequential stream segregation in the absence of spectral cues,'' J. Acoust. Soc. Am. **105**, 339–346.

Temporal effects in simultaneous masking with on- and off-frequency noise maskers: Effects of signal frequency and masker level

Sid P. Bacon and Sophie Savel
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 13 January 2003; revised 24 September 2003; accepted 26 January 2004)

Temporal effects in simultaneous masking were measured as a function of masker level for an on-frequency broadband masker and an off-frequency narrow-band masker for signal frequencies of 750, 1730, and 4000 Hz. The on-frequency masker was 10 equivalent rectangular bandwidths (ERBs) wide and centered at the signal frequency; the off-frequency masker was 500 Hz wide and its lower frequency edge was 1.038 ERBs higher in frequency than the signal. The primary goal of the study was to determine whether previously observed differences regarding the effects of signal frequency and masker level on the temporal effect for these two different types of masker might be due to considerably different signal levels at threshold. Despite similar masked thresholds, the effects of signal frequency and masker level in the present study were different for the two masker types. The temporal effect was significant for the two highest frequencies and absent for the lowest frequency in the presence of the broadband masker, but was more or less independent of frequency for the narrow-band masker. The temporal effect increased but then decreased as a function of level for the broadband masker (at the two higher signal frequencies, where there was a temporal effect), but increased and reached an asymptote for the narrow-band masker. Despite the different effects of signal frequency and masker level, the temporal effects for both types of masker can be understood in terms of a basilar-membrane input–output function that becomes more linear during the course of masker stimulation. [DOI: 10.1121/1.1689344]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ba [MRL] Pages: 1674–1683

I. INTRODUCTION

The threshold for a brief signal masked by a long-duration broadband noise can be 10–20 dB higher when that signal is located near the beginning of the masker than when it is located near the temporal center or end of the masker. Zwicker (1965a, b) dubbed this effect ''overshoot.'' Several possible explanations have been proposed for overshoot. One is that it represents a sharpening of the critical band with time (Green, 1969). According to this explanation, the decrease in threshold over time arises from the decrease in overall noise in the critical band. Although frequency selectivity measured psychophysically does indeed narrow with time (e.g., Bacon and Viemeister, 1985; Wright, 1997), the degree of sharpening is insufficient to account for such a large temporal effect. For example, recent estimates (Bacon *et al.*, 2002) indicate that the critical band or auditory filter decreases by less than a factor of 2, which would account for less than 3 dB of the temporal effect.

Another explanation for overshoot is that it represents peripheral adaptation. This is based on the fact that the response of an auditory-nerve fiber to a long-duration masker will decrease with time, whereas the response to a brief signal will not (Smith and Zwislocki, 1975; Smith, 1979), and thus the neural signal-to-masker ratio will increase with time. However, adaptation would not be expected to account for more than about 3–5 dB of the temporal effect (Smith and Zwislocki, 1975). Furthermore, because overshoot depends upon the masker energy outside the critical band centered at the signal frequency (Zwicker, 1965a; McFadden, 1989; Bacon and Smith, 1991), it is unclear whether adaptation of auditory-nerve fibers tuned to the signal frequency can account for *any* of the effect.

More recently, it has been suggested that overshoot is related to the compressive input–output (I–O) function of the basilar membrane (von Klitzing and Kohlrausch, 1994). This function is often characterized as linear or nearly linear at low and high levels, but compressive at levels between about 40 and 80 dB SPL (e.g., Ruggero and Rich, 1991). Because the level of the brief signal is usually much higher than the overall level of the noise in the auditory filter centered at the signal frequency, the signal and masker sometimes will be on different portions of the I–O function. The especially high thresholds for a signal at or near masker onset (Bacon, 1990) appear to occur when the masker is on the lower, linear portion of the function and the signal is on the middle, compressive portion (von Klitzing and Kohlrausch, 1994; Strickland, 2001; Bacon and Oxenham, 2004). Although compression may play an important role in overshoot, it is important to note that basilar-membrane mechanics do not—by themselves—change over time, and thus some other factor must be involved in the decrease in threshold over time. One possibility is that the amount of basilar-membrane compression decreases by means of an efferent feedback loop mediated via the influence of the medial olivocochlear bundle on the outer hair cells (Schmidt and Zwicker, 1991; Turner and Doherty, 1997; Bacon and Liu, 2000; Strickland, 2001).

The compression explanation is appealing because it can account for several aspects of overshoot, including the effects of signal duration, signal frequency, masker bandwidth, masker level,[1] and cochlear hearing loss (see von Klitzing and Kohlrausch, 1994). Some of these effects are discussed below in more detail.

A large temporal effect in simultaneous masking also exists for off-frequency narrow-band noise or tonal maskers (e.g., Green, 1969; Fastl, 1977; Bacon and Viemeister, 1985), and there are certain similarities between that effect and overshoot (see Bacon and Healy, 2000, for a review). For example, they both (1) have time courses on the order of 100–200 ms; (2) depend primarily upon masker energy above the signal frequency; and (3) are reduced by cochlear hearing loss. These similarities might be construed as evidence for similar underlying mechanisms, and hence as evidence that basilar-membrane compression also plays a role in the temporal effect with tonal maskers. However, there are several differences in the temporal effect with the two different types of masker which seemingly cast doubt on that possibility. Two of those are of particular interest here, namely the effects of signal frequency and masker level.

The amount of overshoot depends dramatically upon signal frequency. It is typically negligible for frequencies below about 1000 Hz (but see McFadden, 1988), but can be 20 dB or larger for frequencies considerably higher than that (Zwicker, 1965a; Fastl, 1976; Elliott, 1965; Bacon and Takahashi, 1992). This frequency effect is consistent with physiological (Cooper and Yates, 1994; Cooper and Rhode, 1995) and psychophysical (Lee and Bacon, 1998; Hicks and Bacon, 1999; Plack and Oxenham, 2000) evidence that the degree of nonlinearity (and hence compression) is reduced in the apical region of the cochlea (for recent evidence that compression may not be reduced at low frequencies, however, see Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003). The temporal effect with off-frequency tonal maskers, on the other hand, is essentially independent of signal frequency for frequencies from 500 to 4000 Hz (Bacon and Moore, 1986), and thus inconsistent with a frequency- or place-dependent nonlinearity.

As the level of a broadband noise masker increases from low to high levels, overshoot first increases but then decreases and often disappears at the highest masker levels (Bacon, 1990; von Klitzing and Kohlrausch, 1994). The temporal effect with off-frequency tonal maskers, on the other hand, increases with increases in masker level and either continues to increase or reaches an asymptote at the highest masker levels tested (Green, 1969; Bacon and Viemeister, 1985; Bacon and Moore, 1986). The level effects with broadband noise maskers have been explained in terms of a basilar-membrane I–O function that is compressive at moderate levels but linear or at least less compressive at high levels (Strickland, 2001; Bacon and Oxenham, 2004). It is unclear whether the very different effect of level with off-frequency tonal maskers can be similarly understood.

Despite the differing effects of signal frequency and masker level for the two different types of masker, it may be premature to rule out the possibility that basilar-membrane compression is also involved in the temporal effect with off-frequency narrow-band maskers. In particular, it is possible that the conditions were not optimal in the previous studies of tonal masking to observe effects of signal frequency and masker level similar to those observed with noise maskers. For example, to observe an effect of signal frequency, it seems likely that the signal should be at a level where compression is maximal at each frequency. This condition has probably been met in the overshoot experiments, as the signal levels typically have been within the 60–80 dB range. In the tonal masking study by Bacon and Moore (1986), however, the signal level at masked threshold was either just below or only somewhat above the estimated compression threshold of 40 dB (averaged across subjects, the average signal level was 39.6 dB SPL for the 60-dB masker and 51.8 dB SPL for the 80-dB masker). Because the degree of compression tends to reach its maximum at levels well above the compression threshold (Ruggero *et al.*, 1997; Oxenham and Plack, 1997; Wojtczak *et al.*, 2001; Nelson *et al.*, 2001), it is highly likely that the signals were not at the optimal level to observe the potential influence of compression on the temporal effect with off-frequency narrow-band maskers.

Along those same lines, it is possible that the temporal effect with tonal maskers has not declined at high levels because the signal has never been at a sufficiently high level. The decline in overshoot at high masker levels generally corresponds to signal levels above about 75–80 dB SPL, where the signal is likely near or on the more-linear portion of the I–O function while the masker is on the compressive portion. The signal in the tonal masking experiments, on the other hand, rarely reaches a level near 60 dB SPL. This is because the frequency of the masker is typically well above the frequency of the signal. This maximizes the temporal effect but minimizes the amount of masking. Another possible reason why the effect of level is different for on- and off-frequency maskers is related to the form of the underlying I–O functions. Although it is reasonable to assume that the masker and signal operate on the same I–O function (albeit at different levels) when the masker spectrally overlaps the signal, the same may not be true when the masker and signal do not overlap spectrally. Thus, if the signal and off-frequency masker operate on I–O functions that are different in shape from one another, then the influence of level on the temporal effect with an off-frequency masker would not be expected to mimic that seen with broadband noise maskers.

The primary purpose of the present study was to determine whether the differences in the effects of signal frequency and masker level described above for the two masker types are due to differences in the level of the signal at masked threshold. To achieve high signal levels in the off-frequency masking condition, a narrow-band noise masker was used, because noise maskers can produce more masking than tonal maskers when the masker is higher in frequency than the signal (Mott and Feth, 1986; Glasberg and Moore, 1994). Furthermore, to achieve as much masking as possible, the noise was positioned closer in frequency to the signal than is typically done with tonal maskers. A secondary purpose was to evaluate whether the temporal effect with both types of masker might be understood in terms of I–O func-

tions that become less compressive during the course of masker stimulation.

II. METHOD

A. Apparatus and stimuli

All stimuli were digitally generated and produced at a 50-kHz sampling rate using a digital array processing card (TDT AP2) and one channel of a digital-to-analog converter, or DAC (TDT DD1). The output of the DAC was low-pass filtered at 8 kHz (TDT FT6), attenuated (TDT PA4), and routed via a headphone buffer (TDT HB6) to a Sennheiser HD250 headphone.

The signal was a pure tone of one of three frequencies equally spaced on a logarithmic scale (750, 1730, or 4000 Hz). For each signal frequency, two maskers were used. The on-frequency masker was arithmetically centered at the signal frequency with a width of 10 critical bands or equivalent rectangular bandwidths (ERBs; Glasberg and Moore, 1990). The frequency ranges of the maskers for the three signal frequencies were 220–1280 Hz, 675–2785 Hz, and 1720–6280 Hz, respectively. These bandwidths should be sufficiently wide to yield maximum overshoot (Zwicker, 1965a; Bacon and Smith, 1991). The off-frequency masker was 500 Hz wide and was positioned such that its lower frequency edge was 1.038 ERBs above the signal frequency. The frequency ranges of the maskers were 866–1366 Hz, 1962–2462 Hz, and 4500–5000 Hz, respectively. For both the on- and off-frequency noises, a new sample was presented during each observation interval.

The 10-ms signal was presented either at the beginning (0-ms delay) or at the temporal center (195-ms delay) of the 400-ms masker. All stimulus durations included 5-ms $\cos^2$ rise/fall times, and all durations and delays were determined from zero-voltage points. For the 4000-Hz signal, the masker spectrum level ranged from -5 to 45 dB SPL for the on-frequency masker and from 15 to 69 dB SPL for the off-frequency masker. Those ranges were truncated somewhat on the low end for the other signal frequencies.

B. Procedure

Subjects were tested in a single-walled, sound-attenuating booth located within a sound-treated room. Thresholds were measured using an adaptive two-interval, forced-choice procedure that estimates the 79.4%-correct point on the psychometric function (Levitt, 1971). The signal was presented in one of the two intervals, and the subject's task was to indicate the interval that contained the signal. Lights were used to indicate when the signal might occur and to provide correct-answer feedback. The two observation intervals were separated by 500 ms. A run started with the signal level 10–15 dB above the estimated threshold; the level was decreased following three correct responses and increased following one incorrect response. The step size was initially 5 dB, but was decreased to 2 dB following the second reversal. Each run consisted of 12 reversals; the threshold estimate for that run was the signal level averaged over the last ten reversals. A run was included only if the standard deviation of the threshold estimate was 5 dB or less. Each threshold reported is the mean of at least three runs, obtained over 3 separate days. If the standard deviation of this mean exceeded 3 dB, an additional estimate was obtained and included in the mean. This continued until the standard deviation was less than 3 dB or until six estimates were obtained. Ninety-four percent (271/288) of all the means presented here had standard deviations less than 3 dB.

C. Subjects

Four individuals participated. They ranged in age from 21–28 years, and had thresholds of 15 dB HL or lower (ANSI, 1996) for octave frequencies from 250 to 8000 Hz. Subject S1 (the second author) had prior experience as a subject in similar experiments on temporal effects in simultaneous masking. The other subjects had 6 h of practice (2 h for each of 3 days) prior to data collection. The practice consisted of detecting a signal at masker onset, for all masker levels and both masker types.

III. RESULTS AND DISCUSSION

A. Experimental results

The first two figures show the individual results in terms of growth-of-masking functions for the conditions in which the signal was either at the temporal center (Fig. 1) or at the onset (Fig. 2) of the masker. Each column represents a different masker type and each row a different signal frequency. These figures serve two purposes. First, they illustrate how the slope or even the form of the masking functions might be influenced by signal delay, signal frequency, and masker type. Second, they allow a comparison of the signal level at threshold for the two masker types.

For the most part, the functions appear to be described reasonably well by a single straight line. However, there are clear exceptions to this in Fig. 2, where some of the results appear to be better represented by at least two segments. Consequently, the masking functions in both figures were fitted with the one, two, or three straight lines chosen to minimize the squared deviations. An additional line was included in the fit only if the variance accounted for increased by at least 4.0%. The results of these regression analyses are summarized in Table I. For the most part, the fit chosen was a single straight line, although occasionally it was two straight lines; in no instance did a third line improve the fit by the criterion amount.

When the signal was at the temporal center of the masker (Fig. 1), all the masking functions were fitted quite well by a single line. For the on-frequency masker, the slope of the masking function for all signal frequencies was about 1.0. This is typical of simultaneous-masking results with broadband maskers in steady-state conditions [i.e., with long signals and long maskers (e.g., Hawkins and Stevens, 1950) or short signals presented well after the onset of a long masker (e.g., Bacon, 1990)]. For the off-frequency masker, the slope of the masking function was typically less than 1.0, and it decreased with increasing signal frequency. The slope averaged across subjects was 0.94, 0.87, and 0.73 for the signal frequencies of 750, 1730, and 4000 Hz, respectively. Others also have found that the slope of the masking func-

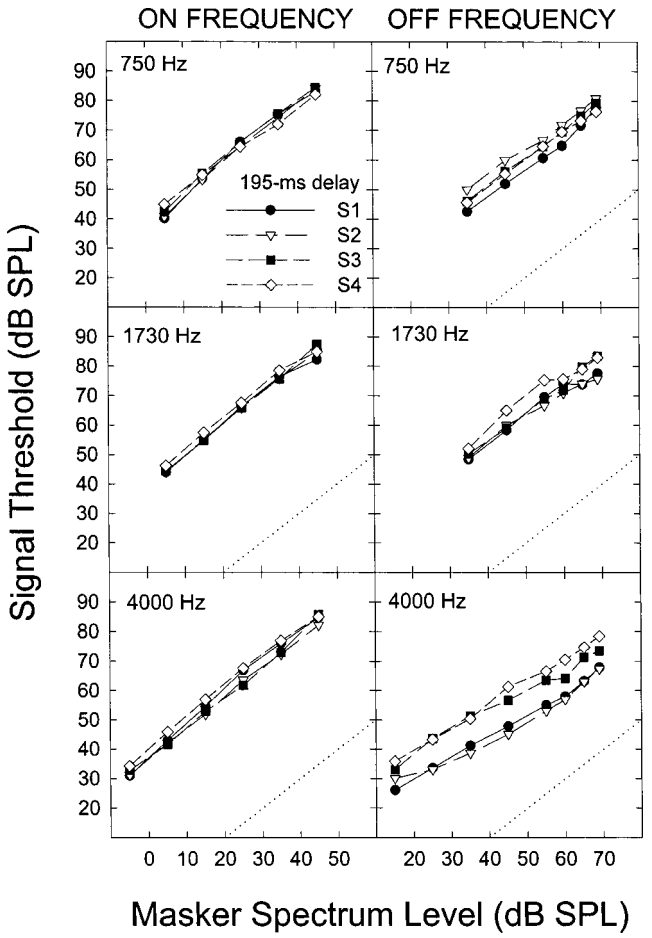


FIG. 1. Individual growth-of-masking functions for a 10-ms signal located at the temporal center of a 400-ms masker. In the left column, the masker was a noise centered at the signal frequency with a width of 10 critical bands (see the text for cutoff frequencies). In the right column, the masker was a noise with a width of 500 Hz; the lower cutoff frequency of the noise was 1.038 ERBs higher in frequency than the signal. The signal frequency was 750 Hz (top row), 1730 Hz (middle row), or 4000 Hz (bottom row). The quiet thresholds (dB SPL) for the four subjects at those three frequencies were: 16.9, 23.7, 23.1, 23.3 (750 Hz); 17.3, 22.2, 28.7, 26.0 (1730 Hz); and 15.1, 20.1, 25.3, 23.3 (4000 Hz). The dotted line without symbols is a linear reference.

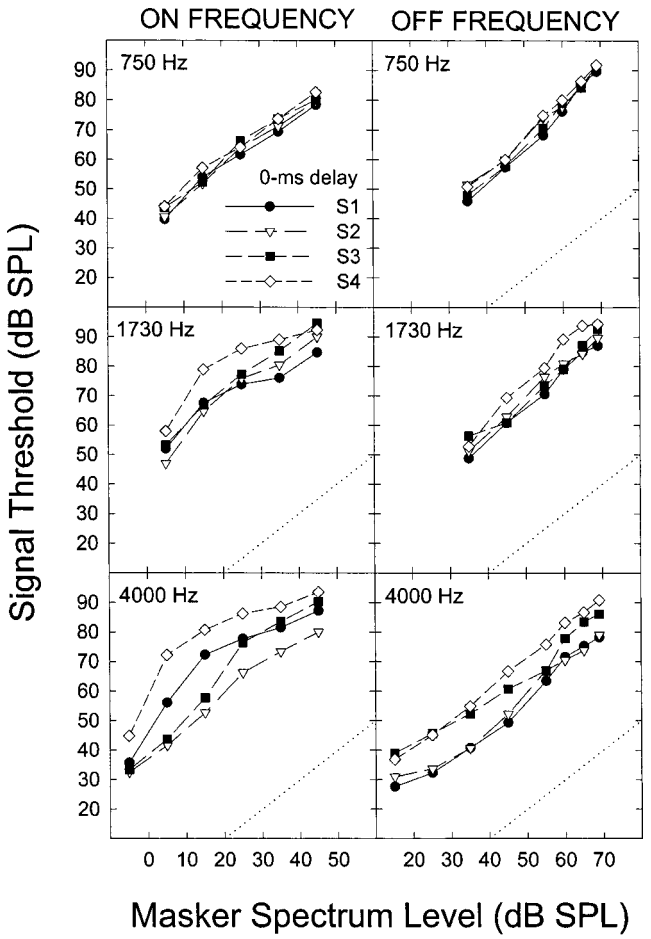


FIG. 2. As in Fig. 1, but for a signal located at the onset of the masker.

tion is less than 1.0 when the masker is higher in frequency than the signal (e.g., Wegel and Lane, 1924; Egan and Hake, 1950).

When the signal was at the onset of the masker (Fig. 2), a few of the masking functions were fitted better by two straight lines than by one; these were restricted to the on-frequency condition at the two highest signal frequencies. For the off-frequency masker, the group average slope was 1.24, 1.16, and 0.97 for the signal frequencies of 750, 1730, and 4000 Hz, respectively. Thus, the slope of the off-frequency masking function decreased with increasing signal frequency, as observed when the signal was at the temporal center of the masker. This is in contrast to the situation with the on-frequency masker, where the slope (or the initial slope) was greater at the two highest frequencies (1.61 and 1.69) than at the lowest frequency (0.95). For the cases in which the masking function was fitted best with two lines, the fits included a line with a slope considerably greater than 1.0 at moderate signal levels and a line with a slope less than 1.0 at high signal levels. Others (e.g., Bacon, 1990) have observed similar, multisegment masking functions for a signal at the beginning of a broadband masker.

As seen in Figs. 1 and 2, there was considerable overlap in the masked thresholds among subjects for some conditions, whereas in others the thresholds diverged considerably. For the off-frequency masker, the masked thresholds more or less overlapped (750 Hz) or were displaced from one another by the same amount at both delays (1730 and 4000 Hz). Thus, any individual differences were more or less independent of signal delay. This was not true for the on-frequency masker. In this case, the masked thresholds for the four subjects overlapped at all three signal frequencies when the signal was at the temporal center of the masker (Fig. 1), but diverged from one another when the signal was at the onset of the masker (Fig. 2). Interestingly, they diverged only at the two highest signal frequencies, and most prominently at the highest frequency (4000 Hz). Others have noted considerable individual differences in conditions in which the signal is at or near the onset of a masker (e.g., Bacon, 1990; Wright, 1995, 1997).

The primary purpose of the present study was to compare the effects of signal frequency and masker level on the temporal effect for relatively broad on-frequency maskers with that for relatively narrow off-frequency maskers under conditions where the two maskers produce similar masked thresholds. As can be seen in Figs. 1 and 2, masked thresh-

TABLE I. Slope(s) and r^2 values corresponding to the linear regression fits to the masking functions in Figs. 1 and 2. A second segment (slope 2) was included in the fit if the variance accounted for increased by at least 4% (i.e., if r^2 increased by at least 0.04). See the text for more details.

Signal freq (Hz)		Masker type: On frequency S1	On frequency S2	On frequency S3	Off frequency S4	Off frequency S1	Off frequency S2	Off frequency S3	Off frequency S4
Delay = 195 ms									
750	Slope 1	1.08	1.08	1.05	0.91	1.00	0.89	0.97	0.91
	r^2	0.987	0.997	0.995	0.998	0.988	0.995	0.999	0.998
1730	Slope 1	0.98	1.02	1.07	0.99	0.87	0.78	0.97	0.87
	r^2	0.988	0.998	0.999	0.992	0.967	0.983	0.992	0.962
4000	Slope 1	1.09	1.02	1.05	1.02	0.75	0.70	0.71	0.79
	r^2	0.995	0.999	0.997	0.996	0.996	0.972	0.986	0.996
Delay = 0 ms									
750	Slope 1	0.93	0.98	0.94	0.94	1.30	1.17	1.26	1.23
	r^2	0.983	0.989	0.981	0.992	0.992	0.990	0.995	0.995
1730	Slope 1	1.56	1.78	1.01	2.09	1.16	1.13	1.10	1.26
	Slope 2	0.54	0.80	⋯	0.43	⋯	⋯	⋯	⋯
	r^2	0.988	0.994	0.98	0.996	0.994	0.992	0.969	0.979
4000	Slope 1	1.83	0.99	1.21	2.74	1.01	0.96	0.89	1.02
	Slope 2	0.49	⋯	⋯	0.50	⋯	⋯	⋯	⋯
	r^2	0.998	0.988	0.976	0.995	0.980	0.975	0.976	0.998

olds were indeed similar for the two types of masker. The temporal effect for these two maskers is shown in Fig. 3. Despite the individual differences discussed above, the effects of signal frequency and masker level on the temporal effect for the two masker types were generally similar across subjects, and thus the results are presented as an average across all four subjects. (The exception to this was the pattern of results for S2, who showed little or no temporal effect at 4000 Hz in the on-frequency condition. The group average was nearly the same, however, regardless of whether or not S2 was included.)

For the on-frequency masker, the temporal effect (overshoot) was negligible at the lowest signal frequency and was as large as 10–15 dB at the two highest signal frequencies. Furthermore, at the two highest frequencies, overshoot increased but then decreased at high levels. For the off-frequency masker, the temporal effect was more or less independent of signal frequency (the somewhat smaller effect at 1730 Hz was due to S1 alone), and it increased with increasing masker level until it reached an asymptote at a value of about 10–12 dB at the highest levels. For both masker types, the effects of signal frequency and masker level were similar to those seen in the literature (broadband noise masker: Zwicker, 1965a; Fastl, 1976; Elliott, 1965; Bacon, 1990; Bacon and Takahashi, 1992; von Klitzing and Kohlrausch, 1994; off-frequency narrow-band masker: Green, 1969; Bacon and Viemeister, 1985; Bacon and Moore, 1986), indicating that the differences between the two masker types cannot be explained by differences in signal level at threshold.

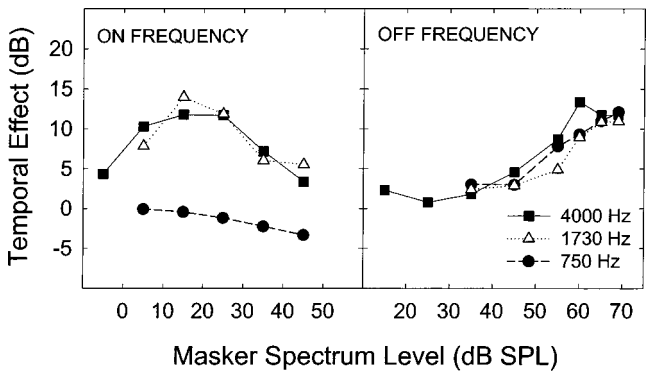


FIG. 3. Temporal effect averaged across all four subjects.

B. Model fits

A secondary purpose of the present study was to evaluate whether the temporal effect with both types of masker might be understood in terms of I–O functions that become less compressive over time as a function of masker stimulation. This was addressed by using essentially the same approach that was used successfully by Strickland (2001) to evaluate the possible role of basilar-membrane nonlinearity in overshoot. In this approach, internal signal and masker levels are the outputs of an I–O function representing the basilar-membrane response, and thresholds represent some criterion difference between those two outputs. In particular, the I–O function is described with the following equation:

$$L_{\text{out}} = (0.5 + (x \times 0.1)) \times L_{\text{in}} + 50 - (x \times 10) + 8.5 \times \left(1 - \left(\frac{1}{1 + \exp(-0.09 \times (L_{\text{in}} - 60))}\right)\right), \quad (1)$$

where L_{in} and L_{out} are input and output levels in dB, and x is a free parameter that determines the degree of compression reflected in the I–O function. Using this equation, Strickland showed that overshoot could be understood as a decrease in the amount of compression during the course of masker stimulation (reflected in the equation as an increase in the value of x). The idea is shown schematically in Fig. 4. The squares correspond to a broadband (on-frequency) masker,

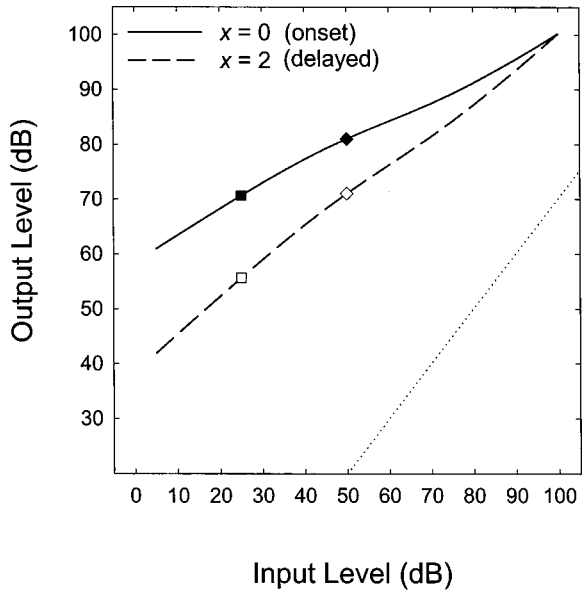


FIG. 4. Schematic showing how the linearization of an I–O function can result in a temporal effect for an on-frequency masker. The I–O functions are based on Eq. (1) in the text. The onset and delayed conditions are represented by the solid and dashed lines, respectively, and the broadband masker and high-frequency signal are represented by squares and diamonds, respectively. The dotted line without symbols is a linear reference. The I–O function is assumed to be more compressive ($x=0$) at masker onset and less compressive ($x=2$) later in time. At onset (filled symbols, solid line), the input signal and masker levels result in a signal-to-masker ratio at the output of about 10 dB. Later in time (unfilled symbols, dashed line), those same input levels result in a larger signal-to-masker ratio at the output. Thus, to yield a constant internal signal-to-masker ratio, the level of the signal can be decreased, and that decrease represents the temporal effect.

whereas the diamonds correspond to a high-frequency signal, which is greater in level than the masker because it is so brief. The solid curve ($x=0$) is assumed to reflect the I–O function at onset, and the filled symbols represent the masker and signal levels at threshold in the onset condition (the input levels correspond to the stimulus levels themselves, whereas the output levels correspond to the internal levels at the basilar membrane). The *internal* signal-to-masker ratio (SMR_i, the difference in output levels between signal and masker) at threshold is about 10 dB. If the I–O function were to become more linear during the course of masker stimulation (dashed line, $x=2$), perhaps as a result of efferent feedback on the outer hair cells, then those same masker and signal levels would result in a much larger SMR_i. Thus, to yield a constant SMR_i, the level of the signal can be decreased, and that decrease represents overshoot. Thus, in this scheme, overshoot reflects the fact that the decrease in compression affects the signal and masker differentially, because they are at different points on the I–O function.

Equation (1) was used to determine, among other things, whether the temporal effect for both masker types could be predicted by a change over time in the degree of compression at the basilar membrane. The details of the fitting procedure are described below. In all instances, the goal was to find the value of x that produced the smallest rms error between the obtained and predicted masked thresholds. Rather than using individual masking functions, the equation was fitted to the group average functions. Furthermore, in order to use the same range of masker levels for all signal frequencies, the lowest one (on-frequency) or two (off-frequency) masker levels at 4000 Hz were excluded from the analyses. This had the additional consequence of keeping the range of signal levels in the fits nearly the same for all conditions; that range was about 45–90 dB SPL. Because the fitting procedure was slightly different for the on- and off-frequency maskers, they are considered separately.

For the on-frequency masker, the signal and masker were assumed to be processed by the same I–O function (i.e., the same value of x was used for the signal and masker), although the masker was at a lower point on that function (as in Fig. 4). In these cases, the level of the masker was expressed as the overall level of the noise in the auditory filter centered at the signal frequency; the auditory filter bandwidth was assumed to be 106, 211, and 456 Hz at 750, 1730, and 4000 Hz, respectively (Glasberg and Moore, 1990). The first step in the fitting process was to estimate the SMR_i needed to detect the signal in the presence of the masker. This criterion SMR (SMR_{cr}) is thought to reflect a type of central processing efficiency that is presumably independent of masking condition. The SMR_{cr} was estimated by determining the SMR_i for the condition where the masker spectrum level was 5 dB, the signal frequency was 4000 Hz, and x was 0. Thus, Eq. (1) was solved for the masker when L_{in} was 31.6 dB (overall level in the auditory filter centered at 4000 Hz) and it was solved for the signal when L_{in} was 53.4 dB (the threshold averaged across the four subjects). The SMR_{cr} was the difference between the output (L_{out}) for the signal and the output for the masker. The criterion value obtained here (8.5 dB) was similar to that (10 dB) obtained by Strickland (2001), and it was used for all fitting procedures reported below.

The following iterative procedure was used to obtain the best-fitting x value for each signal frequency and signal delay. For each value of x in the iteration process, a predicted masking function was generated and evaluated in terms of how well it fitted the obtained masking function by: (1) determining the masker output for each successive masker level; (2) adding SNR_{cr} to each output value to obtain the series of signal outputs; (3) determining the input signal level (predicted signal level) necessary to achieve each signal output; and (4) comparing the predicted signal levels with the measured signal levels. The best-fitting x values and the associated rms values are shown in the top half of Table II. As can be seen from the small values in the rms column, the fits generally were quite good. Two points are worth noting regarding the values of x. First, in the onset condition, those values were smaller at the two highest signal frequencies than at the lowest signal frequency, suggesting that the I–O functions are more compressive at the higher frequencies (but see below). Second, for the two highest signal frequencies, where there was a clear temporal effect in the experimental data (Fig. 3), the value of x was larger (the I–O function was less compressive) at the longer signal delay than it was at the shorter delay. The reverse was true at the lowest signal frequency, where there was no temporal effect (or, if anything, the temporal effect was negative). Thus, as shown by Strickland (2001; also see Bacon and Oxenham,

TABLE II. The best-fitting values of x from Eq. (1) to fit the average growth-of-masking functions for both the on- and off-frequency maskers. For the on-frequency masker, x was forced to be the same for the signal (x_s) and the masker (x_m). For the off-frequency masker, the masking functions were fitted twice, once when x_s was forced to be the same as x_m and once where x_s and x_m could differ.

Signal freq (Hz)	Delay (ms)	$x(x_s=x_m)$	rms	x_s	x_m	rms
		On frequency				
750	0	1.4	1.76	⋯	⋯	⋯
750	195	0.9	0.71	⋯	⋯	⋯
1730	0	−0.5	1.84	⋯	⋯	⋯
1730	195	1.3	0.63	⋯	⋯	⋯
4000	0	0.3	1.96	⋯	⋯	⋯
4000	195	3.2	0.42	⋯	⋯	⋯
		Off frequency				
750	0	−0.2	2.88	−0.4	−0.6	2.77
750	195	1.2	2.15	3.0	4.0	0.82
1730	0	0.10	2.43	−0.3	−0.7	2.07
1730	195	1.8	1.89	4.5	5.8	0.55
4000	0	0.6	1.28	0.6	0.6	1.28
4000	195	3.60	3.63	5.1	5.8	2.71

2004), overshoot can be understood as a reduction in basilar-membrane compression during the course of masker stimulation.

An important question is whether the temporal effect with off-frequency maskers can be understood in a similar way. The approach to fitting the masking functions for the off-frequency masker was similar to that described above, with a few exceptions. The first difference was related to the masker level. In order to describe the masker level in terms of the overall level of the noise in the auditory filter centered at the signal frequency, it was necessary to take into account the attenuation of the off-frequency masker provided by that filter. The amount of that attenuation was estimated in the following way. The group mean masking functions in the long-delay condition for the on- and off-frequency maskers were each fitted with a straight line. The masker levels needed to mask signal levels of 40, 50, 60, 70, and 80 dB SPL were estimated from those fits. The average difference in level between the two maskers defined the attenuation provided by the auditory filter. [Because those masking functions were not parallel (see the individual masking functions in Figs. 1 and 2) it was necessary to average over a range of levels.] The estimated attenuation values were 28.7, 25.8, and 33.7 dB for the signal frequencies of 750, 1730, and 4000 Hz. Thus, for a given masker level, the input levels (L_{in}) for the off-frequency maskers were reduced relative to those for the on-frequency maskers by these attenuation values. The second difference between fitting the on- and off-frequency functions was related to whether x was allowed to differ for the signal and masker. For the on-frequency case, x was the same for the signal and masker, as it is reasonable to assume that the two are processed via the same I–O function. For the off-frequency case, however, this may not be a reasonable assumption, as it is unclear whether the shape of the basilar-membrane I–O function would be the same for the two stimuli (e.g., Ruggero *et al.*, 1997; Cooper, 2004). Thus, the off-frequency functions were fitted twice, once when x was the same for the masker and signal (as was done for the on-frequency masker) and once when it was allowed to differ.[2] The results of these fits are shown in the bottom half of Table II. The fits were reasonably good, but generally not as good as those in the on-frequency case. As with the on-frequency fits, there are two points worth noting regarding the values of x. First, in the onset condition, the values of x were all 0.6 or less, and they did not vary dramatically across frequency. The values were, however, slightly larger at 4000 Hz than at the other frequencies. These results are different from those in the on-frequency condition, where compression was greatest (x was smallest) at the two highest frequencies. Second, regardless of whether x was forced to be the same for the masker and signal or not, the value of x was larger (the I–O function was less compressive) at the longer signal delay. This is consistent with the on-frequency results, and suggests that the temporal effect with off-frequency maskers also might be understood in terms of a decrease in compression of the underlying basilar-membrane I–O function as a function of masker stimulation.

Thus far this section has focused on the fits to the masking functions. Figure 5 shows how well the temporal effect was predicted from those fits. The symbols in each panel represent the experimental results and the lines represent the model fits. The solid line represents the case where x was the same for the signal and masker, whereas the dashed line represents the case where x could differ for the two. For the on-frequency masker, the predicted values follow the general form of the data, although they tend to overestimate the size of the temporal effect at high masker levels (due to the overestimation of threshold at high masker levels in the onset condition). For the off-frequency masker, when x was allowed to differ for the signal and masker the predicted values followed the general form of the data reasonably well, although they underestimated the temporal effect at high levels. [The underestimation results from slightly inaccurate predictions at high levels in the onset condition (750 and 1730 Hz) or in both the onset and centered conditions (4000 Hz).] When x was forced to be the same for the signal and masker, on the other hand, the prediction did not fit the data.

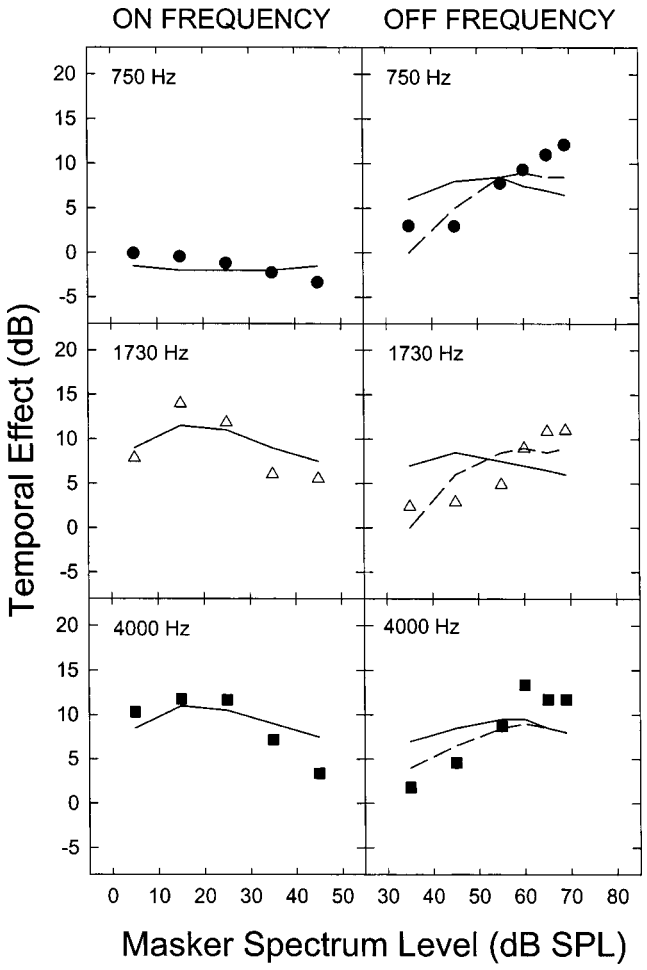


FIG. 5. The temporal effect averaged across all four subjects (symbols). The lines represent model fits to the data. Solid line: x [see Eq. (1)] was forced to be the same for the signal and masker; dashed line: x was allowed to differ for the signal and masker. See Table II for the best-fitting x values.

IV. GENERAL DISCUSSION

The primary purpose of the present study was to determine whether reported differences in the effects of signal frequency and masker level on the temporal effect for broadband on-frequency maskers versus narrow-band off-frequency maskers were due to differences in the level of the signal at masked threshold. The results clearly indicate that this was not the case. The differences that can be seen in the literature were also observed here despite similar masked thresholds for the two different types of masker.

A secondary purpose was to determine whether the temporal effect with both maskers might be understood in terms of I–O functions that become less compressive during the course of masker stimulation. The results indicate that they can. The remainder of this section discusses why the effects of signal frequency and masker level can differ despite the possibility that the two temporal effects may share a common mechanism, and then concludes with a brief discussion of the individual differences that are so common in these kinds of experiments.

A. Effects of signal frequency

One possible explanation for the lack of a temporal effect with broadband noise maskers (commonly referred to as overshoot) at the lowest signal frequency (750 Hz) is that there is less compression at the apical region of the cochlea. This is certainly consistent with some physiological (Cooper and Yates, 1994; Cooper and Rhode, 1995) and psychophysical (Lee and Bacon, 1998; Hicks and Bacon, 1999; Plack and Oxenham, 2000) data indicating that the low-frequency region is more linear than the high-frequency region. However, more recent psychophysical data (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003) suggest that the low-frequency region is just as compressive as the high-frequency region. If there is indeed as much compression at 750 Hz as there is at the higher frequencies, why is there typically little or no overshoot at 750 Hz? Two possible explanations are considered below. They are not mutually exclusive.

The first explanation is related to where the signal and masker are located on the I–O function. A decrease in compression over time will result in a temporal effect only if the signal and masker are affected differently, as a consequence of being on sufficiently different parts of the I–O function (see Fig. 4). Although the threshold level of the signal at masker onset was higher than the overall level of the masker within the auditory filter centered at 750 Hz, the signal and masker were closer in level to one another at that frequency than at either of the two higher frequencies. This may have contributed to the lack of overshoot at 750 Hz. If so, then manipulations designed to elevate signal threshold further might reveal an overshoot effect at low frequencies. This could be accomplished, for example, by using an even shorter signal duration, although one must be concerned with splatter effects, especially at low frequencies where the auditory filter is narrow.

The second explanation for little or no overshoot at 750 Hz is that the degree of compression may change considerably less as a function of masker stimulation at low frequencies than it does at higher frequencies. Assuming that the change in compression is due to the activation of the efferent system (in particular, the medial olivocochlear or MOC system), this frequency effect may reflect a difference in the pattern of efferent innervation in the cochlea. This possibility is consistent with the fact that the efferent innervation and the influence of the MOC system is greater in the high frequencies than it is in the low (Liberman, 1988; Liberman and Guinan, 1998). Although this explanation may seem inconsistent with the fact that there is a temporal effect at 750 Hz in the presence of the off-frequency masker, this apparent inconsistency may be understood in the following way. For the on-frequency masker, the signal and masker are presumably operating on the same I–O function emanating from the 750-Hz region. A relatively sparse efferent innervation in that region would leave the processing of the signal and masker relatively unaffected by ongoing masker stimulation. For the off-frequency condition, on the other hand, the masker is located in a higher frequency region than the signal. Thus, it is possible that the I–O function for the masker could change more than that for the signal during the course of masker stimulation. (This is, in fact, consistent with the results of the model fits, as shown in the bottom right section of Table II.) The decrease in compression would be akin to a decrease in the gain for the masker, and to compensate for

this the signal would need to be reduced in level at the longer delay (see Strickland, 2001, for a similar argument regarding the temporal effect with notched-noise maskers).

B. Effects of masker level

As discussed in the Introduction, the effect of level on the temporal effect with the on-frequency masker can be understood in terms of the signal and masker being at different points along the same compressive I–O function (von Klitzing and Kohlrausch, 1994; Strickland, 2001; Bacon and Oxenham, 2004). To understand the quite different effect of level with the off-frequency masker, one must assume that the signal and off-frequency masker operate on separate I–O functions that differ in shape from one another (see Fig. 5). As seen in the bottom portion of Table II, although the x values (x_s,x_m) were not especially different from one another, the best-fitting I–O function for the masker was less compressive than that for the signal at the longer signal delay. It is possible that the functions would be even more different from one another if the off-frequency masker were displaced even further away from the signal in frequency.

C. Individual differences

Individual differences are common in studies of temporal effects in simultaneous masking (e.g., Zwicker, 1965a; Bacon, 1990; Bacon and Takahashi, 1992; Wright, 1995, 1997). For the most part, they are restricted to the conditions in which the signal is located at or near the onset of the masker, although in the present study differences were observed at both delays for the off-frequency masker. The source of these individual differences is unclear. Bacon *et al.* (2002) suggested that they may reflect the influence of some form of central processing. Wright and her colleagues (Wright *et al.*, 1993; Wright, 1996), on the other hand, have suggested that the size of the temporal effect (essentially the threshold at onset, given that there is little individual variability in the long-delay conditions) may be negatively correlated with two-tone suppression or the sharpness of frequency tuning, two phenomena that presumably depend heavily upon peripheral processing and, more importantly, the integrity of nonlinear processing in the cochlea. Although the present study cannot shed any new light on the source of these individual differences, it is interesting to note that Wright's explanation for individual differences is seemingly inconsistent with the compression-based explanation for the temporal effect, in that one would expect the size of the temporal effect to be *positively* correlated with the amount of compression and hence the integrity of nonlinear cochlear processing. It would be interesting to determine whether such a correlation does indeed exist between the temporal effect and various psychophysical estimates of basilar-membrane compression.

V. FINAL COMMENT

Although the results from the present study suggest that the temporal effect with both on-frequency broadband maskers and off-frequency narrow-band maskers can be understood in terms of a decrease in compression during the course of masker stimulation, this does not preclude the possibility that other factors are involved. Further research will be necessary to determine to what extent both peripheral and central factors play a role.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute of Deafness and Other Communication Disorders (NIDCD Grant No. DC01376). This manuscript benefited from the valuable suggestions provided on earlier versions by Dennis McFadden and an anonymous reviewer.

[1]von Klitzing and Kohlrausch (1994) described how the basilar-membrane I–O function could account for the substantial increase in overshoot as the masker level is increased from a relatively low to a moderate level. Although they did not discuss the decline in overshoot at even higher masker levels, this same I–O function can account for this effect as well (see Strickland, 2001; Bacon and Oxenham, 2004).

[2]Another approach would have been to use the results from the on-frequency analyses to restrict the value of x for the off-frequency fits, under the assumption that the shape of the I–O function in response to the signal should be the same for both masker types. In other words, when x was the same for the signal and masker, it could have been restricted to the best-fitting value obtained when fitting the on-frequency masking functions. And, when x was allowed to differ for the signal and masker, the x value for the signal could have been restricted to that value. This was not done because the basilar-membrane I–O function for a tone whose frequency is equal to the characteristic frequency of the measurement site can be affected by the presence of additional sounds (e.g., Ruggero *et al.*, 1992). This means that the shape of the I–O function for the signal may not be the same in the presence of the off-frequency masker as it is in the presence of the on-frequency masker.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for Audiometers'' (American National Standards Institute, New York).

Bacon, S. P. (**1990**). ''Effect of masker level on overshoot,'' J. Acoust. Soc. Am. **88**, 698–702.

Bacon, S. P., and Healy, E. W. (**2000**). ''Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones,'' J. Acoust. Soc. Am. **107**, 1589–1597.

Bacon, S. P., and Liu, L. (**2000**). ''Effects of ipsilateral and contralateral precursors on overshoot,'' J. Acoust. Soc. Am. **108**, 1811–1818.

Bacon, S. P., and Moore, B. C. J. (**1986**). ''Temporal effects in simultaneous pure-tone masking: Effects of signal frequency, masker/signal frequency ratio, and masker level,'' Hear. Res. **23**, 257–266.

Bacon, S. P., and Oxenham, A. J. (**2004**). ''Psychophysical manifestations of compression: Hearing-impaired listeners,'' in *Compression: From Cochlea to Cochlear Implant*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 107–152.

Bacon, S. P., Repovsch-Duffey, J. L., and Liu, L. (**2002**). ''Effects of signal delay on auditory filter shapes derived from psychophysical tuning curves and notched-noise data obtained in simultaneous masking,'' J. Acoust. Soc. Am. **112**, 227–237.

Bacon, S. P., and Smith, M. A. (**1991**). ''Spectral, intensive, and temporal factors influencing overshoot,'' Q. J. Exp. Psychol. **43A**(3), 373–399.

Bacon, S. P., and Takahashi, G. A. (**1992**). ''Overshoot in normal-hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **91**, 2865–2871.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Simultaneous masking by gated and continuous sinusoidal maskers,'' J. Acoust. Soc. Am. **78**, 1220–1230.

Cooper, N. P. (**2004**). ''Compression in the peripheral auditory system,'' in *Compression: From Cochlea to Cochlear Implant*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 18–61.

Cooper, N. P., and Rhode, W. S. (**1995**). ''Nonlinear mechanics at the apex of the guinea-pig cochlea,'' Hear. Res. **82**, 225–243.

Cooper, N. P., and Yates, G. K. (**1994**). ''Nonlinear input–output functions derived from the response of guinea-pig cochlear nerve fibers: Variations with characteristic frequency,'' Hear. Res. **78**, 221–234.

Egan, J. P., and Hake, H. W. (**1950**). ''On the masking pattern of a simple auditory stimulus,'' J. Acoust. Soc. Am. **22**, 622–630.

Elliott, L. L. (**1965**). ''Changes in the simultaneous masked threshold of brief tones,'' J. Acoust. Soc. Am. **38**, 738–746.
Fastl, H. (**1976**). ''Temporal masking effects. I. Broad band noise masker,'' Acustica **35**, 287–302.
Fastl, H. (**1977**). ''Temporal masking effects. II. Critical band noise masker,'' Acustica **36**, 317–331.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Glasberg, B. R., and Moore, B. C. J. (**1994**). ''Growth-of-masking functions for several types of maskers,'' J. Acoust. Soc. Am. **96**, 134–144.
Green, D. M. (**1969**). ''Masking with continuous and pulsed sinusoids,'' J. Acoust. Soc. Am. **46**, 939–946.
Hawkins, Jr., J. E., and Stevens, S. S. (**1950**). ''The masking of pure tones and of speech by white noise,'' J. Acoust. Soc. Am. **22**, 6–13.
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.
Lee, J., and Bacon, S. P. (**1998**). ''Psychophysical suppression as a function of signal frequency: Noise and tonal maskers,'' J. Acoust. Soc. Am. **104**, 1013–1022.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Liberman, M. C. (**1988**). ''Response properties of cochlear efferent neurons: Monaural vs binaural stimulation and the effects of noise,'' J. Neurophysiol. **60**, 1779–1798.
Liberman, M. C., and Guinan, Jr., J. J., (**1998**). ''Feedback control of the auditory periphery: Antimasking effects of middle ear muscles vs olivocochlear efferents,'' J. Commun. Disord. **31**, 471–483.
Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). ''Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing,'' J. Acoust. Soc. Am. **113**, 951–960.
McFadden, D. (**1988**). ''Absence of overshoot in a dichotic masking condition,'' J. Acoust. Soc. Am. **83**, 1685–1687.
McFadden, D. (**1989**). ''Spectral differences in the ability of temporal gaps to reset the mechanisms underlying overshoot,'' J. Acoust. Soc. Am. **85**, 254–261.
Mott, J. B., and Feth, L. L. (**1986**). ''Effects of the temporal properties of a masker upon simultaneous masking patterns,'' in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 381–386.
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.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Plack, C. J., and Oxenham, A. J. (**2000**). ''Basilar-membrane nonlinearity estimated by pulsation threshold,'' J. Acoust. Soc. Am. **107**, 501–507.
Plack, C. J., and Drga, V. (**2003**). ''Psychophysical evidence for auditory compression at low characteristic frequencies,'' J. Acoust. Soc. Am. **113**, 1574–1586.
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.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Ruggero, M. A., Robles, L., Rich, N. C., and Recio, A. (**1992**). ''Basilar membrane responses to two-tone and broadband stimuli,'' Philos. Trans. R. Soc. London, Ser. B **336**, 307–315; also in R. P. Carlyon, C. J. Darwin, and I. J. Russell, editors, *Processing of Complex Sounds by the Auditory System* (Oxford University Press, Oxford).
Schmidt, S., and Zwicker, E. (**1991**). ''The effect of masker spectral asymmetry on overshoot in simultaneous masking,'' J. Acoust. Soc. Am. **89**, 1324–1330.
Smith, R. L. (**1979**). ''Adaptation, saturation, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Smith, R. L., and Zwislocki, J. J. (**1975**). ''Short-term adaptation and incremental responses of single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.
Strickland, E. A. (**2001**). ''The relationship between frequency selectivity and overshoot,'' J. Acoust. Soc. Am. **109**, 2062–2073.
Turner, C. W., and Doherty, K. A. (**1997**). ''Temporal masking and the 'active process' in normal and hearing-impaired listeners,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ), pp. 387–396.
von Klitzing, R., and Kohlrausch, A. (**1994**). ''Effect of masker level on overshoot in running- and frozen-noise maskers,'' J. Acoust. Soc. Am. **95**, 2192–2201.
Wegel, R. L., and Lane, C. E. (**1924**). ''The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear,'' Phys. Rev. **23**, 266–285.
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.
Wright, B. A. (**1995**). ''Detectability of simultaneously masked signals as a function of signal bandwidth for different signal delays,'' J. Acoust. Soc. Am. **98**, 2493–2503.
Wright, B. A. (**1996**). ''Correlated individual differences in conditions used to measure psychophysical suppression and signal enhancement,'' J. Acoust. Soc. Am. **100**, 3295–3303.
Wright, B. A. (**1997**). ''Detectability of simultaneously masked signals as a function of masker bandwidth and configuration for different signal delays,'' J. Acoust. Soc. Am. **101**, 420–429.
Wright, B. A., McFadden, D., and Champlin, C. A. (**1993**). ''Adaptation of suppression as an explanation of enhancement effects,'' J. Acoust. Soc. Am. **94**, 72–82.
Zwicker, E. (**1965a**). ''Temporal effects in simultaneous masking by white-noise bursts,'' J. Acoust. Soc. Am. **37**, 653–663.
Zwicker, E. (**1965b**). ''Temporal effects in simultaneous masking and loudness,'' J. Acoust. Soc. Am. **38**, 132–141.

Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss

Christopher J. Plack[a] and Vit Drga
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, England

Enrique A. Lopez-Poveda
Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Avda. Alfonso X "El Sabio" s/n, 37007 Salamanca, Spain

(Received 3 September 2003; revised 5 January 2004; accepted 17 January 2004)

Psychophysical estimates of cochlear function suggest that normal-hearing listeners exhibit a compressive basilar-membrane (BM) response. Listeners with moderate to severe sensorineural hearing loss may exhibit a linearized BM response along with reduced gain, suggesting the loss of an active cochlear mechanism. This study investigated how the BM response changes with increasing hearing loss by comparing psychophysical measures of BM compression and gain for normal-hearing listeners with those for listeners who have mild to moderate sensorineural hearing loss. Data were collected from 16 normal-hearing listeners and 12 ears from 9 hearing-impaired listeners. The forward masker level required to mask a fixed low-level, 4000-Hz signal was measured as a function of the masker–signal interval using a masker frequency of either 2200 or 4000 Hz. These plots are known as temporal masking curves (TMCs). BM response functions derived from the TMCs showed a systematic reduction in gain with degree of hearing loss. Contrary to current thinking, however, no clear relationship was found between maximum compression and absolute threshold. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1675812]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Sr [NFV] Pages: 1684–1695

I. INTRODUCTION

Cochlear hearing loss is associated with an increase in absolute threshold, an abnormally rapid growth in loudness with level, and a loss of frequency selectivity (see Moore, 1995 for a review). These characteristics may result from dysfunction of the outer hair cells (OHCs) in the organ of Corti. The OHCs are thought to be involved in an ''active'' mechanism that effectively applies gain to stimulation at frequencies close to the characteristic frequency (CF) of each place on the basilar membrane (BM) (see Yates, 1995 for a review). The gain is greatest at low stimulation levels, and decreases with increasing level. This frequency- and level-dependent gain sharpens BM tuning at low to moderate levels, and also results in a highly *compressive* BM response to mid- and possibly high-level tones close to CF (Robles *et al.*, 1986; Ruggero *et al.*, 1997). Measurements of BM vibration in nonhuman mammals have confirmed that interfering with the function of the OHCs, for example by furosemide injection (Ruggero and Rich, 1991), results in a steeper, more linear response to a tone at CF.

Psychophysical techniques based on forward masking have been used to estimate the growth of response of the human BM. Forward masking is used to avoid simultaneous interactions on the BM (e.g., suppression) that complicate the interpretation of the results (Oxenham and Plack, 1997). Most of these techniques have involved comparisons of the effects of maskers at and below the signal frequency. Since the BM response to a masker well below CF is linear, the off-frequency masking function can be used as a linear reference to derive the BM response to a tone at CF. Oxenham and Plack (1997) measured the forward masker level required to mask a brief signal as a function of the level of the signal (referred to here as the growth of masking, or GOM, technique). When the masker was an octave below the signal frequency of 2000 or 6000 Hz, a given change in signal level required a much smaller change in masker level for the signal to remain at threshold. This is thought to be because the response to the signal is compressive and the response to the masker is linear. Indeed, the shallow off-frequency masking function (masker level plotted against signal level) is an estimate of the BM response function to a tone at CF.

A different technique was developed by Nelson *et al.* (2001). The signal was fixed at a level just above absolute threshold. The masker level required to mask the signal was measured as a function of the masker–signal interval to produce a *temporal masking curve* (TMC). For an *off*-frequency masker, the TMC is assumed to reflect the decay with time of the internal effect of the masker: As the masker–signal interval is increased, the masker level has to increase to compensate for the decay. For an *on*-frequency masker, the TMC reflects the decay of masking, *and* the compression applied to the masker: If the response is compressive, a larger change in physical masker level will be required to produce a given change in BM excitation. It follows that an on-frequency TMC that is steep compared to the off-frequency TMC is indicative of compression. It is also possible to derive response functions from TMC data. It is assumed that, for a given masker–signal interval, the BM excitation level at the signal place in response to the masker is a constant at threshold, regardless of the masker frequency. For a given masker–signal interval, the level of the off-frequency masker re-

[a]Electronic mail cplack@essex.ac.uk

 0001-4966/2004/115(4)/1684/12/$20.00

quired is an estimate of the BM excitation required at the signal place (give or take an additive constant on a dB scale). Therefore, a plot of the on-frequency masker level (input level) against the off-frequency masker level (output level) is an estimate of the BM response function for the on-frequency masker.

The results from the GOM and TMC studies for normal-hearing listeners at high frequencies are broadly consistent with the rate of growth of BM velocity at high CFs, as measured in other mammals (Lopez-Poveda *et al.*, 2003; Nelson *et al.*, 2001; Oxenham and Plack, 1997; Plack and Drga, 2003). Most GOM and TMC studies report compression exponents (the slopes of the response functions on dB/dB coordinates) in the range 0.15–0.3. This corresponds to compression ratios (the inverses of the compression exponents) of between 6.7:1 and 3.3:1. Furthermore, the shapes of the estimated response functions, with linear low-level regions and compressive midlevel regions, are also consistent with the physiology, suggesting that both behavioral techniques measure cochlear processes. However, there are two good reasons for favoring the TMC technique. As signal level is increased in the GOM technique, the peak of the traveling wave produced by a high-frequency signal will shift basally on the BM (McFadden, 1986). This means that the GOM technique is probably not measuring the response of a single place on the BM, but rather the growth of the peak of the traveling wave with level. In addition, as signal level is increased excitation will spread to higher CFs. To prevent listeners using information from the high-frequency side of the excitation pattern (where the response growth is much more linear than at the peak), a high-pass noise needs to be added to the stimulus (Oxenham and Plack, 1997). Nelson *et al.* (2001) demonstrated that GOM curves in the absence of a high-pass noise exhibit about half the compression of GOM curves in the presence of the noise. This finding is consistent with the greater compression exponents measured in GOM studies that did not include high-pass noise (Hicks and Bacon, 1999b; Moore *et al.*, 1999; Plack and Oxenham, 2000). Selection of the appropriate noise level is problematic, especially for impaired listeners. The TMC technique avoids both these complications. In the TMC technique the signal is fixed at a low level, and hence presumably causes excitation above detection threshold over a fixed, relatively small, region of the BM. The region of the BM measured does not change with masker level, and since the spread of excitation is limited there is no need for a high-pass noise.

Both the GOM and TMC techniques have been used to estimate the BM response for listeners with cochlear hearing loss. The results suggest that a hearing loss of greater than about 50 dB is associated with an almost linear BM response (Nelson *et al.*, 2001; Oxenham and Plack, 1997): The slopes of the GOM functions and TMCs do not vary with masker frequency in these cases (providing support for the contention that the psychophysical techniques measure cochlear processes). For ears with less severe losses, the results are mixed. Moore *et al.* (1999) used the ratio of the slopes of off- and on-frequency GOM functions as an estimate of compression. They found that the compression exponent only began to increase markedly as hearing loss increased above 35 dB. However, the results may have been compromised by the fact that Moore *et al.* did not use high-pass noise to mask spread of excitation (see above). The compression exponents estimated by Moore *et al.* for normal-hearing listeners were at least twice as great as those from GOM studies that included the high-pass noise (Nelson *et al.*, 2001; Oxenham and Plack, 1997). Hicks and Bacon (1999a), again using the GOM technique without high-pass noise, found that mild temporary hearing loss induced by aspirin was associated with a change in slope, consistent with a reduction in compression. Two listeners with mild permanent sensorineural hearing loss showed similar effects. In a recent study measuring GOM for *simultaneous* notched-noise maskers, Baker and Rosen (2002) reported a reduction in gain and compression for a listener with a hearing loss of only 20 dB. However, compression estimates were generally quite low in this study, possibly because of suppressive interactions between the masker and the signal.

In the present study, the TMC technique was used to estimate the BM response to a tone at CF for listeners with a range of impairments, from no impairment to mild to moderate. The aim was to determine how the shape of the response function changes with severity of hearing loss, and to test the hypothesis that mild hearing loss is associated with a reduction in compression.

II. METHOD

A. Listeners

Sixteen normal-hearing listeners and nine listeners with mild to moderate hearing impairment participated in the study. Normal-hearing listeners (ten females and six males, aged 19–37 years old) were mostly students from the University of Essex. All had normal audiogram thresholds (within 15 dB ANSI, 1996) in octave steps from 250–8000 Hz.

Hearing-impaired listeners (five females and four males), were aged 54–68 years old, except for listener RD, who was 42 years old. Hearing-impaired listeners reported the onset of hearing difficulties between 2 and 15 years ago and had mild-to-moderate amounts of hearing loss. This was most likely sensorineural hearing loss since it came on gradually and was unrelated to any acute trauma or known disease. It was most likely age related, except for listeners PJ, SG, and RD, who reported repeated exposure to noisy environments when younger. RD also had a family history of hearing loss. On average, audiogram levels for the hearing-impaired group were higher than laboratory norms for normal hearing by 20, 30, and 38 dB at 1000, 2000, and 4000 Hz, respectively. Except for ED, the thresholds at lower frequencies were normal or near normal, suggesting that the impairments did not have a substantial conductive component. Although bone-conduction tests were not performed, the close spacing of the on- and off-frequency TMCs for the impaired listeners is also inconsistent with a conductive loss (see Sec. IV A). RD had borderline normal hearing at 4000 Hz, but elevated thresholds (35–42-dB loss) at 6000 and 8000 Hz. His absolute threshold for the brief 4000-Hz signal in the experiment was 5 dB above the highest absolute

TABLE I. Absolute thresholds, stimulus parameters, and estimated BM response parameters for normal-hearing (upper) and hearing-impaired (lower) listeners, ordered according to the absolute threshold for the signal. Listeners were tested in their right ears unless indicated otherwise. Gain estimates are only included when the low-level portion of the response function is defined by at least two points. Compression and gain could not be sensibly estimated for RG and ES due to the variability of their data. Values marked with asterisks are from response functions generated by interpolation of the off-frequency TMCs.

Listener	Signal absolute threshold (dB SPL)	Audiogram threshold at 4000 Hz (dB SPL)	Noise spectrum level (dB)	Gain (dB)	Compression exponent
EK	3.5	−8.8	−16.5	42.2	0.18
VD	4.9	−6.2	−15.1	52.8	0.13
RG	5.2	−5.0	−14.8	⋯	⋯
PP	6.0	−5.0	−14.0	37.2	0.12
RS	6.2	−5.7	−13.8	48.9	0.38
IG	6.4	−8.2	−13.6	46.0	0.00
RB	7.0	−3.5	−13.0	48.1	0.19
TP	9.2	2.2	−10.8	46.5	0.29
AC	9.4	−4.3	−10.6	39.8	0.23
CO	9.8	−5.7	−10.2	⋯	0.30
ES	9.9	2.2	−10.1	⋯	⋯
NB	10.5	2.2	−9.5	40.5	0.18
CG	10.8	−0.7	−9.2	⋯	0.22
JS	14.3	3.7	−5.7	46.2	0.14
IY	15.1	8.2	−4.9	35.3	0.37
CN	17.1	5.0	−2.9	38.0	0.11
RD	22.7	10.0	2.7	34.2	0.29
JC	25.9	24.7	−4.1	⋯	0.18/0.26*
ED(l)	36.2	28.2	−3.8	23.3	0.15
SG	40.3	37.0	0.3	⋯	0.25/0.29*
DJ(r)	41.7	36.2	1.7	27.8*	0.21*
RC(r)	45.1	37.0	−4.9	13.9/13.9*	0.28/0.28*
ED(r)	47.4	40.0	−2.6	17.4	0.37
MB	52.9	44.3	2.9	17.3*	0.14*
RC(l)	54.4	48.3	4.4	6.5*	0.02*
PJ	56.5	51.0	6.5	18.5*	0.32*
DJ(l)	59.3	49.2	−0.7	13.0*	0.46*
BH	60.8	52.3	10.8	10.2*	0.10/0.33*

threshold for the normal-hearing group. Audiogram thresholds, and absolute thresholds for the signal used in the experiment, are given in Table I.

All of the listeners were naive except for EK, IY, PP, and VD. The normal-hearing listeners, except for AC, CN, ES, and RB, had 4–8 h practice in pilot studies for the current experiment. AC, CN, ES, and RB, and all of the hearing-impaired listeners received 1–2 h practice in blocks used to determine parameters for them in the main experiment. There were no systematic improvements in thresholds in the experimental sessions. Listeners were paid £5 per hour for their participation.

B. Stimuli and equipment

The experiment involved forward masking of tonal signals by tonal maskers. The signal had a frequency of $f_s = 4000$ Hz and an absolute duration of 8 ms (4-ms raised-cosine ramps, 0-ms steady state). The masker had a frequency of either $f_m = 2200$ or 4000 Hz and an absolute duration of 104 ms (2-ms raised-cosine ramps, 100-ms steady state). Silent masker–signal intervals (masker envelope offset to signal envelope onset) ranged from 0–100 ms in steps of 5 or 10 ms, with the set of intervals used dependent on each listener's performance. The signal level was set to 10 dB SL, which ranged from 15–27-dB SPL for normal-hearing listeners and from 33–71-dB SPL for hearing-impaired listeners. Masker level was varied trial by trial.

A low-level notched noise was gated on and off with the masker. This was intended as a temporal cue to help reduce possible confusion effects (Neff, 1986) and not as a source of masking. The noise was white except for a notch at the signal frequency (filter cutoffs at $0.883 f_s$ and $1.117 f_s$, with 90-dB/oct filter slope). For normal-hearing listeners and listener RD, the spectrum level in the passband was set 30 dB below the signal level (i.e., 20 dB below signal absolute threshold). For the remaining hearing-impaired listeners the spectrum level was set to either 40, 50, or 60 dB below the signal level so that it fell in the range −5 to 11 dB. For most hearing-impaired listeners, setting the spectrum level to 30 dB below the signal meant having spectrum levels almost up to 40 dB, resulting in masking due to the noise. Setting the spectrum level to a level between −5 and 11 dB was practical in that hearing-impaired listeners could make use of temporal information at lower frequencies without the noise contributing to masking at 4000 Hz. As described above, most of the hearing-impaired listeners had elevated thresholds at 4000 Hz but normal or near-normal absolute thresholds for frequencies below 2000 Hz. Notched-noise levels for each individual are given in Table I.

The experiment was run using custom-made software on a PC workstation located outside a double-walled sound-

attenuating booth. Stimuli were digitally generated and were produced using an RME Digi96/8 PAD 24-bit soundcard set at a clocking rate of 48 000 Hz. The soundcard includes an antialiasing filter. The headphone output of the soundcard was fed via a patch panel in the sound-booth wall, without filtering or amplification, to Sennheiser HD 580 circumaural headphones. Each listener sat in the booth and decisions were recorded via a computer keyboard. Listeners viewed a computer monitor through a window in the sound booth. Lights on the monitor display flashed on and off concurrently with each stimulus presentation and provided feedback at the end of each trial.

C. Procedure

All stimuli were presented monaurally. Normal-hearing listeners were tested in their right ear. Hearing-impaired listeners were tested in their right ear, or in both left and right ears if their audiogram thresholds at 4000 Hz differed across ears by more than 10 dB. Those tested in both ears wore an earplug in their contralateral ear to prevent the possibility of airborne sound aiding performance.

The experiment used a two-interval, forced-choice adaptive tracking procedure with the interstimulus interval set to 500 ms. The signal level was fixed at 10 dB SL and the masker level was varied adaptively using a two-up, one-down rule to obtain the masker level needed to achieve 70.7 percent correct (Levitt, 1971). The masker frequency and masker–signal interval were fixed in any given block of trials. For normal-hearing listeners, the initial masker level was set to 0 dB SPL. The step size of the adaptive track was 8 dB for the first four turnpoints and 2 dB for 12 subsequent turnpoints. Data for listeners RD and BH were collected using these settings, but we found it was desirable to slightly modify the procedure for the rest of the hearing-impaired listeners, due to limitations in the equipment's maximum output (102-dB SPL rms). The modifications were that the initial masker level was set to 20–40 dB below estimated threshold, and the step size was 4 dB for the first four turnpoints and 2 dB for the 12 subsequent turnpoints. For all listeners, the mean of the last 12 turnpoints was taken as the threshold estimate for each block of trials. If the standard deviation of the turnpoints was greater than 6 dB the estimate was discarded and the block was later repeated. Data were also discarded and repeated if possible for any blocks in which the masker clipped more than twice at levels above 102 dB SPL.

Listeners ran blocks of trials lasting 2–4 min per block and spent 15–60 min in the sound booth at any one time, taking breaks as needed. A replication consisted of a complete run of 10–20 blocks per listener, depending on the range of the masker–signal interval at each f_m, which was determined during each listener's practice trials. The order of blocks was randomized across masker–signal interval and f_m until all blocks in a replication had been completed. Unless otherwise indicated, the mean threshold across four replications was taken as the threshold estimate for each combination of masker–signal interval and f_m. Absolute thresholds for the signal were measured using the same basic procedure, except that signal level was varied using a two-down, one-up adaptive rule.

III. RESULTS AND ANALYSES

A. Temporal masking curves

Individual TMCs are presented in Fig. 1 for normal-hearing listeners and in Fig. 2 for hearing-impaired listeners. In general, on-frequency TMCs (triangles) are steeper, in part or in whole, than the accompanying off-frequency TMCs. Assuming the BM response to off-frequency maskers is linear at the signal place, then steeper portions of the on-frequency TMC indicate a compressive response to the on-frequency masker. Such results were found for both the normal-hearing and the hearing-impaired groups.

For three impaired ears, RC(l), MB, and PJ, it was not possible to reliably measure on-frequency thresholds at the longest masker–signal interval. In two successive runs, they *consistently* detected the signal when the 4000-Hz masker was at 100 dB SPL. The equipment clipped at the next higher level in the adaptive track (102 dB SPL), so the data point for the longest masker–signal interval presented in Fig. 2 for these three listeners was set at 100 dB SPL, and in each case this value was used (when required) in the analyses described below. Limiting the level in this way resulted in *underestimates* of their thresholds and consequently, based on linear extrapolation of their off-frequency TMCs, underestimates of the amount of compression (overestimates of the compression exponents). For several impaired ears, the off-frequency masker levels at the longer intervals also clipped at some stage during the adaptive track on every replication, and these measurements were aborted [see MB, PJ, DJ(l), and BH in Fig. 2]. However, unlike the on-frequency measurements for the three ears described above, the signal was only *occasionally* detected with the masker at the clipping threshold (the adaptive track touched the clipping point before retreating). It may be assumed that the ''true'' off-frequency masker level for these missing points lies somewhat below the clipping threshold.

A surprising aspect of the data is that the slopes of the *off*-frequency TMCs appear to be different for the normal-hearing and hearing-impaired listeners. Two analyses were conducted to illustrate and quantify this difference. First, the slopes of the straight lines connecting consecutive points on the off-frequency TMCs were calculated for each listener. The slope values for all the listeners were then combined and ordered by masker–signal interval or by masker level. The upper panels of Fig. 3 show running averages of these values, plotted against running averages of masker–signal interval (upper left) and masker level (upper right). The running averages were calculated separately for the normal-hearing and hearing-impaired groups. The graphs indicate that there are some trends in the data. First, the off-frequency TMC slopes for the normal-hearing group show a tendency to decrease with increasing masker–signal interval. The correlation between masker–signal interval and slope is significant for the normal-hearing group ($r=-0.231$, $n=106$, $p=0.017$). On the other hand, there is little variation in slope with masker–signal interval for the hearing-impaired group,

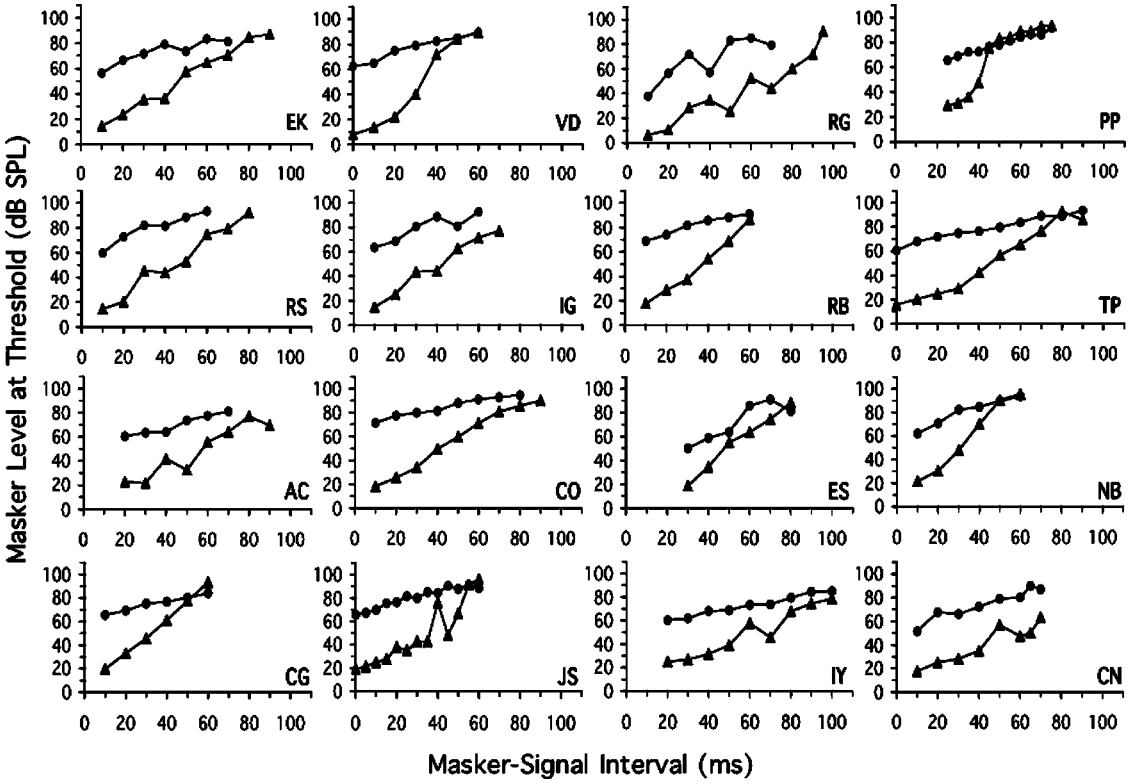


FIG. 1. TMCs for normal-hearing listeners, showing mean masker level at threshold as a function of masker–signal interval for on-frequency (4000-Hz) maskers (triangles) and off-frequency (2200-Hz) maskers (circles).

and the correlation is not significant ($r=-0.155$, $n=48$, $p=0.293$). Second, the off-frequency TMC slopes for the normal-hearing group show a tendency to decrease with increasing masker level, and the correlation between level and slope is significant ($r=-0.271$, $n=106$, $p=0.005$). The correlation between level and slope is not significant for the hearing-impaired group ($r=-0.159$, $n=48$, $p=0.282$), although the range of levels for this group is much less than for the normal-hearing group.

A second analysis investigated the relationship between absolute threshold and off-frequency TMC slope at a specific masker–signal interval and at a specific masker level. Second-order polynomials were fit to the off-frequency TMCs for each listener, and the slopes of the functions calculated (analytically) at a masker–signal interval of 30 ms and, separately, at a masker level of 85 dB SPL. A slope value was only included when the masker–signal interval (either 30 ms, or the calculated masker–signal interval for an 85-dB SPL masker) fell within the range of intervals tested for that listener. The calculated slopes are shown in the lower

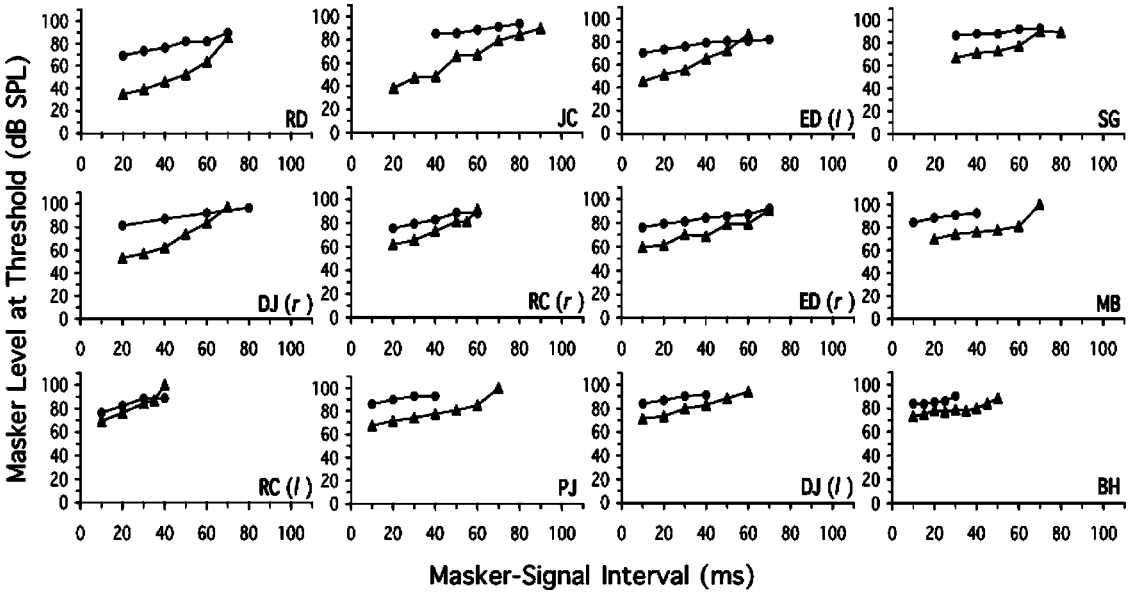


FIG. 2. As Fig. 1, except showing TMCs for hearing-impaired listeners.

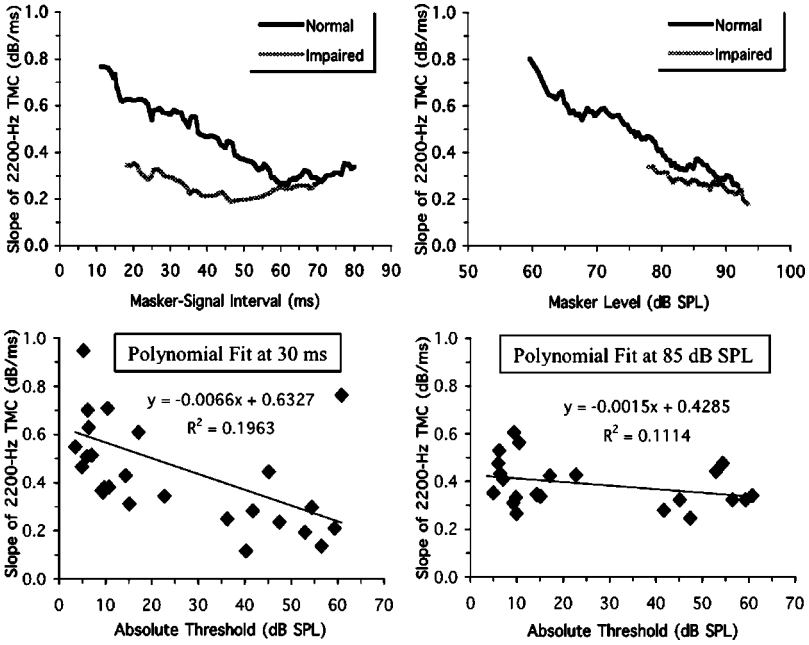


FIG. 3. The upper panels show the slopes of the 2200-Hz TMCs, collapsed across listeners and plotted as running averages over 20 consecutive points. The slopes are shown as a function of masker–signal interval (upper left) and masker level (upper right). The lower panels show scatterplots of 2200-Hz TMC slope against signal absolute threshold. The slopes were derived by fitting second-order polynomials to the TMC data for each listener. The slopes at a masker–signal interval of 30 ms (lower left) and at a masker level of 85 dB SPL (lower right) were calculated from the fitted functions. The lines show linear regression fits, with equations and R^2 values displayed on the figure.

two panels of Fig. 3, plotted against the absolute thresholds for the signal. There is a significant negative correlation between absolute threshold and slope at a 30-ms masker–signal interval ($r=-0.443$, $n=27$, $p=0.021$), but no significant correlation between absolute threshold and slope at an 85-dB SPL masker level ($r=-0.334$, $n=22$, $p=0.129$). In summary, the results show that at short masker–signal intervals the off-frequency TMC slopes are shallower for the hearing-impaired listeners than for the normal-hearing listeners. However, the difference may be related to the fact that the masker thresholds were at a higher level for the hearing-impaired listeners. There appears to be little difference in slope between the normal-hearing and hearing-impaired groups when the TMCs are matched for masker level.

It is conceivable that the results for the shorter masker–signal intervals were influenced by the notched noise that was presented as a cue to the offset of the masker. The noise level was generally higher relative to the signal level for the normal-hearing group compared to the hearing-impaired group (see Table I). Although in all conditions the signal was clearly above threshold in the presence of the noise alone, if the noise had contributed to the masking of the signal at the short masker–signal intervals, then the masker level at threshold may have been artificially lowered for the normal-hearing group at the short intervals (both on and off frequency) leading to an increase in TMC slope. One way to test this hypothesis is to examine the masker levels at threshold for the on-frequency masker. It is assumed that, for a given masker–signal interval at threshold, the ratio of BM velocity in response to the signal to BM velocity in response to the masker is constant. Furthermore, the BM response to a low-level on-frequency masker should have been affected in a similar way as the response to the 10-dB SL signal by any attenuation (or loss of gain) resulting from the hearing impairment: If the hearing impairment resulted in the response to the signal being attenuated by x dB, then the response to the masker should also have been attenuated by x dB. It follows that the difference between the *physical* levels of the signal and the masker at threshold should have been unaffected by the hearing loss, *if* there was no additional source of masking for the normal-hearing listeners.

At a masker–signal interval of 10 ms, the mean difference between signal level and on-frequency masker level is −1.7 dB for the normal-hearing group and 2.0 dB for the hearing-impaired group. At an interval of 20 ms, the values are 6.1 and 5.5 dB, respectively. So, although there is a suggestion that the noise may have contributed to masking at the 10-ms gap, there appears to have been no effect at a 20-ms gap, for which there is a clear difference in off-frequency TMC slope between the two groups.

B. Response functions

Following the approach of Nelson *et al.* (2001), TMCs for each listener were converted into BM response functions by plotting the off-frequency masker threshold against on-frequency masker threshold, paired according to masker–signal interval. For several hearing-impaired ears, off-frequency masker levels were not available at all the masker–signal intervals for which on-frequency masker levels were measured. For these ears, it was assumed that the slopes of the off-frequency TMCs do not change significantly with masker–signal interval. Given that there was no significant correlation between masker–signal interval and slope for the hearing-impaired listeners (see Sec. III A), this was felt to be a reasonable assumption. For those masker–signal intervals missing a measured off-frequency masker level, an off-frequency masker level was generated by interpolation, using a linear fit to the off-frequency data.

The response functions for normal-hearing listeners and for hearing-impaired listeners are plotted in Figs. 4 and 5, respectively. These show the growth of level with masker–signal interval for the off-frequency masker relative to the growth for the on-frequency masker. Open symbols indicate those points that were generated by interpolating the off-frequency TMCs. The positive diagonal ($y=x$) is included

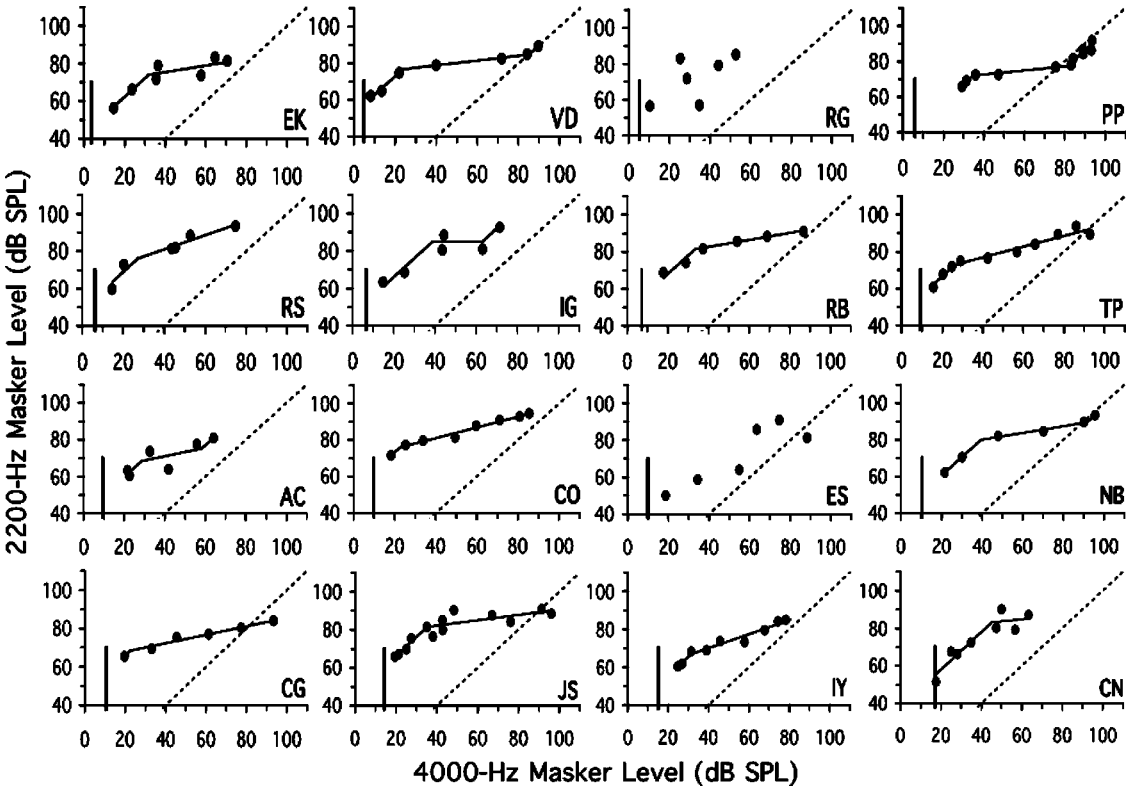


FIG. 4. Estimated response functions for normal-hearing listeners, showing the 2200-Hz masker level at threshold plotted against the 4000-Hz masker level at threshold, paired according to masker–signal interval. The positive diagonal (straight line) indicates the expected response of a passive, linear system. The vertical line indicates signal absolute threshold (*re*: the *x* axis). The kinked line shows the piecewise linear fit (see the text for details).

in Figs. 4 and 5 to indicate the expected response of a linear system with 0-dB gain (see below). Masker–signal intervals over which the on-frequency TMC is parallel to the off-frequency TMC translate to portions of the response functions with slope unity, implying a linear BM response. Masker–signal intervals over which the on-frequency TMC is steeper than the off-frequency TMC translate to portions of the response functions with slope less than unity (compression).

Most of the normal-hearing listeners show evidence of compression, having shallow portions in their response functions in Fig. 4, although this is not so clear for listeners RG and ES, whose data are variable. Most of the hearing-impaired listeners also show evidence of compression, as

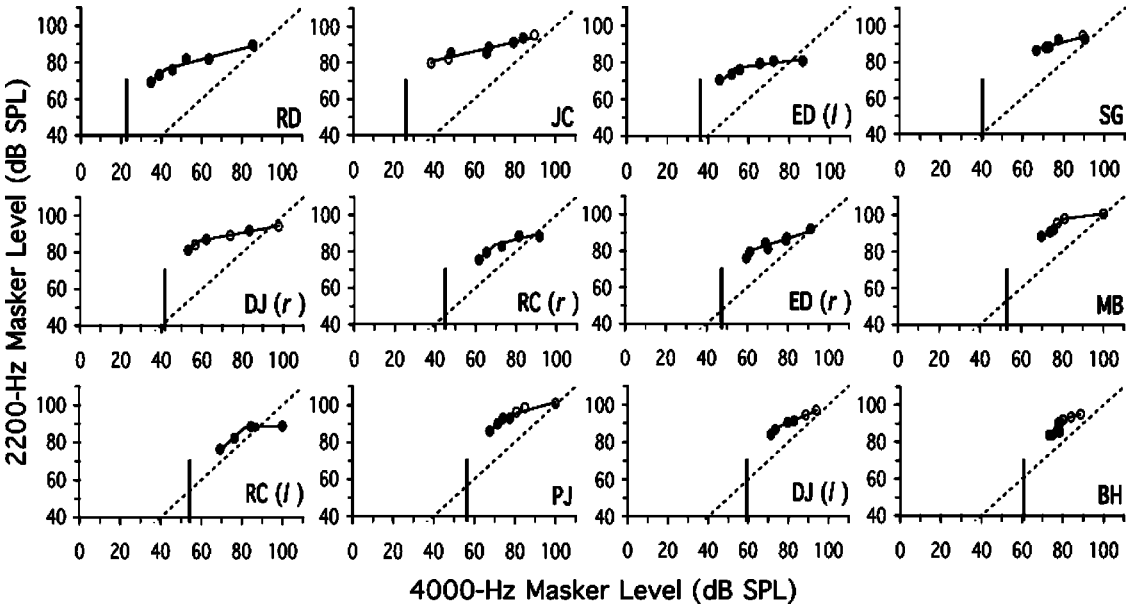


FIG. 5. As Fig. 4, except for hearing-impaired listeners. Open circles show points on the response functions that were estimated by interpolation of the off-frequency data, using straight-line fits to the off-frequency TMCs. Left and right ears are indicated by (*l*) and (*r*), respectively.

seen in the shallow portions of their response functions in Fig. 5. Indeed, many of the hearing-impaired response functions appear at least as compressive as those seen for normal-hearing listeners, although the compression extends over a smaller range of levels for the former. This suggests that BM compression is present even when sensitivity is much reduced compared to normal-hearing listeners. Some normal-hearing listeners (VD,PP,IG,NB) show evidence of a return to linearity at the highest masker levels, with on-frequency TMCs becoming shallow at long masker-signal intervals, and hence the response functions becoming steeper at high levels. This is not the case in general for hearing-impaired listeners. If such a pattern is to be found with hearing-impaired listeners, it must occur at very high levels (>100 dB SPL).

C. Estimates of gain and compression

The maximum gain of the active mechanism can be estimated using the horizontal distance between the left-most, linear portion of a response function and the positive diagonal. This is an estimate of the difference between the levels of a 4000-Hz masker and a 2200-Hz masker required to produce the same BM response. If it is assumed that no gain is applied to the off-frequency masker (so that the positive diagonal approximates the off-frequency response even at low input levels), then this estimate is equivalent to the difference in the BM response between a low-level masker at 4000 Hz and a low-level masker at 2200 Hz. This measure of gain is similar to that used by others, e.g., Ruggero *et al.* (1997), in that it estimates the difference between the active BM response and the maximum passive response at signal place. The data of Ruggero *et al.* suggest that the off-frequency response to a tone at 0.55×CF is slightly less (by 2–4 dB) than that of the maximum passive response. Since the passive response has broad tuning, it is assumed that any discrepancy between the off-frequency response and the maximum passive response is approximately constant across individuals. If so, this should not affect any correlation of gain with absolute threshold in our analyses below.

To help quantify gain and compression, a three-section fit was applied to each listener's BM response function (Lopez-Poveda *et al.*, 2003; Yasin and Plack, 2003). The function comprised a linear low-level region [Eq. (1)], a compressive mid-level region [Eq. (2)], and a linear high-level region [Eq. (3)]. The three sections were joined by two breakpoints, a lower breakpoint (BP_1) joining sections 1 and 2, and an upper breakpoint (BP_2) joining sections 2 and 3. The equations for the three sections are given by

$$L_{\text{out}} = L_{\text{in}} + G \quad (L_{\text{in}} \leqslant \text{BP}_1), \tag{1}$$

$$L_{\text{out}} = cL_{\text{in}} + k_1 + G \quad (\text{BP}_1 < L_{\text{in}} \leqslant \text{BP}_2), \tag{2}$$

$$L_{\text{out}} = L_{\text{in}} + k_2 + G \quad (L_{\text{in}} > \text{BP}_2), \tag{3}$$

where G is the gain (dB), c is the slope of the compressive region (dB/dB) or the compression exponent, $k_1 = \text{BP}_1(1-c)$, $k_2 = \text{BP}_2(c-1) + k_1$. L_{in} (input level, level of 4000-Hz masker) and L_{out} (level of BM response, level of 2200-Hz masker) are both expressed in dB. The function was fit to the data using the *fminsearch* function in MATLAB to satisfy a least-squares regression criterion. The slopes of the lower and upper sections were fixed at unity (linear response), while the slope of the middle section (c) and the breakpoints (BP_1 and BP_2) were varied by the fitting procedure. The only constraint on the fitting procedure was that the compression exponent c was not allowed to be negative; otherwise, the parameters were allowed to vary freely. In some cases either the lower breakpoint or the upper breakpoint estimated by the fitting procedure was beyond the range of the data, so that effectively the data were fit using a reduced number of parameters. The fits were first made to the response function data without any off-frequency interpolation (just the filled symbols in Figs. 4 and 5), and then separately including the interpolated data (i.e., filled *and* open symbols). Fits were not included when there were less than five points on the response function [i.e., for DJ(r), MB, RC(l), PJ, and DJ(l), without interpolated data]. The lines fit to the response functions including the interpolated data are shown in Figs. 4 and 5. The fits for listeners RG and ES had high rms errors (9 and 7 dB, respectively) and are not included in the figures or in subsequent analyses. For the other listeners, however, the fits generally provide a good description of the data (rms errors less than 4 dB).

Scatterplots of maximum gain against signal absolute threshold and compression against signal absolute threshold are shown in Fig. 6. Gain and compression estimates were taken from the three-section fits. A gain estimate was included only if there were at least two points on the response function that were below BP_1. The results are shown separately for fits to the response-function data without interpolation (left-hand panels) and for fits to the data including the interpolated off-frequency values (right-hand panels). The fits that involved interpolated values are shown as open symbols. Data from both normal-hearing and hearing-impaired listeners are presented together in each graph. A straight-line fit to the scatterplot data, and the squared correlation coefficient, are also displayed. Individual gain and compression estimates are given in Table I.

As shown by Fig. 6, the estimated gain of the cochlear amplifier decreases systematically with increasing hearing loss. The correlation is statistically significant both for the data without interpolation ($r=-0.930$, $n=16$, $p<0.0005$; slope$=-0.736$ dB/dB) and for the data with interpolation ($r=-0.951$, $n=22$, $p<0.0005$; slope$=-0.647$ dB/dB). Figure 6 also shows that, although there is a range of compression values for both normal-hearing listeners (left half of the data) and hearing-impaired listeners (right half of the data), there is little correlation between compression and absolute threshold. The correlations are not significant either for the data without interpolation ($r=0.101$, $n=21$, $p=0.662$) or for the data with interpolation ($r=0.284$, $n=26$, $p=0.159$). For some of the more impaired ears the compressive region is defined by only a couple of points (often interpolated). However, the ears with milder impairments [RD, JC, ED(l), SG, DJ(r), RC(r), and ED(r)] have well-defined shallow sections in their response functions and compression exponents within the normal range (0.29, 0.26,

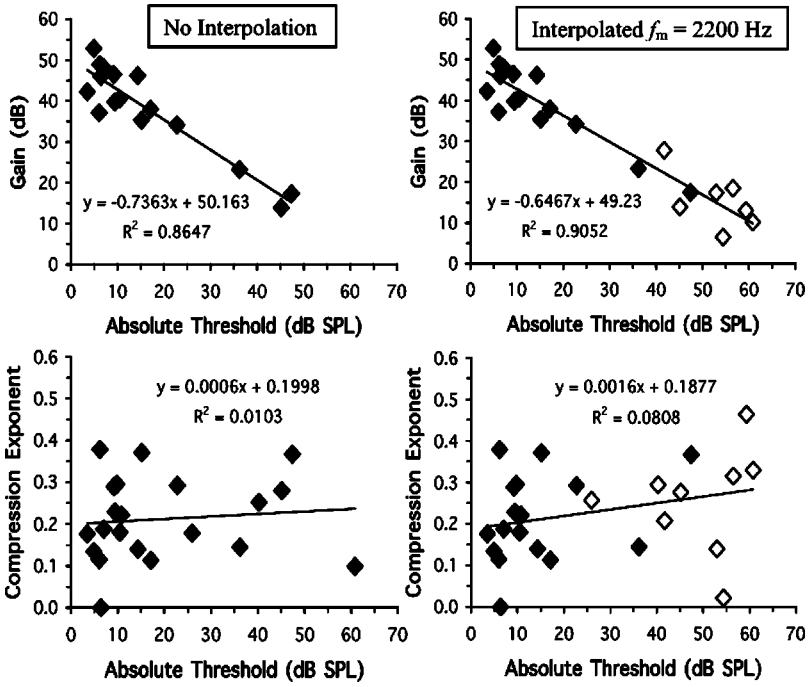


FIG. 6. Scatterplots of maximum gain against signal absolute threshold (upper panels), and response-function slope against signal absolute threshold (lower panels), at 4000 Hz. The left-hand panels show the results for response functions generated without interpolation. The right-hand panels include the results for response functions that were generated using interpolated off-frequency masker levels (indicated by the open symbols). Gain estimates are only included when the low-level portion of the response function is defined by at least two points (see Table I).

0.15, 0.29, 0.21, 0.28, and 0.37, respectively, for the interpolated data). The mean compression exponent for the interpolated data is 0.20 for the normal-hearing listeners and 0.26 for the hearing-impaired listeners.

There is an issue regarding whether the three-section straight-line fits provide an accurate characterization of the response functions. Third-order polynomials have been used previously to fit response functions (Nelson *et al.*, 2001; Plack and Drga, 2003). Such fits were attempted on the present data but gave very inconsistent results, with negative slopes in some cases. It was felt that the three-section fits, although not ideal, do capture the main features of the response functions for the normal-hearing and hearing-impaired listeners: The reduction in gain with hearing loss (also reflected in the vertical spacing between the on- and off-frequency TMCs), and the preservation of a shallow slope in the response function (compression), even for ears showing the largest hearing loss.

IV. DISCUSSION

A. Consequences of inner and outer hair cell dysfunction

An elevation in absolute threshold may result from a dysfunction of the inner hair cells (IHCs) or of the OHCs (see Moore, 1995). IHC dysfunction reduces the efficiency of transduction of BM vibration (hence reducing sensitivity) but is not thought to affect the mechanical properties of the BM itself (Liberman *et al.*, 1986). OHC dysfunction affects the response of the BM, but not the transduction process *per se*. Now, imagine a situation in which a cochlea has damage to the IHCs only. For a given input level, the response properties of the BM should be *identical* to that for a healthy ear. According to the measures described earlier, the maximum gain of the active mechanism, the input level at the first breakpoint in the response function, and the compression exponent, should all be unaffected by the hearing loss, and hence independent of the degree of hearing loss. For an ear with purely OHC dysfunction, however, the maximum gain should be strongly related to the absolute threshold. Specifically, a plot of gain (in dB) against absolute threshold (in dB) should be a straight line with a slope of -1. Turning down the gain by 10 dB should result in an increase in absolute threshold by 10 dB.

If the hearing loss experienced by our listeners were simply a matter of non-frequency-specific attenuation *prior* to the BM, as might result from conductive hearing loss, then both on- and off-frequency TMCs would increase by the same amount and, consequently, the estimated maximum gain would remain unchanged as a function of hearing loss. Similarly, pure IHC loss (post-BM attenuation) should result in no change in the estimated maximum gain (although it is only possible to measure the lower breakpoint on the response function if it is above absolute threshold). These predictions do not seem to describe the present data. The strong relation between maximum gain and absolute threshold (Fig. 6) suggests that the hearing loss in the impaired listeners tested here was mainly the result of OHC dysfunction. Off-frequency TMCs were somewhat higher in level for the hearing-impaired compared to the normal-hearing group, but it was the general increase in level for the on-frequency TMCs relative to the off-frequency TMCs that characterized the impaired group, again consistent with OHC damage. Including the interpolated data, the plot of gain against absolute threshold has a slope of -0.65, which suggests that most of the threshold elevation can be attributed to a reduction in gain. The fact that the slope was not -1 suggests that there may have been some IHC dysfunction among the listeners (contributing to perhaps 35% of the threshold elevation for those ears for which the gain was measurable). Indeed, from the results of other studies examining the relative proportions of IHC and OHC hearing loss, it would be surprising if there

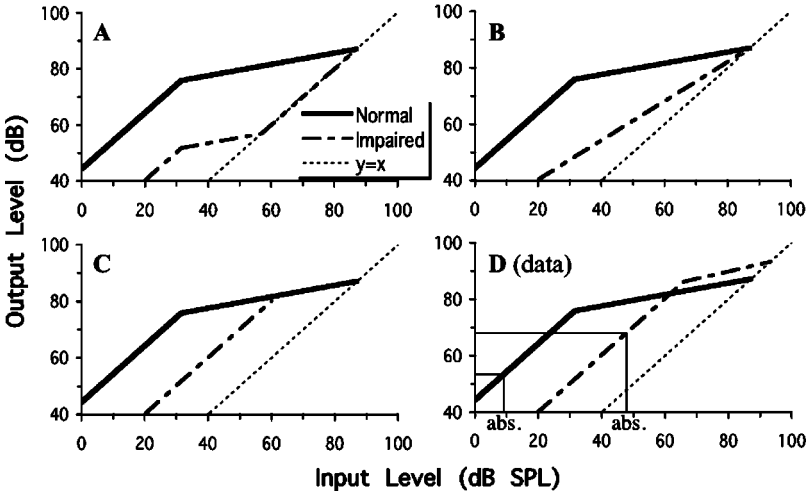


FIG. 7. BM response functions for normal-hearing and hearing-impaired listeners. The normal response function is an average function generated from the present data (see the text for details). In panels (A) to (C) the response functions for the hearing-impaired listeners are hypothetical. In panel (A), the gain for hearing-impaired listeners is reduced equally at all input levels up to the passive response. In panel (B), the reduction in gain is greatest at low input levels, with a diminishing reduction as level is increased. In panel (C), the gain is reduced at low input levels, but unaffected at high input levels. Panel (D) shows average response functions, for both normal-hearing and hearing-impaired listeners, generated from the present data (see the text for details). In panel (D), the thin vertical lines show the average absolute thresholds for the signal, and the thin horizontal lines the associated response levels, for the normal-hearing and hearing-impaired listeners used to generate the response functions.

were not some evidence of IHC dysfunction (Lopez-Poveda *et al.*, 2004; Moore and Glasberg, 1997; Moore *et al.*, 1999). IHC dysfunction will raise the output level on the BM response function that corresponds to absolute threshold, and may explain why the low-level linear segment on the response function is generally shorter for the hearing-impaired listeners than for the normal-hearing listeners (see Figs. 4 and 5, and Sec. IV B). However, given that OHC dysfunction was probably the main cause of the reduction in sensitivity, what do the present results tell us about the effects of OHC dysfunction on the response of the BM?

B. Effects of outer hair cell dysfunction on the BM response

To aid the discussion of the effects of OHC dysfunction on the shape of the BM response function, three different hypothetical scenarios are illustrated in Fig. 7. Each panel of the figure shows the normal response function (continuous black line, generated from the present data as described below), the passive response function with no active mechanism (thin dotted line), and a hypothetical response for a listener with mild hearing loss (alternate dashes and dots). In the upper left panel (A) the gain is reduced equally at all input levels up to the passive response. In the upper right panel (B) the reduction in gain is greatest at low input levels, with a diminishing reduction as level is increased. In the lower-left panel (C) the gain is reduced at low input levels, but unaffected at high input levels. The lower-right panel (D) shows average response functions generated from the present data. These functions were obtained by averaging the x and y values of the lower breakpoints and the compression exponents derived from the fitting procedure, across the normal-hearing listeners and across the hearing-impaired listeners. Upper breakpoints were omitted because they often could not be specified (the values from the fitting procedure were above the highest points on the response functions), and (as for the gain estimates) the results for ears with less than two points on the response function below the lower breakpoint were also omitted. For those ears that remained, it seems clear that the data are best summarized by option C. Although there are individual differences in the response functions, overall it appears that mild cochlear hearing loss is associated with a reduction in the gain at the *lower input levels only*, and not across the whole range of input levels that are affected by the active mechanism. This is why the slope of the compressive part of the response function did not vary significantly with absolute threshold (see Fig. 6).

In this respect, the present data do not appear to be consistent with some physiological models of hearing loss. The BM response function of a chinchilla injected 40 min previously with furosemide showed a reduction in gain at all levels (Ruggero and Rich, 1991), more similar to option (A) in Fig. 7. As with the present data, however, BM compression was relatively unaffected by this mild hearing loss. Murugasu and Russell (1995) report guinea pig displacement measurements during salicylate perfusion. Some of their response functions show a reduction in gain at all levels, but some show an effect only at low levels, similar to the functions reported here. Recent auditory-nerve recordings from cats with noise-induced hearing loss also seem consistent with the present data (Heinz *et al.*, in press). For mild hearing loss, a measure of the total auditory-nerve activity showed a reduction in response at low levels but not at high levels, consistent with a reduction in gain at low levels only.

The thin vertical and horizontal lines in panel (D) of Fig. 7 show the average absolute threshold for the signal, and the associated response level, for those normal-hearing and hearing-impaired listeners used to generate the average response functions. Notice that the BM response level at threshold is higher for the hearing-impaired group. This may be interpreted as a reduction in sensitivity resulting from IHC dysfunction. These lines also illustrate the point made in Sec. IV A that the linear segment of the response function measurable in the experiment (threshold to first breakpoint) is shorter for the hearing-impaired listeners.

C. Off-frequency temporal masking curves

A surprising incidental finding of the experiment was that the slopes of the off-frequency TMCs were *shallower* for hearing-impaired listeners than for normal-hearing listeners when compared at short masker–signal intervals. Recently, Rosengard *et al.* (2003) have also reported shallow off-frequency TMCs in hearing-impaired listeners compared to normals. The analysis described in Sec. III A suggests that

the difference may be related to the higher off-frequency masker levels for the hearing-impaired listeners. When the off-frequency TMC slopes were compared at the same masker level (or at long masker–signal intervals) there was little difference between the groups.

According to the interpretation of TMCs outlined in the Introduction, the shape of the off-frequency TMC should depend only on the internal decay of forward masking, a measure of temporal resolution. There is little evidence to suggest that hearing-impaired listeners in general have a deficit in temporal resolution, at least as measured by tasks such as gap detection with sinusoidal markers (Moore and Glasberg, 1988) and modulation detection (Bacon and Gleitman, 1992; Moore *et al.*, 1992). However, it is the case that the hearing-impaired listeners in the present study were older and received less training than the normal-hearing listeners, and this may have influenced temporal processing. If the effects of hearing impairment on the off-frequency TMCs were *not* a consequence of a general temporal resolution deficit, two possibilities remain. The first is that the aspect of temporal resolution measured by forward masking is unconnected with the aspect (or aspects) measured in other tasks, and that hearing-impaired listeners have a specific deficit in forward masking. This might be possible if forward masking is a consequence of adaptation at the IHC/auditory nerve synapse (Furukawa and Matsuura, 1978; but see Oxenham, 2001; Smith, 1979), and that IHC dysfunction affects this in some way. The second possibility is that the auditory system responds nonlinearly to an off-frequency masker, either at the level of the BM or more centrally. To account for the difference between the normal and impaired ears, the off-frequency compression exponent may be invariant with level but increased in impaired ears (leading to a shallow TMC slope in impaired ears), or the compression exponent may be increased at high levels (leading to a shallow TMC slope at high levels). The finding that the TMC slopes were similar for normal and impaired ears at long masker–signal intervals favors the latter explanation.

V. CONCLUSIONS

(i) BM response functions derived from on- and off-frequency TMCs generally show a linear low-level region and a compressive midlevel region. With hearing loss the low-level region shifts to the right, reflecting the reduction in sensitivity, but there is little evidence for a change in the slope of the compressive region with losses up to 50 dB or so.

(ii) The results suggest that mild to moderate sensorineural hearing loss is associated with a reduction in the gain for low-level CF tones, but little change in the gain for higher-level tones, and consequently little change in the maximum compression.

(iii) Hearing-impaired listeners show shallower off-frequency TMCs than normal-hearing listeners when measured at short time intervals. However, there is little effect of impairment on TMC slope at long time intervals, or if the TMCs are matched for off-frequency masker level.

ACKNOWLEDGMENTS

The authors thank the Associate Editor and two anonymous reviewers for very helpful comments on an earlier draft of the manuscript, and especially for spotting that the off-frequency TMC slopes decrease at high levels. The authors also thank Andrew Oxenham for comments on an earlier draft of the manuscript, and Ray Meddis for valuable discussions regarding the interpretation of the data. The research was supported by EPSRC Grant GR/N07219. Author EALP was supported by FIS PI020343 and G03/203.

Bacon, S. P., and Gleitman, R. M. (**1992**). ''Modulation detection in subjects with relatively flat hearing losses,'' J. Speech Hear. Res. **35**, 642–653.

Baker, R. J., and Rosen, S. (**2002**). ''Auditory filter nonlinearity in mild/moderate hearing impairment,'' J. Acoust. Soc. Am. **111**, 1330–1339.

Furukawa, T., and Matsuura, S. (**1978**). ''Adaptive rundown of excitatory post-synaptic potentials at synapses between hair cells and eighth-nerve fibers in the goldfish,'' J. Physiol. (London) **276**, 193–209.

Heinz, M. G., Scepanovic, D., Sachs, M. B., and Young, E. D. (**in press**). ''Normal and impaired level encoding: Effects of noise-induced hearing loss,'' in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York).

Hicks, M. L., and Bacon, S. P. (**1999a**). ''Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking,'' J. Acoust. Soc. Am. **106**, 1436–1451.

Hicks, M. L., and Bacon, S. P. (**1999b**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Liberman, M. C., Dodds, L. W., and Learson, D. A. (**1986**). ''Structure–function correlation in noise-damaged ears: A light and electron-microscopic study,'' in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, edited by R. J. Salvi, D. Henderson, R. P. Hamernik, and V. Colletti (Plenum, New York).

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). ''Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing,'' J. Acoust. Soc. Am. **113**, 951–960.

Lopez-Poveda, E. A., Plack, C. J., Meddis, R., and Blanco, J. L. (**2004**). ''Cochlear nonlinearity between 500 and 8000 Hz in listeners with moderate cochlear hearing loss,'' Abstracts of the Twenty-Seventh Annual Midwinter Meeting of the Ass. Res. Otolaryngol., Daytona Beach, FL.

McFadden, D. (**1986**). ''The curious half octave shift: Evidence for a basalward migration of the travelling-wave envelope with increasing intensity,'' in *Basic and Applied Aspects of Noise-Induced Hearing Loss*, edited by R. J. Salvi, D. Henderson, R. P. Hamernik, and V. Colletti (Plenum, New York), pp. 295–312.

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, Oxford).

Moore, B. C. J., and Glasberg, B. R. (**1988**). ''Gap detection with sinusoids and noise in normal, impaired and electrically stimulated ears,'' J. Acoust. Soc. Am. **83**, 1093–1101.

Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.

Moore, B. C. J., Shailer, M. J., and Schooneveldt, G. P. (**1992**). ''Temporal modulation transfer functions for bandlimited noise in subjects with cochlear hearing loss,'' Br. J. Audiol. **26**, 229–237.

Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). ''Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**, 2761–2778.

Murugasu, E., and Russell, I. J. (**1995**). ''Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea,'' Aud. Neurosci. **1**, 139–150.

Neff, D. L. (**1986**). ''Confusion effects with sinusoidal and narrow-band-noise forward maskers,'' J. Acoust. Soc. Am. **79**, 1519–1529.

Nelson, D. A., 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.

Oxenham, A. J. (**2001**). ''Forward masking: Adaptation or integration?'' J. Acoust. Soc. Am. **109**, 732–741.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Plack, C. J., and Drga, V. (**2003**). ''Psychophysical evidence for auditory compression at low characteristic frequencies,'' J. Acoust. Soc. Am. **113**, 1574–1586.
Plack, C. J., and Oxenham, A. J. (**2000**). ''Basilar-membrane nonlinearity estimated by pulsation threshold,'' J. Acoust. Soc. Am. **107**, 501–507.
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 phase responses,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2003**). ''Estimates of basilar-membrane compression in listeners with normal hearing derived from growth-of-masking functions and temporal masking curves,'' Abstracts of the Twenty-Sixth Annual Midwinter Research Meeting of the Ass. Res. Otolaryngol., Daytona Beach, FL.
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.
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.
Smith, R. L. (**1979**). ''Adaptation, saturation, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Yasin, I., and Plack, C. J. (**2003**). ''The effects of a high-frequency suppressor on tuning curves and derived basilar-membrane response functions,'' J. Acoust. Soc. Am. **114**, 322–332.
Yates, G. K. (**1995**). ''Cochlear structure and function,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 41–73.

Mechanisms determining the salience of coloration in echoed sound: Influence of interaural time and level differences

Katrin Krumbholz[a)]
Institute of Medicine (IME), Research Center Jülich, D-52425 Jülich, Germany, and Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

Kate Maresh, Jennifer Tomlinson, and Roy D. Patterson
Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

Annemarie Seither-Preisler and Bernd Lütkenhöner
Department of Experimental Audiology, ENT Clinic, Münster University Hospital, Kardinal-von-Galen-Ring 10, D-48149 Münster, Germany

(Received 22 May 2003; accepted for publication 30 December 2003)

This study investigates whether the salience of the pitch associated with a single reflection of a broadband sound, such as noise, is determined by the monaural information mediated by the stimuli at the two ears, or by the relative locations of the primary sound and the reflection. Pitch strength was measured as a function of the reflection delay and the lateral displacement between the primary sound and the reflection. Thereby, lateral displacement was produced by means of interaural time differences (ITDs) in experiment 1 and interaural level differences (ILDs) in experiment 3. The results from both experiments are in accordance with the assumption that the strength of the pitch associated with a reflection is based on a central average of the internal representations of the stimuli at the two ears. This notion was corroborated by experiment 2, which showed that the results from experiment 1 could be mimicked by simply adding the stimuli from the two ears and presenting the merged stimulus identically to both ears. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1649334]

PACS numbers: 43.66.Hg, 43.66.Pn [AK] Pages: 1696–1704

I. INTRODUCTION

In reverberant environments, sounds often produce multiple reflections from surrounding surfaces. Usually, the primary sound and the reflections are perceived as a single, fused auditory event whose apparent location is dominated by the localization cues in the primary sound [for a review, see Blauert (1997) and Litovsky *et al.* (1999)]. However, a reflection can noticeably change other perceptual qualities of the composite sound, such as its loudness, its spatial extent, or its timbre. In particular, a reflection may impart a pitch, or ''coloration,'' to the sound (Bilsen and Ritsma, 1969; Zurek, 1979).[1]

To a first approximation, a reflection is a delayed and attenuated copy of the primary sound, which is added to the primary sound before reaching the listener's ears. When a copy of a random noise is delayed by d ms, multiplied with a gain factor, g, and added back to the original noise, the resulting sound elicits a two-component perception with a hiss, typical of the original noise, and a weak pitch, which approximately corresponds to the reciprocal of the delay, $1/d$ (in kHz). The salience of the pitch is maximal when the delayed noise and the original noise are equally intense ($g=1$), and decreases when *either* noise is attenuated relative to the other. In the frequency domain, the delay-and-add process creates a cosinusoidal ripple on the power spectrum of the original noise, with peaks at integer multiples of $1/d$, so the stimulus has been termed ''rippled noise'' (RN) by previous authors (Bilsen, 1966; Bilsen and Ritsma, 1969; Yost and Hill, 1978). The cosinusoidal power transfer function of the delay-and-add process is often referred to as a ''comb-filter'' function.

Informal observations by Koenig (1950) suggest that the perceptual effects caused by the reverberation of speech or incidental noises in a room are less noticeable when listening binaurally than when listening monaurally. This finding was explored further by Zurek (1979), who measured the detection threshold for a single reflection when the primary sound was a random noise and the stimuli were presented over headphones. In the first, the *diotic* condition, the primary sound and the reflection were presented identically to both ears. In the second condition, the reflection was presented *dichotically* with an interaural time difference (ITD) of 500 μs. Thus, the stimuli at the two ears were essentially RNs, with identical delays in the diotic condition, and with delays differing by 500 μs in the dichotic condition. Zurek found that, when the reflection delay was less than about 10 ms, detection threshold for the reflection was lower in the diotic than in the dichotic condition; that is, listeners found it harder to detect the reflection when the stimuli at the two ears were different. On the assumption that coloration was the detection cue for reflection delays shorter than 10 ms, Zurek's results corroborate Koenig's informal observations.

a)Electronic mail: k.krumbholz@fz-juelich.de

 0001-4966/2004/115(4)/1696/9/$20.00

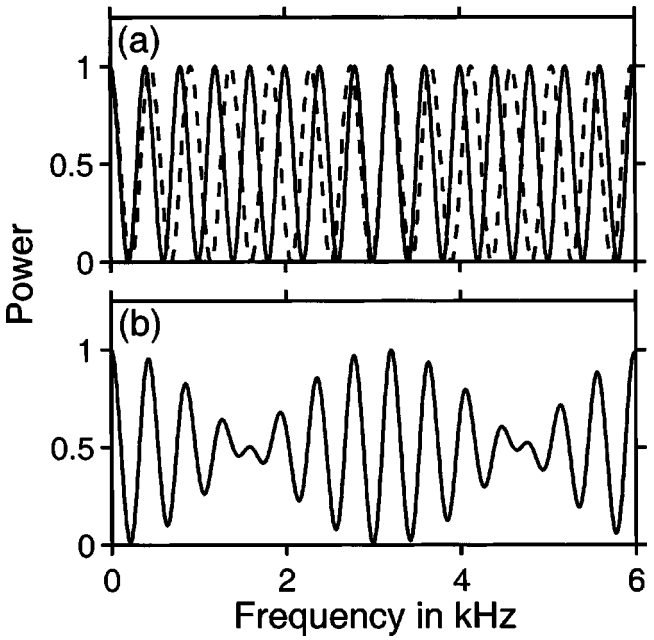


FIG. 1. Comb-filter functions of two RNs (a) with delays of 2.18 (dashed line) and 2.5 ms (solid line), and the average of the two functions (b).

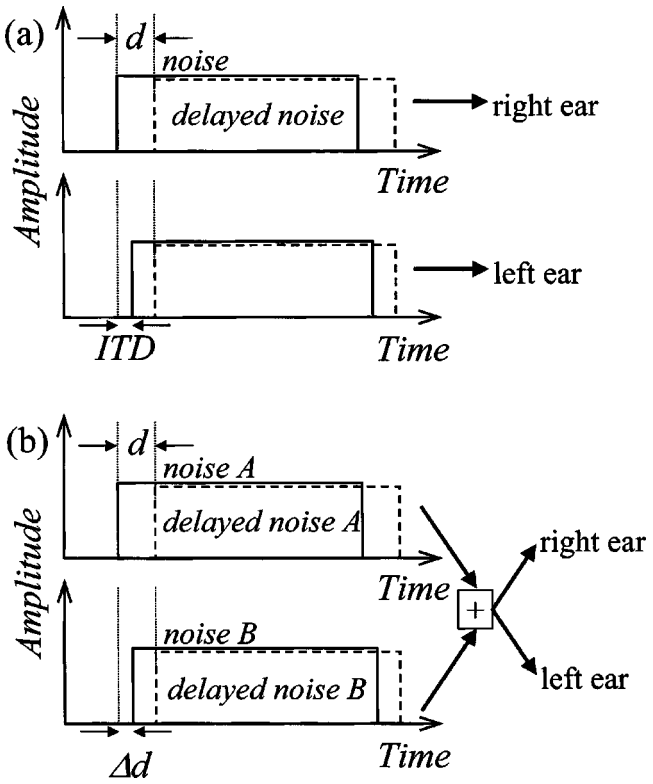


FIG. 2. Schematic representation of the stimuli used in experiments 1 and 2. In both cases, the undelayed and delayed noises are represented by solid and dashed lines, respectively; d is the nominal reflection delay. In experiment 1, the undelayed noise was presented with varying ITDs (a). In experiment 2, two RNs with varying delay differences, Δd, were added digitally and the merged stimulus presented identically to both ears (b).

Zurek proposed a qualitative explanation of his results based on the power spectra of the stimuli. He suggested that the auditory system averages the auditory spectra[2] of the stimuli from the two ears, and that the salience of the coloration of the composite stimulus is determined by the depth of the ripples in the central averaged spectrum. In the diotic case, the spectra from the two ears were identical. In the dichotic condition, on the other hand, the effective delays of the RNs at the left and right ears differed by the ITD in the reflection, so the cosine ripples in their comb-filter functions had slightly different repetition rates. The dashed and solid lines in Fig. 1(a) show the comb-filter functions of two RNs with delays differing by 320 μs, and Fig. 1(b) shows the average of the two functions. In this example, the cosine ripples in the individual functions are out of phase in the frequency regions just below 2 kHz and around 5 kHz. In these frequency regions, the ripples cancel out and the ripple depth is reduced in the averaged spectrum. Zurek assumed that the reduction in ripple depth in the averaged spectrum causes the reduction in the salience of coloration in the composite stimulus.

An alternative explanation of Zurek's (1979) findings is that the salience of coloration depends on the *relative directions* of the primary sound and the reflection. It has been suggested that the strength of the perceptual suppression of echoes (''precedence effect'') varies with the spatial separation between the primary sound and the reflections (Litovsky *et al.*, 1999; see also Shinn-Cunningham *et al.*, 1993; Krumbholz and Nobbe, 2002). In Zurek's dichotic condition, the primary sound and the reflection had different ITDs (0 and 500 μs), so, when sounded separately, they would be perceived in different lateral positions, and it may be that the salience of coloration decreases with increasing lateral displacement between primary sound and reflection. Theile (1978, 1980, cited in Blauert, 1997) proposed that ''under certain conditions, the auditory system recognizes the positions of all of the sound sources... and undertakes appropriate inverse filtering based on this information'' (Blauert, 1997, p. 327). Here, ''inverse filtering'' refers to the ripple spectrum imparted by the reflections. According to this hypothesis, which will be referred to as the ''directionality hypothesis,'' the function relating coloration salience to the spatial separation between primary sound and reflection would be expected to be similar for different reflection delays. In contrast, Zurek's central-averaging hypothesis posits that the salience of coloration is determined by the relative positions of the peaks and troughs in the auditory spectra from the two ears, which crucially depend on the reflection delay and on the ITDs of the primary sound and the reflection.

The current experiments extend Zurek's (1979) original experiment. In experiment 1, coloration salience was measured for a range of lateral separations between primary sound and reflection, mediated by different ITDs, and for different reflection delays. The results revealed that the coloration salience is not simply determined by the lateral displacement between the primary sound and the reflection. Experiment 2 investigates whether coloration salience alternatively depends on a central averaging of the stimulus representations from the two ears. Experiment 3 examines whether the central averaging hypothesis can account for conditions, in which lateral displacement between primary sound and reflection is mediated by ILDs rather than ITDs.

II. EXPERIMENT 1

Experiment 1 investigates the salience of the pitch associated with a single reflection of a random noise as a function of the lateral displacement between the primary sound and the reflection. In this experiment, lateral displacement was produced by means of ITDs. A schematic representation of the stimuli is shown in Fig. 2(a). The primary sound (solid lines) was a random noise that was presented with varying ITDs. The reflection (dashed lines) was a delayed copy of the original noise that was multiplied with a gain factor (g) and presented identically to both ears. Thus, in the current experiment, the reflection was diotic, whereas the primary sound was dichotic with an ITD. In Zurek's (1979) experiment, the situation had been reverse, in that the primary sound was diotic, whereas the reflection was dichotic with an ITD. The ordering of the sounds would not be expected to have any effect on the data from experiment 1. It was dictated by another study, in which we measured the salience of the *interaural spatial* cues mediated by the reflection. In that experiment, we did not want the measures to be confounded by variations in the laterality of the reflection. The data from that study will be presented elsewhere.

In the current experiment, the salience of the pitch produced by the composite sound was estimated by measuring the minimum gain, g, of the reflection, for which the listener could just discriminate a 12% difference in the nominal reflection delay. Bilsen and Ritsma (1970) and later Yost and Hill (1978) used a similar procedure to measure the pitch strength of simple RNs.

A. Methods

1. Stimuli

The stimuli of experiment 1 consisted of a primary sound and a single reflection. The primary sound was a burst of Gaussian noise with a total duration of 350 ms (between the 0-Volt points) and an overall level of 59 dB SPL. The noise was generated afresh on each trial and gated on and off with 2.5-ms squared-sine and -cosine ramps. The primary sound was presented with varying ITDs, ranging from 0 to 800 μs in 160-μs steps. In the following, a positive ITD or ILD means that the noise at the left ear was delayed or attenuated relative to the noise at the right ear, so the intracranial image was lateralized on the right. The range of ITDs tested roughly corresponds to that encountered in realistic environments (Wightman and Kistler, 1993). When the ITD is zero, the intracranial image is perceived approximately in the middle of the head; with an ITD of 800 μs, the intracranial image is completely lateralized towards the right ear. In order to simulate a reflection, a copy of the original noise was delayed, multiplied with a gain factor, g, and added to the primary sound. The reflection was identical at both ears. Thus, the stimuli at the two ears were effectively RNs with delays d at the right ear and d-ITD at the left ear, and with the same amplitude ratio between the original noise and the delayed noise at both ears [see Fig. 2(a)]. Four different reflection delays were tested, ranging from 2.5 to 20 ms in doublings.

Stimuli were generated digitally with 16-bit amplitude resolution and a sampling rate of 25 kHz (TDT AP2). The stimuli for the left and right ears were digital-to-analog converted through separate channels of the D/A converter (TDT DD1), antialiasing filtered at 10 kHz (TDT FT6-2) and attenuated separately using two programmable attenuators (TDT PA4). The attenuated stimuli were passed to two channels of a headphone buffer (TDT HB6) and presented through headphones (AKG K 240 DF) to the listener, who was seated in a double-walled, sound-attenuating booth. The level of the stimuli was regularly checked with a spectrum analyzer (SRS SR780).

2. Procedure

The salience of the pitch produced by the reflection in the present stimuli was estimated by measuring the ''pitch-difference detection threshold'' (PDDT). On each trial in a two-alternative, forced-choice procedure, two stimuli consisting of primary sound and reflection were presented with a silent interval of 500 ms. The reflection delays of the two stimuli differed by 12% of their average, which was the nominal reflection delay, d, of the current run. Either the first or the second stimulus had the longer delay. A 12% delay difference corresponds to a pitch difference of about two semitones. The listener was asked to choose the stimulus with the higher pitch, *viz.*, the shorter delay, by pressing one of two buttons on a response box. Feedback was given at the end of each trial. The primary sound and the reflection had equal levels at the start of each threshold run. After three consecutive correct responses, the gain, g, of the reflection was decreased, and after each incorrect response, the gain was increased, tracking the gain that yields 79% correct responses (Levitt, 1971). The step size of the gain increments and decrements was 5 dB up to the first reversal of gain, *viz.*, the first incorrect response; it was reduced to 3 dB up to the second reversal and was set to 2 dB for the rest of the eight reversals that made up each threshold run. Each threshold estimate is the mean of the gains (in dB) at the last six reversals. Each listener completed at least three threshold runs for each condition. Where the standard error of these three threshold estimates exceeded 2 dB, one or two more threshold estimates were obtained. Thus, the data points presented in the figures are the means of 3 to 5 threshold estimates and the error bars show their standard error. The order in which conditions were measured was counterbalanced between the threshold runs.

3. Listeners

Four listeners (three female, one male) participated in experiment 1 authors KK, KM, and JT, and one student, RM, who was paid for his services at an hourly rate. The listeners were between 21 and 31 years of age and had no reported hearing deficits.

B. Results

Figure 3 shows the PDDT as a function of the ITD in the primary sound; the four panels show the results for the four reflection delays tested ($d = 2.5$, 5, 10, and 20 ms). Individual thresholds are represented by different symbols, which are shifted relative to each other along the abscissa for

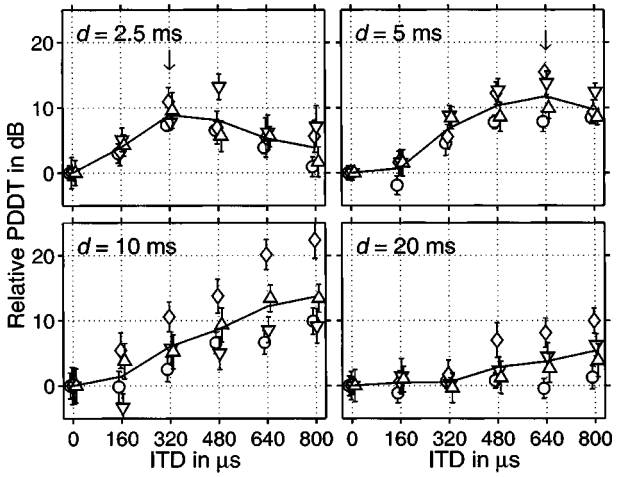


FIG. 3. Relative pitch-difference detection threshold (PDDT) for four different reflection delays, *d*, plotted as a function of the ITD in the primary sound; each panel shows the results for one reflection delay (see the legend in the upper left corner of each panel). The relative PDDT is defined as the PDDT minus the threshold value for ITD = 0 μs (diotic) and the respective reflection delay. Individual thresholds are represented by different symbols, which are shifted relative to each other along the abscissa for clarity; the average thresholds are connected by solid lines.

clarity. The average thresholds are connected by solid lines. The thresholds shown in Fig. 3 were normalized by subtracting the threshold value for ITD = 0 μs (diotic condition) at the respective reflection delay. Normalization was used to eliminate absolute threshold differences between listeners, which would otherwise have obscured common trends in the threshold functions across listeners. Figure 4 shows the absolute thresholds for ITD = 0 μs as a function of the reflection delay; as in Fig. 3, individual thresholds are denoted by different symbols and the average thresholds are connected by a solid line.

For all reflection delays tested, the diotic condition (ITD = 0 μs) yielded the lowest PDDT, indicating that the pitch associated with the reflection was most salient in this condition. The functions relating the PDDT to the ITD in the

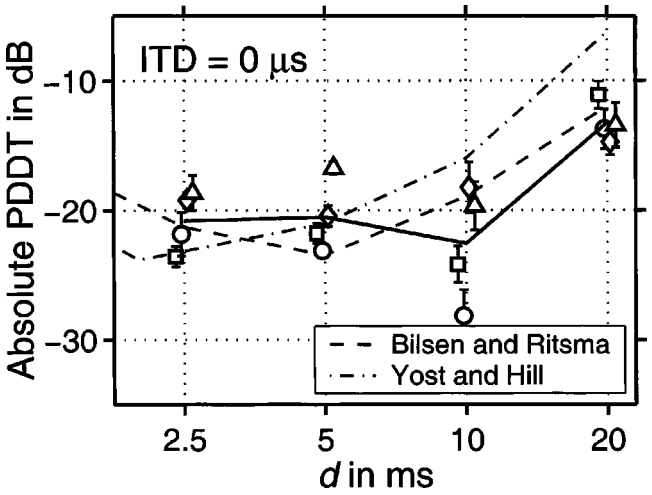


FIG. 4. Absolute PDDT for ITD = 0 μs, plotted as a function of the reflection delay, *d*, (symbols and solid line). The dashed and dash-dotted lines show the PDDT reported by Bilsen and Ritsma (1970) and Yost and Hill (1978) for a comparison.

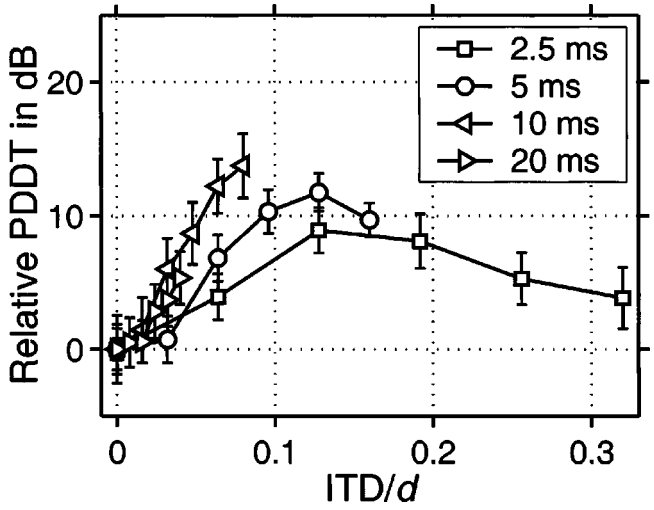


FIG. 5. Average relative PDDT replotted from Fig. 3 as a function of the normalized ITD, that is, ITD divided by the respective reflection delay, *d*. The parameter is the reflection delay (see the legend).

primary sound were nonmonotonic for reflection delays, *d*, of 2.5 and 5 ms. For $d = 2.5$ ms, the average threshold assumed a maximum when the ITD was 320 μs (see the vertical arrow in the upper left panel of Fig. 3). When *d* was doubled to 5 ms (upper right panel), the maximum in the threshold function shifted to ITD = 640 μs. The threshold functions for $d = 10$ and 20 ms (lower panels) increased monotonically with increasing ITD, but the slope of the increase was shallower for $d = 20$ ms than for $d = 10$ ms. This suggests that the threshold functions for different delays are scaled versions of one another, and that the functions for $d = 10$ and 20 ms are also nonmonotonic, but assume their maxima at ITDs outside the range measured in the current experiment. However, plotting the threshold functions against the normalized ITD, that is, ITD divided by the reflection delay, *d*, does not eliminate the differences between them (Fig. 5). Rather, the initial slope of the normalized threshold functions becomes progressively steeper as the delay is increased from 2.5 to 10 ms; the slope of the function for $d = 20$ ms is similar to that for $d = 10$ ms.

The absolute PDDT (Fig. 4) was roughly constant for *d* between 2.5 and 10 ms and then increased abruptly by about 10 dB when *d* was increased to 20 ms. For a comparison, the dashed and dash-dotted lines in Fig. 4 show the PDDT reported by Bilsen and Ritsma (1970) and Yost and Hill (1978), respectively. Despite slight differences in procedure, the present data are in reasonable agreement with these previous data. The largest deviation between the present and the previous data occurs at $d = 10$ ms, where the present PDDT is slightly smaller than those reported by the two previous studies.

III. EXPERIMENT 2

The results of experiment 1 contravene the directionality hypothesis, which predicts that pitch strength decreases with increasing lateral displacement between the primary sound and the reflection, irrespective of the reflection delay. The shape of the function relating pitch strength (as measured by

the PDDT) to the ITD difference between the primary sound and the reflection strongly depended on the reflection delay, and the function was nonmonotonic for the shorter delays.

Experiment 2 was aimed to test whether the combination of pitch-related information across the ears is accomplished by a central averaging of the monaural stimulus representations as hypothesized by Zurek (1979). In this experiment, the averaging process was performed *artificially* by physically adding two RNs and presenting the merged stimulus identically to both ears. The two RNs in the merged stimuli of experiment 2 were similar to the RNs that constituted the stimuli at the two ears in experiment 1. If the auditory system effectively averages the internal representations of the stimuli at the two ears before evaluating the pitch-related information, a merged RN should have a similar pitch strength as the corresponding dichotic stimulus.

A. Methods

1. Stimuli and procedure

In experiment 1, the stimuli at the right and left ears were effectively RNs with delays d and d-ITD [see Fig. 2(a)]. The stimuli used in experiment 2 also consisted of two RNs with differing delays, but in this case the RNs were added digitally before being presented to the listener. A schematic representation of the stimuli is shown in Fig. 2(b). The two RNs were generated using two independent primary noises with equal overall levels [noises A and B in Fig. 2(b)]. A copy of each noise was delayed by d ms, multiplied with a gain factor, g, and added to the respective primary noise. The same gain factor was used for both reflections, so the reflections also had the same levels. The overall level of the two primary noises added together was 59 dB SPL. As in experiment 1, the gain of the reflections was the adaptive parameter. The onset times of the primary noises were staggered by the delay difference, Δd, which is analogous to the ITD in experiment 1 [see Fig. 2(a)]. Thus, the two RNs had delays of d and d-Δd. They were added digitally and presented identically to both ears. The composite stimulus will be referred to as a ''merged'' RN (compare Patterson *et al.*, 2000; Handel and Patterson, 2000). It was necessary to use two independent noise samples (A and B) as primary noises, because using the same noise sample would have resulted in an unwanted ripple spectrum corresponding to the delay difference Δd. The PDDT was measured for four different reflection delays, ranging from 2.5 to 20 ms in doublings, and for varying delay differences, Δd, ranging from 0 to 800 μs in 160-μs steps. Apart from the differences described above, stimulus generation, setup, and procedure were the same as in experiment 1.

2. Listeners

Four listeners (three female, one male) took part in experiment 2, three of whom had already taken part in Experiment 1 (KM, JT, and KK). The fourth listener, RT, was a student who was paid for his services at an hourly rate. The listeners were between 21 and 31 years of age and had no reported history of hearing impairment.

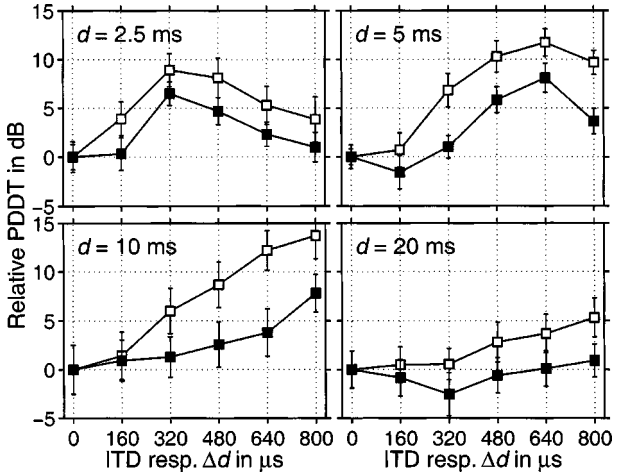


FIG. 6. Relative PDDT from experiment 2 (filled symbols), averaged across four listeners and plotted as a function of the delay difference, Δd. The four panels show the results for the four reflection delays tested (see the legend in the upper left corner of each panel). For a comparison, the open symbols show the average threshold from experiment 1, replotted from Fig. 3.

B. Results

The filled symbols in Fig. 6 show the average PDDT from experiment 2, plotted as a function of the delay difference, Δd, of the merged RNs. The four panels show the results for the four reflection delays tested ($d=2.5$, 5, 10, and 20 ms). As in Fig. 3, the PDDT was normalized by subtracting the threshold value for $\Delta d=0$ μs and the respective reflection delay. For a comparison, the open symbols show the average PDDT from experiment 1, plotted as a function of the ITD in the primary sound as in Fig. 3. As the delay difference, Δd, in experiment 2 is analogous to, and had the same values as the ITD in experiment 1, both data sets could be plotted on the same abscissa. The threshold functions from the two experiments were strikingly similar. As in experiment 1, the threshold functions from experiment 2 were nonmonotonic for reflection delays of 2.5 and 5 ms, with maxima at $\Delta d=320$ and 640 μs, respectively; the threshold functions for $d=10$ and 20 ms increased monotonically, and the slope of the increase was shallower for $d=20$ ms than for $d=10$ ms. However, while the *shape* of threshold functions from the two experiments was similar, the thresholds from experiment 2 varied over a considerably smaller range than those from experiment 1.

IV. EXPERIMENT 3

The results of experiments 1 and 2 indicate that coloration salience is determined by a central averaging of the internal stimulus representations from the two ears (Zurek, 1979) rather than by the lateral separation between the primary sound and the reflection. Experiment 3 was aimed to test whether the central averaging hypothesis can also account for conditions in which the lateral separation between primary sound and reflection is mediated by ILDs rather than by ITDs.

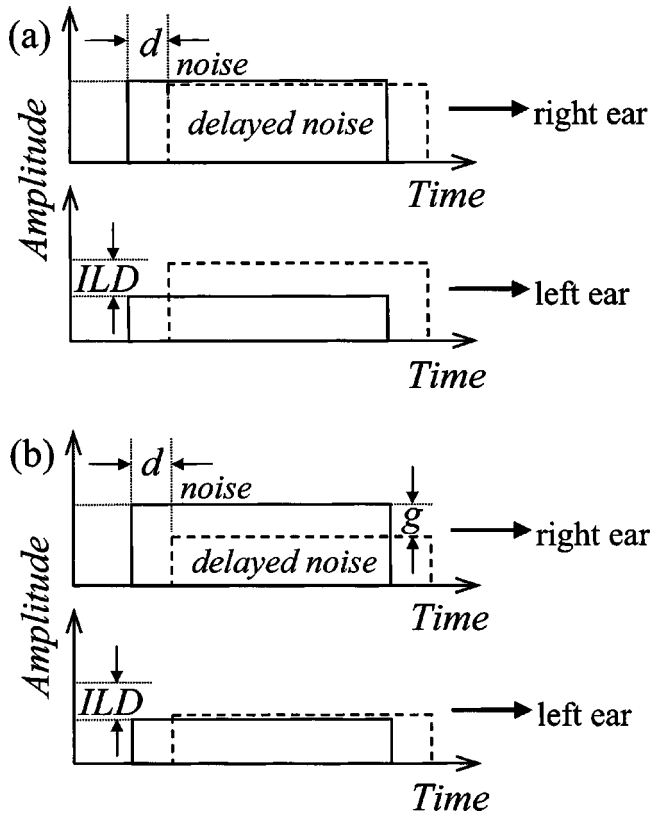


FIG. 7. Schematic representation of the stimuli used in experiment 3 for two different values of the reflection gain, g (a and b). As in experiment 1, the reflection (dashed lines) was presented identically to both ears. The primary noise (solid lines) was lateralized by means of varying ILDs. Panel (a) shows the case where the reflection gain, g, is close to unity; panel (b) shows the situation when g (in dB) has converged to a value close to -ILD.

A. Methods

The stimuli were similar to those used in experiment 1, except that the primary sound was presented with varying ILDs, rather than ITDs. The ILDs ranged from 0 to 30 dB in 6-dB steps. Thereby, the undelayed noise at the left ear was attenuated by the respective ILD, whereas the undelayed noise at the right ear was left unchanged. A schematic representation of the stimuli is shown in Fig. 7(a). As in experiment 1, the delayed noise (dashed lines) was identical at both ears and its amplitude was equal to the amplitude of the undelayed noise (solid lines) at the right (unattenuated) ear times the gain factor, g. Thus, in experiment 3, the stimuli at the two ears were RNs with the same delay, d, at both ears, but with different amplitude ratios between the original noise and the delayed noise, depending on the ILD and the gain, g. In the example shown in Fig. 7(a), the gain is roughly equal to unity ($g=1$); Fig. 7(b) shows the case where the gain (in dB) approximately corresponds to the ILD in the primary sound. In experiment 3, only a single reflection delay of 5 ms was tested.

The listeners, setup, and procedure were the same as in experiment 1. At the beginning of the threshold runs, the reflection started with the level of the primary sound at the ear, at which the primary sound was stronger (that is, the right ear).

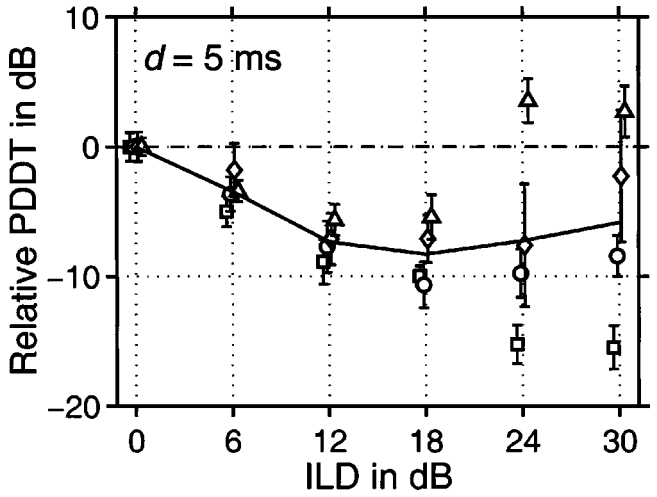


FIG. 8. Relative PDDT as a function of the ILD in the primary sound; the reflection delay, d, was 5 ms in this case. As in Fig. 3, individual thresholds are represented by different symbols, which are shifted relative to each other along the abscissa for clarity, and the average thresholds are connected by a solid line. The thresholds were normalized to the threshold value for ILD $=0$ dB (diotic), which is the same as for the ITD$=0$ μs and $d=5$ ms in experiment 1 (Fig. 4).

B. Results

Figure 8 shows the PDDT as a function of the ILD in the primary sound; the reflection delay was 5 ms in this case. As in Fig. 3, individual thresholds are denoted by different symbols and the average thresholds are connected by a solid line. The thresholds were normalized to the threshold value for ILD$=0$ dB, which is the same as for the ITD$=0$-μs and $d=5$-ms condition from experiment 1 (Fig. 4). In experiment 3, the variability between listeners was much larger than in experiment 1. Inter- and also intraindividual variability was particularly large for the larger ILDs (24 and 30 dB). In contrast to experiment 1, the PDDT for the diotic condition (ILD$=0$ dB) was larger than for most of the dichotic conditions. As the ILD in the primary sound was increased, the PDDT initially decreased, indicating that pitch strength *increased* with increasing lateral displacement between primary sound and reflection. For one listener (KK), the PDDT kept decreasing over the whole range of ILDs tested. For the other three listeners, the slope of the threshold function reversed when the ILD was increased above 18 dB. Simulations based on Zurek's (1979) model (see Sec. V) revealed that the data from experiment 3 were consistent with the central averaging hypothesis.

V. DISCUSSION

The current study was aimed to test whether the salience of the pitch associated with a single reflection of a broadband noise is determined by a central averaging of the monaural information mediated by the stimuli at the two ears or by the relative locations of the primary sound and the reflection. Experiment 1 measured pitch strength as a function of the lateral displacement between the primary sound and the reflection, whereby lateral displacement was produced by ITDs. The results showed that the shape of the function relating pitch strength to the difference in the ITDs of primary

sound and reflection depends on the reflection delay, and that the function is nonmonotonic for the shorter delays (Fig. 3). These findings contravene the directionality hypothesis, which predicts that pitch strength is a fixed, monotonically decreasing function of the displacement between the primary sound and the reflection. In this respect, the current results accord with those of Brüggen (2001), who used a room simulation system and multidimensional perceptual rating to show that, under certain circumstances, the adverse perceptual effects of coloration in realistic reverberant environments are not reduced in binaural compared to monaural listening. In fact, the results of the current experiment 1 could be mimicked by simply adding the stimuli from the two ears and presenting the merged stimulus identically to both ears (experiment 2; Fig. 6), indicating that the pitch strength of the dichotic stimuli is basically determined by the monaural information mediated by the stimuli at the two ears. This is in accordance with Zurek's (1979) central averaging hypothesis, which assumes that the strength of the pitch associated with reverberation is derived from an averaged representation of the stimuli from the two ears.

In experiment 3, where the displacement was produced by means of an ILD in the primary sound, the PDDT decreased with increasing lateral displacement between primary sound and reflection (Fig. 8), when the reflection gain, g, at threshold is expressed relative to a constant reference level, i.e., the level of the unattenuated primary noise at the right ear (see Sec. IV A). In the following we show that Zurek's (1979) central averaging model produces a good fit to the data from the ILD conditions of the current experiment 3.

Zurek implemented a simple quantitative version of the central averaging model and showed that the model produced a good fit to some of his experimental results. In particular, Zurek showed that the effect of the overall levels of the left- and right-ear stimuli could be explained by assuming that the auditory spectra of the stimuli from the two ears undergo nonlinear compression prior to central averaging. In Zurek's simple model, the central spectrum is given by

$$G(f)=[S_L(f)]^p+[S_R(f)]^p, \tag{1}$$

where f is frequency, p is the compression exponent ($p \leqslant 1$), and S_L and S_R are the auditory spectra of the stimuli from the left and right ears, respectively. Instead of calculating the auditory spectra of the stimuli, Zurek simply used their comb-filter power transfer functions. The comb-filter function of a RN consisting of an undelayed noise with amplitude A_1, and a delayed noise with amplitude A_2 is given by

$$S(f)=A_1^2+A_2^2+2A_1A_2\cos(2\pi fd), \tag{2}$$

where d is the delay. The peak-to-trough ratio of the comb-filter function—and thus the pitch strength of the RN—is maximal when the amplitudes of the undelayed and the delayed noises, A_1 and A_2, are equal, and decreases when either noise is attenuated relative to the other. In experiment 3, the relative levels of the original and the delayed noises depended on the ILD in the primary sound and on the gain factor, g. Figures 7(a) and (b) show schematic representations of the stimuli for two different gain factors, corresponding to two different trials in a threshold run. At the beginning of the run, when g is close to unity, the peak-to-trough ratio of the right-ear comb-filter function is large because the undelayed and delayed noises are about equally intense [Fig. 7(a)]. In contrast, the peak-to-trough ratio of the left-ear comb-filter function is small because the undelayed noise is attenuated by the ILD. As the gain, g, of the delayed noise is decreased during the threshold run, the peak-to-trough ratio of the right-ear comb-filter function decreases because the amplitude of the delayed noise decreases relative to that of the undelayed noise [Fig. 7(b)]. In contrast, the peak-to-trough ratio of the left-ear comb-filter function *increases* at first, because the amplitude of the delayed noise approaches that of the undelayed noise. The peak-to-trough ratio of the left-ear comb-filter function reaches a maximum when the undelayed and delayed noises are equally intense, that is, when g (in dB) is equal to $-$ILD. At that point, however, the overall level of the left-ear stimulus is lower than the overall level of the right-ear stimulus, and so, without compression, the peak-to-trough ratio of the averaged spectrum would be expected to be dominated by the right-ear comb-filter function. However, compression reduces the level difference between the ears, and thereby increases the relative weight of the left-ear comb-filter function.

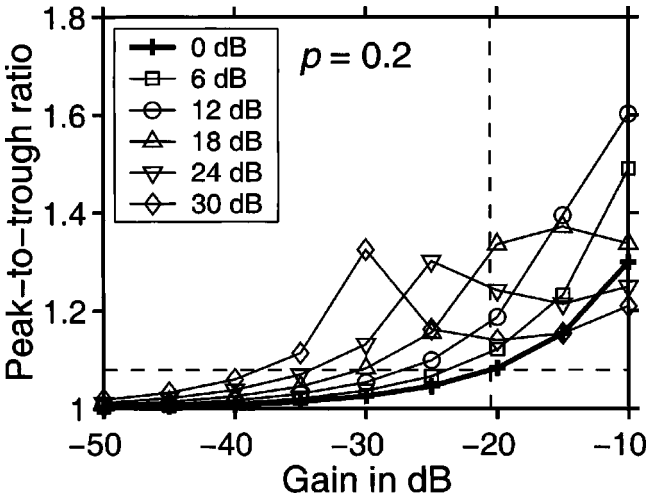


FIG. 9. Peak-to-trough ratio of the central averaged comb-filter functions [Eqs. (1) and (2)] for the stimuli from experiment 3, plotted as a function of the reflection gain, g. The parameter is the ILD (see the legend); the function for the diotic condition (ILD$=0$ dB) is highlighted with a bold line. The vertical dashed line marks the average measured threshold for the diotic condition; the horizontal dashed line shows the peak-to-trough ratio at which the function for the diotic condition intersects the vertical dashed line, and which was used as a threshold criterion (C) for the other ILD conditions. The compression exponent, p, was 0.2 in this example.

We used Eqs. (1) and (2) to calculate the central averaged spectra of the stimuli from experiment 3 for varying compression exponents, p. For each of the ILDs tested in experiment 3, and for each value of p, the central averaged spectrum was calculated for a range of reflection gains, g, comprising both sub- and suprathreshold values. The peak-to-trough ratio of the central averaged spectrum was defined as the ratio of the largest peak to the smallest trough plus 0.0001 (to avoid dividing through zero). Figure 9 shows the

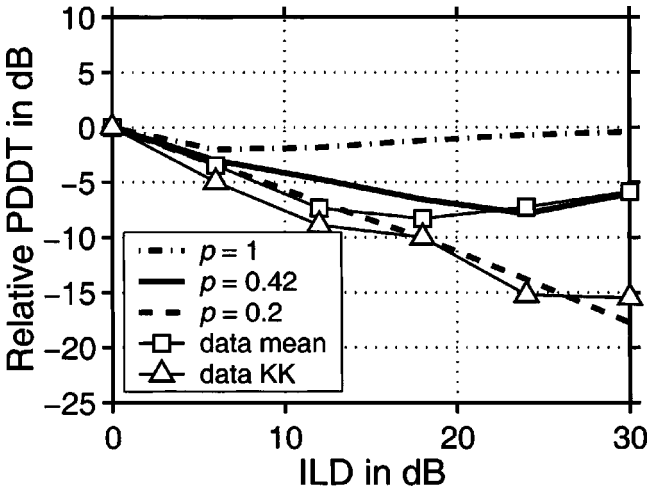


FIG. 10. Simulated PDDT (bold lines) for the ILD conditions from experiment 3 calculated with Zurek's (1979) simple version of the central averaging model. Different line types show the modeling results for different compression exponents, p (see the legend). The faint squares connected by a faint, solid line show the average data from the experiment; the triangles show listener KK's data, whose threshold function deviated from the other listeners' threshold functions in that it decreased monotonically with increasing ILD.

peak-to-trough ratio for a compression exponent $p=0.2$, plotted as a function of the reflection gain, g. The parameter is the ILD (see the legend); the bold line depicts the function for the diotic condition (ILD$=0$ dB). Threshold was taken to be that value of g, at which the peak-to-trough ratio exceeded a fixed criterion value, C. The criterion, C, represented by a horizontal dashed line in Fig. 9, was taken to be the peak-to-trough ratio of the central spectrum for the diotic condition (bold in Fig. 9) when g is equal to the average measured threshold for this condition. The average measured threshold for the diotic condition amounted to about -20.5 dB and is represented by a vertical dashed line in Fig. 9. Thus, the simulated threshold for the diotic condition was equal to the average measured threshold for this condition.

The bold lines in Fig. 10 show the simulated PDDT for three different compression exponents, p. The open squares connected by a faint solid line show the average measured threshold from experiment 3. When $p=1$ (no compression), the simulated threshold depends little on the ILD (bold, dash-dotted line), because the central averaged spectrum is dominated by the right-ear comb-filter function. When $p=0.2$ (strong compression), on the other hand, the simulated threshold decreases roughly linearly with increasing ILD up to the largest ILDs tested (bold, dashed line), which the average threshold does not. For intermediate compression, however, the model predicts the nonmonotonic pattern of the average threshold data fairly well. The deviation between simulated and average thresholds was minimal when $p=0.42$ (bold, solid line), which is comparable to the compression actually exerted by the cochlea and the auditory nerve (Oxenham and Moore, 1995). Further, Fig. 10 suggests that the interindividual differences in the shape of the threshold function from experiment 3 can be accounted for by changing the compression exponent, p, in the model. The simulated threshold function for $p=0.2$ (bold, dashed line) produces a reasonable fit to listener KK's threshold function (open triangles), which deviated from the other listeners' threshold functions in that it decreased monotonically over the whole range of ILDs tested. A good deal of the interindividual variability in experiment 3 was probably due to differences in the listeners' ability to focus on the ''better'' ear, that is the ear with the stronger pitch, even when the overall sound level at that ear was much weaker than at the other ear. We could have modeled the effect of directing attention to the better ear by multiplying its power spectrum with a weighting factor before averaging. A differential weighting of the two ears would have a similar net effect as the compression exponent in Zurek's model, namely to reduce the effect of the level difference between the ears.

The threshold functions from experiment 2 had similar shapes as the respective threshold functions from experiment 1 (Fig. 6), but varied over consistently smaller ranges, indicating that the way in which the auditory system combines pitch information from the two ears is different from a simple summation. While we have no conclusive explanation for the difference in threshold range, we think that it must result from some kind of nonlinear transformation in the auditory processing up to the pitch extraction stage. Preliminary modeling results indicate, however, that the difference cannot be explained by a simple compressive nonlinearity, like the one used in Zurek's central averaging model.

While Zurek's (1979) simple version of the central averaging model produces a good fit to the average data from the ILD conditions of experiment 3 and even explains some of the interindividual variability in those data, the model is too simplistic to be able to explain the results from the ITD conditions measured in experiment 1. That is because the comb-filter functions used in Zurek's original version of the model don't take into account that the resolution of cochlear filtering decreases with increasing filter frequency. The dashed and solid lines in Fig. 1(a) show the comb-filter functions of the RNs at the left and right ears, respectively, when the reflection delay is 2.5 ms, the reflection gain, g, is unity, and the ITD in the primary sound is 320 μs. Panel (b) shows the average of the two functions. The cosine ripples in the two comb-filter functions are out of phase in the frequency region just below 2 kHz, so in this frequency region they cancel out in the averaged function. Nevertheless, the *maximum* peak-to-trough ratio of the averaged function is similar to that of the individual functions, because the depth of the cosine ripples in the individual functions is the same at all frequencies and the ripples are in phase at very low and at higher frequencies (e.g., between 2 and 4 kHz). Thus, with the assumption that pitch strength is determined by the maximum peak-to-trough ratio of the central averaged spectrum, Zurek's original model would predict that pitch strength is independent of the ITD in the primary sound. That is in contradiction to the data and thus highlights the need to use physiologically more plausible spectral representations of the stimuli (auditory spectra). Alternatively, the monaural representations of the pitch-related information in the stimuli might by based on the timing of auditory-nerve action potentials rather than the spatial distribution of neural activity across the tonotopic array (Yost *et al.*, 1996; Patterson *et al.*,

1996; Krumbholz *et al.*, 2001, 2003), so it may be more appropriate to average the monaural time-interval representations (Patterson *et al.*, 1994) of the stimuli at the two ears rather than their auditory spectra. This will be investigated in a future study.

ACKNOWLEDGMENTS

This research was supported by the Medical Research Council (G9901257), the Deutsche Forschungsgemeinschaft (Lu342/4-2), the Alexander von Humboldt-Stiftung, and the Austrian Academy of Sciences (APART Project 524).

[1]In this paper, the terms ''pitch'' and ''coloration'' will be used interchangeably.

[2]The term ''auditory spectrum'' refers to the internal representation of the stimulus power spectrum, that is, the distribution of neural activity across the tonotopic array. The internal spectral representation of the stimulus is often also referred to as the excitation pattern (see, e.g., Moore and Glasberg, 1987).

Bilsen, F. A. (**1966**). ''Repetition pitch: Monaural interaction of a sound with the repetition of the same, but phase-shifted sound,'' Acustica **17**, 295–300.

Bilsen, F. A., and Ritsma, R. J. (**1969/70**). ''Repetition pitch and its implication for hearing theory,'' Acustica **22**, 63–73.

Bilsen, F. A., and Ritsma, R. J. (**1970**). ''Some parameters influencing the perceptibility of pitch,'' J. Acoust. Soc. Am. **47**, 469–475.

Blauert, J. (**1997**). *Spatial Hearing*, revised edition (MIT Press, Cambridge, MA).

Brüggen, M. (**2001**). ''Coloration and binaural decoloration in natural environments,'' Acustica **87**, 400–406.

Handel, S., and Patterson, R. D. (**2000**). ''The perceptual tone/noise ratio of merged, iterated rippled noises with octave, harmonic, and nonharmonic delay ratios,'' J. Acoust. Soc. Am. **108**, 692–695.

Koenig, W. (**1950**). ''Subjective effects in binaural hearing,'' J. Acoust. Soc. Am. **22**, 61–62.

Krumbholz, K., and Nobbe, A. (**2002**). ''Buildup and breakdown of echo suppression for stimuli presented over headphones—the effects of interaural time and level differences,'' J. Acoust. Soc. Am. **112**, 654–663.

Krumbholz, K., Patterson, R. D., Nobbe, A. and Fastl, H. (**2003**). ''Microsecond temporal resolution in monaural without spectral cues?'' J. Acoust. Soc. Am. **113**, 2790–2800.

Krumbholz, K., Patterson, R. D., and Nobbe, A. (**2001**). ''Asymmetry of masking between noise and iterated rippled noise: Evidence for time-interval processing in the auditory system,'' J. Acoust. Soc. Am. **110**, 2096–2107.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.

Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.

Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Overshoot and the severe departure from Weber's law,'' J. Acoust. Soc. Am. **97**, 2442–2453.

Patterson, R. D. (**1994**). ''The sound of a sinusoid: Time-interval models,'' J. Acoust. Soc. Am. **96**, 1419–1428.

Patterson, R. D., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). ''The relative strength of tone and noise components of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 3286–3294.

Patterson, R. D., Yost, W. A., Handel, S., and Datta, A. J. (**2000**). ''The perceptual tone/noise ration of merged iterated rippled noises,'' J. Acoust. Soc. Am. **107**, 1578–1588.

Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). ''Adjustment and discrimination measurements of the precedence effect,'' J. Acoust. Soc. Am. **93**, 2923–2932.

Wightman, F. L., and Kistler, D. J. (**1993**). ''Sound localization,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 155–192.

Yost, W. A., and Hill, R. (**1978**). ''Strength of the pitches associated with rippled noise,'' J. Acoust. Soc. Am. **64**, 485–492.

Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.

Zurek, P. M. (**1979**). ''Measurements of binaural echo suppression,'' J. Acoust. Soc. Am. **66**, 1750–1757.

The influence of duration and level on human sound localization

Joyce Vliegen and A. John Van Opstal[a)]
Department of Medical Physics and Biophysics, University of Nijmegen, Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands

(Received 30 March 2003; revised 12 September 2003; accepted 10 January 2004)

The localization of sounds in the vertical plane (elevation) deteriorates for short-duration wideband sounds at moderate to high intensities. The effect is described by a systematic decrease of the elevation gain (slope of stimulus–response relation) at short sound durations. Two hypotheses have been proposed to explain this finding. Either the sound localization system integrates over a time window that is too short to accurately extract the spectral localization cues (*neural integration hypothesis*), or the effect results from cochlear saturation at high intensities (*adaptation hypothesis*). While the neural integration model predicts that elevation gain is independent of sound level, the adaptation hypothesis holds that low elevation gains for short-duration sounds are only obtained at high intensities. Here, these predictions are tested over a larger range of stimulus parameters than has been done so far. Subjects responded with rapid head movements to noise bursts in the two-dimensional frontal space. Stimulus durations ranged from 3 to 100 ms; sound levels from 26 to 73 dB SPL. Results show that the elevation gain decreases for short noise bursts at all sound levels, a finding that supports the integration model. On the other hand, the short-duration gain also decreases at high sound levels, which is in line with the adaptation hypothesis. The finding that elevation gain was a nonmonotonic function of sound level for all sound durations, however, is predicted by neither model. It is concluded that both mechanisms underlie the elevation gain effect and a conceptual model is proposed to reconcile these findings. [DOI: 10.1121/1.1687423]

PACS numbers: 43.66.Qp, 43.66.Ba, 43.66.Mk [AK] Pages: 1705–1713

I. INTRODUCTION

In order to localize a sound, the auditory system relies on binaural and monaural acoustic cues. Binaural cues result from interaural differences in sound level (ILD) and timing (ITD), which relate to sound position in the horizontal plane (azimuth). Monaural cues consist of direction-dependent spectral shape information caused by reflection and diffraction at torso, head, and pinnae (described by head-related transfer functions, or HRTFs). These spectral cues are essential to resolve front–back confusions and to localize sounds in the vertical plane (elevation; see Blauert, 1996, for a review). Although the binaural difference cues are extracted quite reliably under a wide variety of stimulus conditions and spectra, the transformation of the HRTFs into a reliable estimate of sound-source elevation is a challenging problem for several reasons.

First, the spectrum at the eardrum (which will be denoted by the sensory spectrum) is a linear convolution of the (*a priori* unknown) sound-source spectrum with the particular HRTF associated with the unknown sound direction. Thus, in extracting sound-source elevation, the auditory system is faced with an ill-posed problem. One way to deal with this problem would be to incorporate *a priori* assumptions about potential source spectra. For example, if the source spectrum is assumed flat, the sensory spectrum is identical to the HRTF. Yet, subjects are able to localize a variety of broadband sound spectra that are not flat with remarkable accuracy (Oldfield and Parker, 1984; Wightman and Kistler, 1989; Middlebrooks and Green, 1991; Hofman and Van Opstal, 1998). Apparently, the assumptions about potential source spectra are more relaxed.

If the assumption holds that source spectra do not resemble any of the HRTFs, the spectral correlation between the sensory spectrum and each of the HRTFs can be shown to peak exactly at the correct HRTF (Middlebrooks, 1992; see Hofman and Van Opstal, 1998, for details). Such a strategy would allow accurate localization for a large class of nonflat stimulus spectra. However, when amplitude variations within the source spectrum become too large, the localization accuracy of sound elevation deteriorates (Wightman and Kistler, 1989; Hofman and Van Opstal, 2002).

A second problem concerns the presence of considerable spectro-temporal variations in natural sounds. Until recently, localization studies have typically used long-duration stimuli with stationary spectro-temporal properties. Not much is known as to how nonstationary sounds affect sound localization performance.

Hofman and Van Opstal (1998) studied the effects of different spectro-temporal stimulus properties on sound localization performance in the two-dimensional frontal hemifield. The only response variable that depended systematically on the temporal stimulus parameters was the *slope* of the stimulus–response relation for the elevation components (i.e., the elevation gain). In particular, for stimuli with durations shorter than several tens of ms the gain started to decrease with decreasing burst duration. Neither response variability, nor the azimuth responses depended on the stimulus parameters. Based on their results, Hofman and Van Opstal

a)Electronic mail: johnvo@mbfys.kun.nl

(1998) proposed that the sound localization system needs to integrate about 40–80 ms of broadband input to yield a stable estimate of sound-source elevation (the *neural integration hypothesis*).

Recently, an alternative explanation for these data has been put forward (Macpherson and Middlebrooks, 2000). In that proposal, the decrease in gain is due to the so-called "*negative level effect*" reported earlier by Hartmann and Rakerd (1993). In this earlier study, subjects were unable to localize high-level clicks (>86 dB SPL), with errors decreasing for intermediate (74–86 dB) and lower (68–80 dB) sound levels. Hartmann and Rakerd (1993) suggested that this effect was caused by saturation of cochlear excitation patterns. As a consequence, the auditory system would fail to resolve the spectral details of the clicks. For long-duration stimuli, the system would adapt to the high sound level, so that a reliable elevation estimate could be based on later portions of the signal (the *adaptation hypothesis*).

To elaborate on this possibility, Macpherson and Middlebrooks (2000) presented short- (3 ms) and long-duration (100 ms) noise bursts at sensation levels (SL) between 25 and 60 dB. Like Hofman and Van Opstal (1998), they found that elevation gains were lower for short-duration stimuli than for long-duration stimuli, but only at high sensation levels. Moreover, when the short noise bursts were presented within spatially diffuse noise, elevation gain depended on the level of the masker. Elevation gains increased with increasing masker level until a masked sensation level of about 40 dB. These results are at odds with the neural integration hypothesis, which would predict no effect of signal level. However, they are predicted by the adaptation model, as the background noise would activate the putative adaptive mechanism prior to the onset of the 3-ms noise bursts. At higher masker levels, performance decreased, which could be due to a low signal-to-noise ratio.

Macpherson and Middlebrooks (2000) concluded that the results of all three studies can thus be explained by the negative level effect. Note, however, that this mechanism does not specify how and why only the elevation *gain* would be affected by cochlear saturation, and why other parameters, e.g., response variability, or azimuth localization, remain unaffected.

Note also that the fixed stimulus level of 70 dB SPL employed by Hofman and Van Opstal (1998) corresponds to the low end of intensities used by Hartmann and Rakerd (1993). Moreover, Frens and Van Opstal (1995) had reported similar gain-duration effects for stimuli of only 60 dB SPL.

The results of Hofman and Van Opstal (1998) and Macpherson and Middlebrooks (2000) are difficult to compare directly because of differences in methodology. First, Hofman and Van Opstal (1998) used a variety of stimulus durations, mixed randomly within a single recording session, whereas Macpherson and Middlebrooks (2000) collected responses to two different stimulus durations (3 and 100 ms) in different blocks of trials. Second, while Hofman and Van Opstal (1998) presented all stimuli at 70 dB SPL, Macpherson and Middlebrooks (2000) employed various intensities, but quantified as sensation levels. These two measures are not readily equated. Third, the pointer used to indicate perceived sound direction differed in the two studies: eye movements (restricted to the 35-deg oculomotor range) by Hofman and Van Opstal (1998) vs head movements over a much larger measurement range by Macpherson and Middlebrooks (2000).

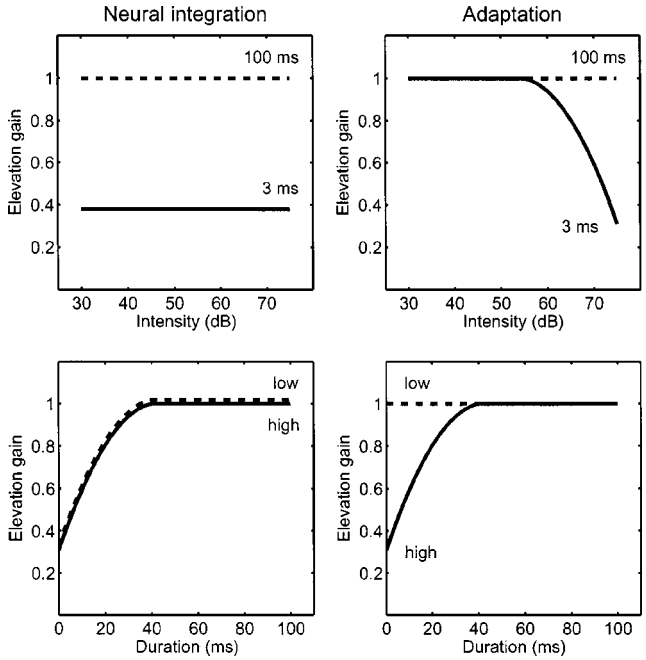


FIG. 1. Predictions of the neural integration model (left) and the cochlear adaptation hypothesis (right). Top row: elevation gain as a function of intensity for two durations. Bottom row: elevation gain as a function of duration for low and high intensities. The adaptation model predicts a decrease of elevation gain for short-duration stimuli at high intensities only, and a stable gain for longer-duration stimuli at all levels. The neural integration model predicts a decrease of the gain with duration at all stimulus levels, while gain is insensitive to stimulus level.

Finally, both studies measured only a small portion of the duration-intensity parameter space, with minor overlap. Therefore, to allow for a better comparison of both data sets, we have included and extended the measurements of both studies by employing a range of noise durations (3–100 ms) and sound levels (26–73 dB SPL). Up to 16 different stimulus conditions were measured within the same recording session, and were randomly interleaved. A summary of the expected results for the two hypotheses is provided in Fig. 1.

II. METHODS

The experiment consisted of three sessions, differing slightly in the parameter values used. In the first session we used durations ranging from 3 to 100 ms. We found that the largest changes in the results occurred for durations between 3 and 30 ms. Therefore, in later sessions we restricted the duration values to this range. These last two sessions consisted of stimuli with the same range in durations, but with different, slightly overlapping, intensity ranges.

A. Subjects

Two female and seven male subjects participated in the experiments. Their age ranged from 22 to 44 years. Two of the subjects (JV and JO) were the authors of this paper. Five other subjects were experienced in sound localization

studies. Subjects FF and JM had no previous localization experience. Before the actual experiment started, these inexperienced subjects were given a short practice session to get familiar with the stimuli and the localization paradigm. All subjects had normal binaural hearing (absolute thresholds within 20 dB HL at frequencies between 250 and 8000 Hz).

Subject JV participated in all three sessions. Subjects JO, FF, and JM participated in the first session only, while the remaining five subjects participated in sessions 2 and 3.

B. Apparatus

Experiments were conducted in a completely dark and sound-attenuated room with dimensions $L \times W \times H = 3.5 \times 2.45 \times 2.45\ m^3$. The room had an ambient background sound level of 20 dBA SPL. Horizontal and vertical head movements were measured with the search-coil technique. Subjects wore a lightweight helmet (about 150 g), consisting of a narrow strap above the ears, which could be adjusted to fit around the subject's head, and a second strap that ran over the head. A small coil was mounted on the latter. Two orthogonal pairs of coils were attached to the room's edges to generate the horizontal (60 kHz) and vertical (80 kHz) magnetic fields. The head-coil signal was amplified and demodulated (Remmel Labs), after which it was low-pass filtered at 150 Hz (Krohn-Hite 4413) and then stored on hard disk at a sampling rate of 500 Hz/channel for subsequent off-line analysis.

Subjects were seated comfortably in the center of the room facing a frontal hemisphere (radius: 1.0 m) that consisted of a thin wooden framework with 12 spokes and five concentric rings. This setup thus defined a polar coordinate system with its origin at the straight-ahead position. Target excentricity, R, is measured as the angle with respect to the straight-ahead position, whereas target direction, ϕ, is measured in relation to the horizontal meridian. For example, $R = 0$ deg corresponds to straight ahead for each ϕ, and $\phi = 0$, 90, 180, and 270 deg (for $R > 0$) corresponds to right, up, left, and down, respectively. On the hemisphere, a total of 58 small broad-range loudspeakers (Monacor MSP-30) were mounted at directions $\phi = 0$, 30, 60, 90,..., 330 deg (corresponding to each of the 12 spokes) and excentricities of $R = 0$, 15, 30, 45, 60, and 75 deg (corresponding to the five rings). At the outer ring ($R = 75$ deg) part of the framework (at downward directions $\phi = 240$, 270, and 300 deg) was removed to allow for space for the subject's legs. A thin glassfiber ended in the center of each speaker, through which a well-defined visual stimulus (0.15-deg diameter, 1.5-Cd/m^2) could be presented that originated from a red and green LED mounted behind the speaker. The peripheral LEDs were used to calibrate the head-coil signals at the start of an experimental session (see below), while the center LED at $(R, \phi) = (0,0)$ deg served as a fixation light at the start of a localization trial. The polar target coordinates (R, ϕ) were transformed into azimuth-elevation angles (α, ε), in the off-line analysis of the data (see Sec. II F and Hofman and Van Opstal, 1998, for details).

In the first experimental session, only the speakers at the first three rings of the hemisphere were used ($R = 0$, 15, 30, 45 deg; $N = 37$ locations). For the second and third session the speakers of all five rings were used, except for the central speaker at straight ahead; $N = 57$.

The height of the chair was adjusted to align the center of the subject's head with the center of the hemisphere. Walls, ceiling, and floor, as well as the spokes and rings of the hemifield, were covered with black sound-absorbing foam that eliminated acoustic reflections down to 500 Hz (Schulpen Schuim, The Netherlands).

TABLE I. Detection thresholds in dB SPL for all subjects for the four stimulus durations employed in the second and third session.

Duration	JV	MW	HV	MZ	WV	FW
3 ms	23	30	31	27	25	31
6 ms	21	25	25	19	18	19
14 ms	20	21	13	17	18	21
30 ms	20	21	13	12	17	20

C. Stimuli

Acoustic stimuli were generated digitally with a Tucker-Davis System II, using a TDT DA1 16-bit digital-to-analog converter (50-kHz sampling rate). Stimuli were then passed to a TDT PA4 programmable attenuator, which controlled the sound level. All stimuli consisted of independently generated Gaussian white noise with 0.5-ms sine-squared on- and off-set ramps.

In the first session, durations of 3, 10, 31, and 100 ms were used, with intensities of 26, 36, 46, and 56 dB SPL (a total of 592 trials per run and two or three runs per subject). In the second and third session, durations of 3, 6, 14, and 30 ms were used. Sound levels were at 33, 43, 53, and 63 dB SPL for the second session (one run of 912 trials) and at 58, 68, and 73 dB SPL for the third session (one run of 684 trials).

D. Sensation levels

For the six subjects that participated in sessions 2 and 3, free-field detection thresholds for broadband noise bursts of 3, 6, 14, and 30 ms were determined. Sounds were presented from the center speaker in the sound-attenuated room. Listeners performed a two-interval, two-alternative, forced-choice task where sound level was controlled by a three-down, one-up adaptive tracking procedure (Levitt, 1971). For all subjects, thresholds decreased with increasing noise duration. Table I summarizes the results of these measurements for all subjects. From these data sensation levels (SL) were computed by subtracting the thresholds from the SPL values of the stimuli as recorded at the level of the subject's head.

E. Recording paradigm

All measurements were performed in darkness. When making a head saccade in darkness, the eyes will typically not remain centered in the head. Especially for peripheral target locations, the position of the eyes in the head will be quite excentric (exceeding 20 deg), resulting in potentially large (and variable) undershoots of the measured head posi-

tion if subjects use both eyes and head to point to the target (Goossens and Van Opstal, 1997). To circumvent this potential problem, a thin aluminum rod with a dim red LED (0.15 Cd/m^2) attached to its end protruded from the helmet's left side. The rod was adjusted such that the LED was positioned in front of the subject's eyes at a distance of about 40 cm. At the start of a trial, the subject had to align this rod LED with the central LED of the hemisphere, while keeping his head in a comfortable straight-ahead position. The rod LED thus served as a head-fixed pointer during the experiments. Pointing with the LED to the perceived location of the target ensured that the eyes remained at a fixed, central position in the head while pointing.

Each recording session started with a calibration run in which the subject had to align the rod LED with each of the LEDs on the hemisphere. After calibration, head position was known with an absolute accuracy of 3% or better over the entire measurement range.

In subsequent blocks, the sound stimuli were presented. Each trial started by presenting the central fixation LED. After a randomly selected fixation period of 1.5 to 2.0 s, the fixation LED was switched off and 400 ms later the sound stimulus was presented at a peripheral location. The subject's task was to point the rod LED as quickly and as accurately as possible towards the perceived sound location. No feedback was given about performance. As stimuli were always extinguished well before the initiation of the head movement (typical reaction times about 200–300 ms), all experiments were conducted under fully open-loop conditions.

For all experiments, the order of stimulus conditions and positions was randomized throughout a session.

F. Data analysis and statistics

The coordinates of the target locations and head-movement responses are described in a double-pole coordinate system, in which the origin coincides with the center of the head. The horizontal component, azimuth α, is defined as the direction relative to the vertical median plane, whereas the vertical component, elevation ϵ, is defined as the direction relative to the horizontal plane through the ears (Knudsen and Konishi, 1979).

From the calibration run, the raw head position signals and the corresponding LED coordinates were used to train two three-layer backpropagation neural networks that mapped the raw data signals to the calibrated head position signals (azimuth and elevation angles, respectively). This was done to account for minor cross talk between horizontal and vertical channels and minor inhomogeneities in the magnetic fields (Goossens and Van Opstal, 1997). Goal-directed head movements were identified in the calibrated response data. The endpoint of the first head movement after stimulus onset, where response azimuth and elevation were stable, was defined as the response position.

Head saccades with a reaction time *re*: stimulus onset of less than 80 ms or above 800 ms were discarded from further analysis. Earlier responses are assumed to be predictive and are usually very inaccurate. Later responses are considered to be caused by inattention of the subject. Typically, less than 2% of the responses had to be discarded on the basis of these criteria.

For each stimulus condition (fixed stimulus duration and sound level), a linear regression line was fitted through the stimulus–response relations for azimuth (α) and elevation (ε) components, respectively, by applying the least-squares error criterion

$$\alpha_R = G_\alpha \cdot \alpha_T + b_\alpha ,$$
$$\varepsilon_R = G_\varepsilon \cdot \varepsilon_T + b_\varepsilon , \tag{1}$$

where (α_R, ϵ_R) are the head-movement response components, $(\alpha_T, \varepsilon_T)$ are the target coordinates; $(G_\alpha, G_\varepsilon)$ are the slopes of the regression lines (here called the response gain), and $(b_\alpha, b_\varepsilon)$ (in deg) are the offsets (response bias). The bootstrap method was used to estimate the standard deviations of the slopes, offsets, and Pearson's linear correlation coefficients (Press *et al.*, 1992).

To quantify the effects of stimulus duration and sound level on the stimulus–response relation, we also performed a nonlinear regression on the entire data set (all stimulus conditions and recording sessions pooled; elevation data only). In this regression, the elevation gain, G_ε was a (nonlinear) function of duration and sound level (five free parameters; see the Appendix for details).

Finally, to enable a quantitative comparison of the relative contributions of stimulus duration and stimulus level on the response elevations across the different stimulus conditions, we also performed two normalized multiple-linear regressions on two relevant cross sections through the data (see Sec. III).

III. RESULTS

Typical localization results of the first experimental session are presented in Fig. 2, which shows the endpoints of the azimuth and elevation components of the head-movement responses of subject FF together with the fitted linear regression lines. We found for all subjects that sound-source azimuth (○) was localized accurately with performance remaining rather stable for all test conditions. In contrast, the elevation response components (▲) depended strongly on the different stimulus parameters. Correlation coefficients for the stimulus–response relations were typically high. Both for the azimuth and elevation response components they were found to be close to 1.0, except for the shortest stimuli at the lowest sound level (26 dB), where correlations dropped to around zero for two subjects for both azimuth and elevation. These stimuli were probably close to, or even below, the detection threshold for these subjects. Azimuth gains were stable for all conditions, except for the 3-ms condition at the lowest intensity, where gains were considerably lower for those same two subjects. For the other two subjects in this stimulus condition azimuth gains decreased only slightly. For the elevation responses, gains appeared to increase with increasing duration for all stimulus levels. For stimulus durations between 30 and 100 ms, the response gain leveled off. For fixed durations the slope of the elevation regression line also varied with stimulus level.

In Fig. 3 the gains for the azimuth response components of sessions 2 and 3 are plotted as a function of stimulus duration for all intensities and all six subjects. For most sub-

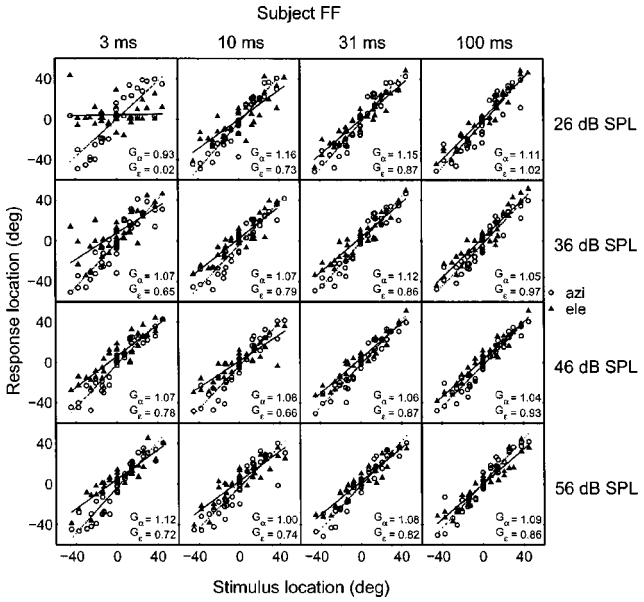


FIG. 2. Stimulus response relations for azimuth (○) and elevation (▲) response components of subject FF for four different stimulus durations (columns) and four different stimulus intensities (rows). Data taken from session 1. Best-fit regression lines (dotted: azimuth, solid: elevation) are also shown, together with the values of the azimuth and elevation gains.

jects, gains were around 1.0 or slightly higher, except for subject HV, whose gains were around 1.4 in the third session. Gains remained stable across the different stimulus conditions. Note also that the gain values could vary considerably between sessions. This is apparent for most of the subjects, for whom the data appear to split into two separate clusters,

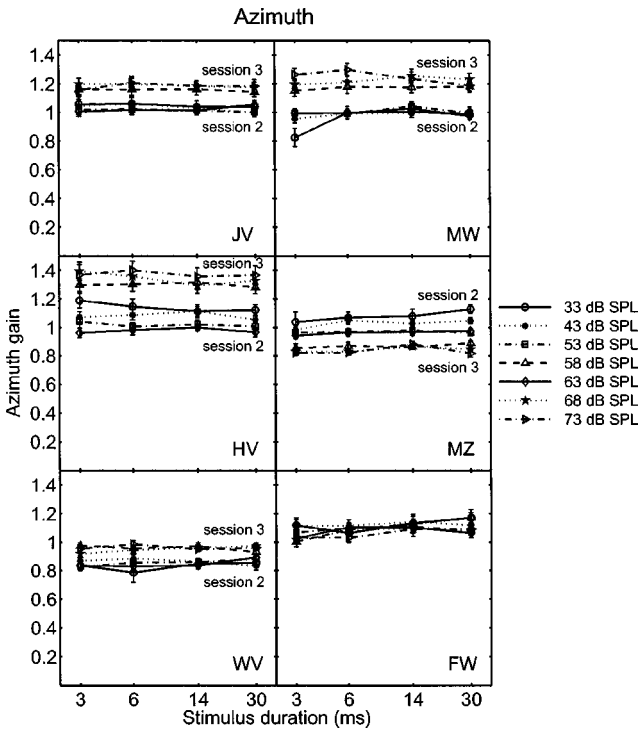


FIG. 3. Azimuth gains as a function of stimulus duration for all six subjects of sessions 2 and 3. The different line styles and symbols in each panel correspond to the different stimulus levels. Note the absence of any consistent trend and apparent separation of the obtained gain values for the two sessions in all but one subjects.

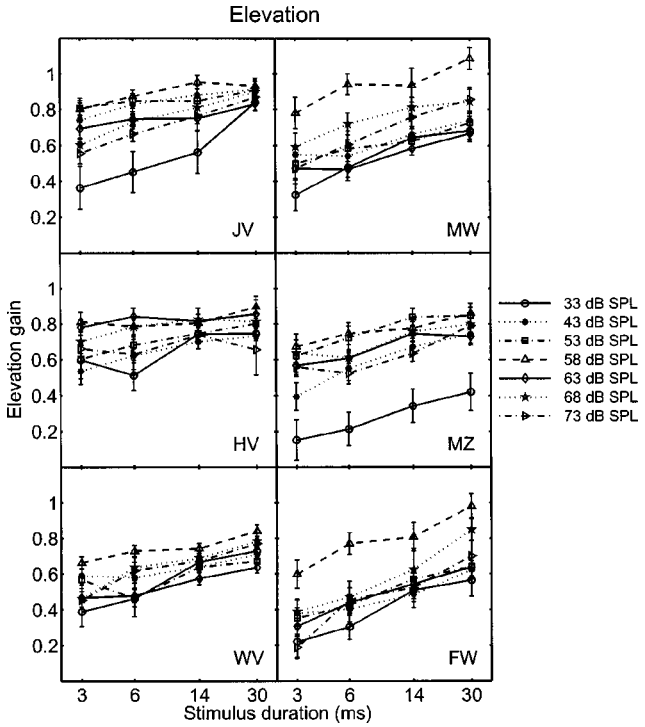


FIG. 4. Elevation gains as a function of stimulus duration for all six subjects of sessions 2 and 3. Same format as Fig. 3. Note the clear effect of stimulus duration on elevation gain for all subjects and at all stimulus levels.

each one corresponding to a different recording session (2 or 3). It might be due to simple day-to-day variation or to the different intensity ranges used in the two sessions.

The data for the elevation gains obtained from these same sessions are shown in Fig. 4 in the same format as Fig. 3. Although the absolute gain values differed between subjects, qualitatively similar patterns emerged for all subjects in both recording sessions. Elevation gain covaried with sound duration for all stimulus intensities, although the effect was most prominent at low and high levels. Gains were lowest for the 3-ms bursts at 33 dB SPL, where elevation gains were typically around 0.2–0.4. The fact that elevation gain increased with increasing sound duration for all stimulus levels, and not just for the highest stimulus levels, provides support for the neural integration hypothesis and is inconsistent with the adaptation hypothesis.

As can be noted in Fig. 4, elevation gain also appeared to vary with stimulus intensity. This feature is better illustrated in Fig. 5, which shows elevation gain as a function of absolute sound level (in dB SPL) for all stimulus durations and all subjects who participated in sessions 2 and 3. The gains were lowest for the lowest sound intensities, and especially for the shortest noise bursts. For intermediate sound levels, gains increased to a maximum value, to decrease again for higher sound levels. This latter phenomenon is reminiscent of the negative level effect reported by Hartmann and Rakerd (1993) and Macpherson and Middlebrooks (2000). It can be seen, however, that gains varied with intensity for all stimulus durations, not only the shortest ones, although the changes tended to be smaller for longer stimulus durations. The fact that elevation gains increased with increasing sound level for low intensities, a positive level

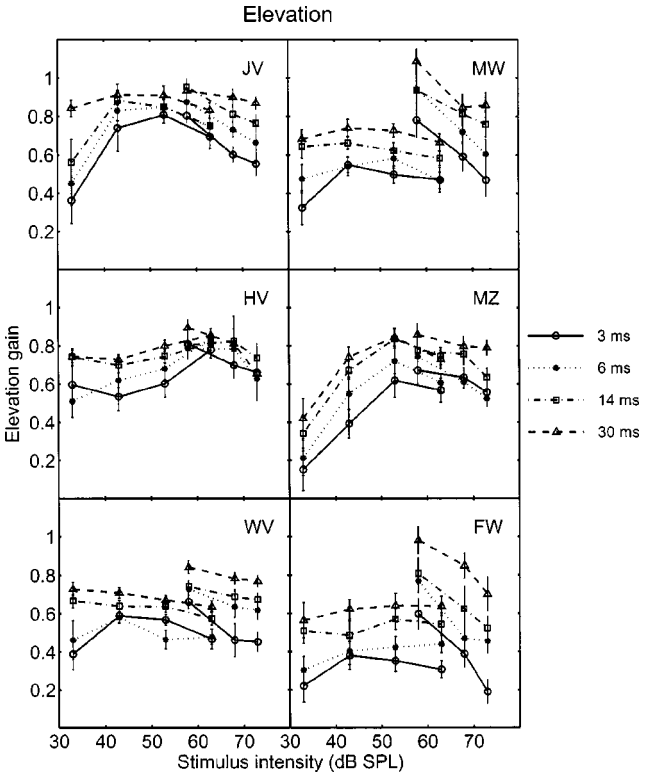


FIG. 5. Elevation gains as a function of intensity (in dB SPL) for all six subjects of sessions 2 and 3. The different line styles and symbols correspond to the different stimulus durations. Note the consistent nonmonotonic changes of elevation gain with stimulus level. Note also that both a positive and a negative level effect were observed for all stimulus durations.

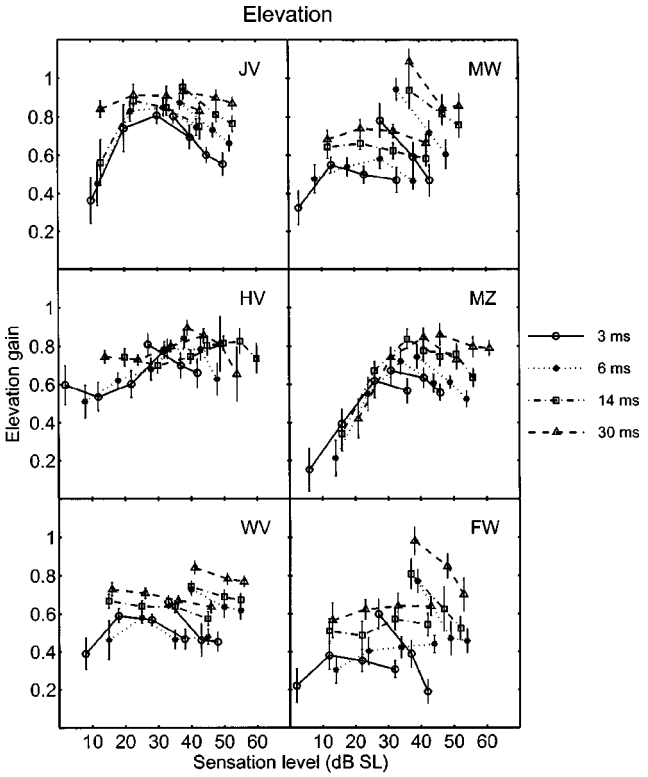


FIG. 6. Elevation gains as a function of sensation level for all six subjects of sessions 2 and 3. Same format as Fig. 5.

effect, was not predicted by either the neural integration hypothesis or the adaptation hypothesis.

It should be noted that, as in Fig. 3, three of the subjects (MW, WV, FW) showed different gain values for similar stimulus conditions in the two sessions, with higher gains in session 3 than in session 2.

For a better comparison with the data of Macpherson and Middlebrooks (2000), elevation gains are plotted as a function of sensation level in Fig. 6. Elevation gains increased strongly at the lower sensation levels; above about 45 dB SL the gains decreased. This trend was obtained for all stimulus durations.

In order to describe the effects of stimulus duration and intensity for the entire data set, we performed a nonlinear regression on all elevation responses of a given subject, pooled across recording sessions and stimulus parameters. To that end, the gain in the regression model of Eq. (1) was taken to be a function of both intensity and duration, yielding $G_\varepsilon(D,I)$. The shape of this function was estimated on the basis of the results shown in Figs. 2, 4, 5, and 6. Thus, the intensity dependence of the elevation response gain was described by a simple parabolic function, to incorporate both the positive and negative level effects. The effect of stimulus duration was described by a saturating exponential, which levels off for long durations. The response bias had a fixed value. The regression model had five free parameters, $\beta_1-\beta_5$, which were found by minimizing the mse (fitting between 1005–3195 data points; see the Appendix for details). The model yielded a good description of the data, with consistent parameter values for the different subjects and recording sessions, and high R^2-values (see Table III for results). On the basis of these results we estimated the stimulus intensity for which the elevation response gains reached a maximum at $-\beta_1/(2\beta_2)$. Values were typically between 50–70 dB SPL, with a median of 62 dB SPL.

According to the adaptation hypothesis, the negative level effect is obtained for short-duration stimuli only. Our data, however, suggest that a negative level effect occurs at all stimulus durations. Although this observation is supported by the nonlinear regression model, it is not possible to quantitatively compare the strength with which each stimulus parameter influences the elevation responses because the different variables are expressed in different units. A simpler way to quantify these effects would therefore be to convert to dimensionless variables (i.e., normalization).

To restrict the analysis to the negative level effect only, it is necessary to incorporate only that section of the data where it occurs (for the highest stimulus levels). To that end, we performed a multiple linear regression on the normalized elevation gains ($N=12$) obtained by linear regression [Eq. (1)] on the data from session 3 only (for which $L=58$, 68, 73 dB SPL)

$$\hat{G}_\epsilon=\beta_L\cdot\hat{L}+\beta_D\cdot\hat{D},\ \text{ with }\ \hat{X}\equiv\frac{X-\mu_X}{\sigma_X},\tag{2}$$

with μ_X and σ_X the mean and variance of the respective variable (L is stimulus level in dB SPL, D is duration in ms, and G_ε is the measured elevation gain). In this regression, β_L and β_D are the (dimensionless) partial regression coefficients. The resulting regression parameters for each subject are listed in Table II (left portion). Note that all coefficients

TABLE II. Normalized partial regression coefficients for stimulus duration and intensity for the negative level effect (data of the third session: 58, 68, 73 dB SPL), and for the positive level effect (data of the second session: 33, 43, 53 dB SPL) [Eq. (2)].

	Negative level			Positive level		
Subject	β_L	β_D	R^2	β_L	β_D	R^2
JV	−0.61	0.68	0.79	0.69	0.50	0.66
MW	−0.68	0.65	0.85	0.27	0.83	0.70
HV	−0.75	0.32	0.59	0.27	0.80	0.64
MZ	−0.54	0.78	0.86	0.86	0.43	0.91
WV	−0.44	0.78	0.76	0.10	0.78	0.54
FW	−0.62	0.73	0.90	0.32	0.88	0.84
Mean	−0.61	0.66	⋯	0.42	0.70	⋯

for stimulus level are indeed negative, while for sound duration they are positive. More importantly, the absolute values of the two parameters are roughly equal, indicating that at high stimulus levels both stimulus factors influence the elevation gain to a comparable degree.

A positive level effect was obtained for lower stimulus levels and for all stimulus durations. To quantify this effect we performed a multiple linear regression [Eq. (2)] on the normalized elevation gains ($N=12$) for the lower stimulus levels ($L=33$, 43, 53 dB SPL). The resulting regression parameters are listed in Table II (right portion). The coefficients for stimulus level are all positive and their absolute values are slightly smaller than for the negative level effect. For sound duration, the values are roughly equal to the duration values for the negative level effect.

If this positive level effect were entirely due to a poor signal-to-noise ratio (SNR), the response variability would be expected to systematically vary with stimulus duration and sound level in a similar way as the response gain. To test for this, Fig. 7 shows the response variability (defined as the mean-squared error around the regression line) of the data from sessions 2 and 3 as a function of stimulus intensity for the different stimulus durations. Note that only for the lowest stimulus intensities and shortest durations was the response variability higher than for the other conditions for most subjects. Only for subjects MW and HV did the variability increase for high intensities, but this was true for all durations. Interestingly, the variability obtained for the high-intensity, short-duration stimuli was indistinguishable from the other stimulus conditions. For the majority of stimulus levels, the variability is quite comparable (around 10 deg).

IV. DISCUSSION

By systematically varying both sound duration and sound level within the same experimental session, the current experiments confirm and extend recent reports by Hofman and Van Opstal (1998) and Macpherson and Middlebrooks (2000), and provide more insight into the combined effects of these stimulus parameters on human sound localization.

The results show that the azimuth response components remained virtually unaffected for all stimulus conditions (Figs. 2, 3) except for stimuli with an intensity around the detection threshold. However, the response *elevation* gain was strongly affected by both stimulus parameters (Figs. 2, 4–6). Neither response bias (not shown), nor response variability (Fig. 7) was systematically related to the stimulus parameters. Our results are summarized in Fig. 8, which plots, in the format of Fig. 1, the prediction of Eq. (A2) (see the Appendix) applied to the pooled elevation gain data of subject JV. A comparison of Figs. 8 and 1 indicates that neither the neural integration model, nor the adaptation model explains the data well.

For all subjects, elevation gains increased with increasing sound duration, until a plateau was reached for durations above 30 ms. Although the effect was most conspicuous at the lowest and highest sound levels, it was apparent for all stimulus intensities tested. These results, especially for the higher intensities, are in good agreement with the results reported by Hofman and Van Opstal (1998), who tested their subjects at 70 dB SPL.

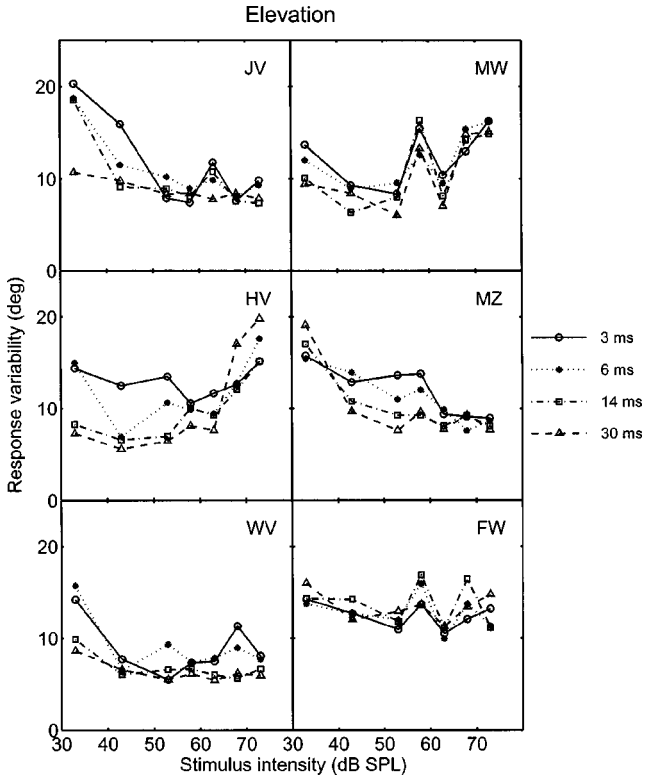


FIG. 7. Variability of elevation responses as a function of stimulus intensity. The different line styles and symbols correspond to the different stimulus durations. For most stimuli the variability is comparable; in contrast to the effects on response gain values, the variability does not change with recording session.

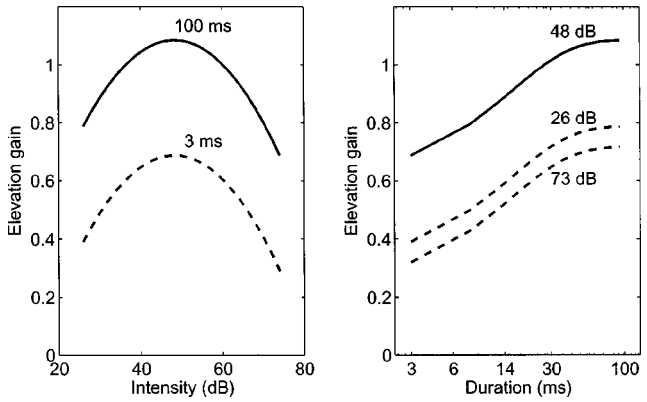


FIG. 8. Schematic summary of the results, presented in the same format as Fig. 2. Curves are based on the parameters of a nonlinear regression [Eq. (A2)] on the elevation gain data of all three sessions for subject JV. Note the logarithmic *x* axis.

Elevation gains varied in a nonmonotonic way with sound intensity (Figs. 5, 6). At low sound levels gains were low; they increased for intermediate sound levels (positive level effect), and decreased again for stimulus levels above about 55–65 dB SPL (negative level effect). When elevation gains are plotted as a function of sensation level, our results are in good agreement with the findings of Macpherson and Middlebrooks (2000).

In contrast to Macpherson and Middlebrooks (2000), however, our results indicate that both stimulus parameters affect the localization of sound-source elevation to a comparable degree (Table II). A possible reason for this difference might be that in the present study all stimulus conditions were randomly interleaved instead of presented in separate blocks of trials with fixed duration. As is illustrated, e.g., in Figs. 3 to 6 there can be considerable day-to-day variation in the absolute values of the obtained gains. Such a variability might potentially mask the effects.

This variability in our results between sessions could be due to simple day-to-day variation, or it could be the result of the differences in the intensity range used (33–63 vs 58–73 dB SPL).

Taken together, our results extend the findings of Hofman and Van Opstal (1998) and Macpherson and Middlebrooks (2000) and provide a more complete picture of the effect of sound duration and intensity on localization behavior. The data indicate that the negative level effect is not sufficient to account for the gain–duration relation which was found to persist for lower stimulus levels too.

We therefore propose that the gain–duration effect is indicative of a neural integration mechanism that accumulates evidence in order to "construct" its best estimate of sound-source elevation. As noted by Macpherson and Middlebrooks (2000), the negative level effect clearly does not fit into such a scheme, but rather provides support for the adaptation model. Note, however, that the consistent effects on elevation gain of other temporal stimulus parameters like sweep duration or interburst interval for long-duration (500 ms) stimuli at 70 dB SPL (Hofman and Van Opstal, 1998) are not readily explained by saturation of cochlear excitation patterns.

The conceptual neural-integration model put forward by Hofman and Van Opstal (1998) provides an explanation for the consistent finding that elevation *gain* is affected by the temporal stimulus parameters. In short, it proposes that the gain reflects the confidence level about the system's final estimate of sound-source elevation. This confidence is obtained by the internal correlation of the sensory spectrum [repeatedly sampled over short (<5 ms) time windows] with learned and stored representations of the subject's spectral cues, and subsequently averaged over a longer time window (several tens of ms). Clearly, this model should be extended to accommodate the level-dependent effects described in the present study.

In the absence of any certainty about stimulus location (e.g., due to low SNR), the default estimate might primarily rely on nonacoustic factors like prior knowledge about potential source locations. For example, in the current experiment this would be on average the straight-ahead location within the frontal hemifield. These factors may thus set the default gain of the internal estimate to zero, as well as an initial response bias (an average expected location). The actual response of the subject would thus be determined by a relative weighting of the prior expectation and the accumulated acoustic evidence for the veridical sound elevation. Idiosyncratic day-to-day variation of the weighting factor could underlie the intersession variability in observed gains.

It is straightforward to appreciate how the dynamic correlation model of Hofman and Van Opstal (1998) could be extended to incorporate the nonlinear influence of stimulus level (Fig. 9). At low stimulus levels and short durations the accumulated evidence remains low; hence, the response gain will be low too. Note that the observed gains were not zero

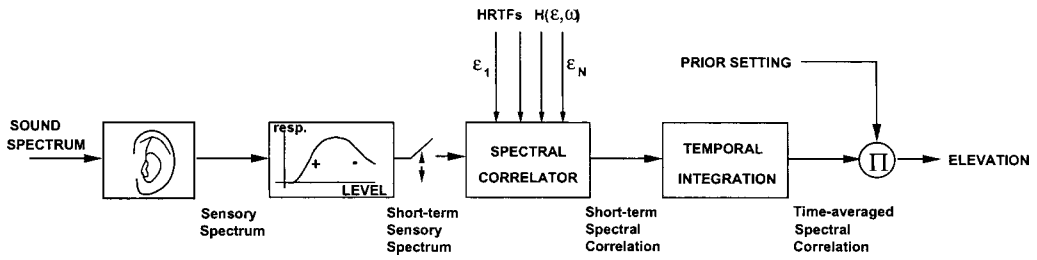


FIG. 9. Extension of the conceptual model of Hofman and Van Opstal (1998) in which the output of the short-term integration stage (which embodies a "multiple look" on the sensory spectrum over short (<5 ms) time windows) depends on sound level. The latter may be due to cochlear nonlinearities and/or neural tuning properties. Following the spectral correlation stage (comparison of the short-term sensory spectrum with stored HRTFs), a dynamic estimate of elevation is generated by averaging over a longer time window of several tens of ms. The output of this final stage is weighted against a preset default estimate, that may be based on prior expectation.

TABLE III. Partial regression coefficients for the multiple nonlinear regression on the data for all subjects and all stimulus conditions [Eq. (A2)].

Subject	Elevation gain				Bias			
	$\beta_1(\cdot 10^{-2})$	$\beta_2(\cdot 10^{-4})$	β_3	$\beta_4(\cdot 10^{-2})$	β_5	R^2	N	Session
JV	2.5	−2.3	0.46	3.3	−3.2	0.75	3195	1,2,3
FF	1.8	−1.8	0.52	8.6	1.8	0.79	1727	1
JM	1.0	−0.73	0.58	5.9	4.1	0.66	1616	1
JO	0.082	1.1	0.44	7.5	7.0	0.58	1005	1
MW	1.5	−0.99	0.35	7.1	9.7	0.71	1584	2,3
HV	2.1	−1.7	0.18	9.8	2.0	0.77	1585	2,3
MZ	1.3	−0.65	0.34	3.3	8.4	0.72	1562	2,3
WV	1.5	−1.2	0.31	8.8	8.6	0.82	1577	2,3
FW	0.72	−0.28	0.46	5.7	11.7	0.50	1592	2,3

for this condition, and that responses appeared to correlate well with the actual stimulus locations. Increasing the stimulus level will in turn improve the correlation, since the signal exceeds the critical SNR sooner. This effect would account for the positive level effect observed in our data. In the same vein, longer stimulus durations accumulate more and more evidence about the veridical sound elevation.

The nonlinear effect of stimulus level (positive and negative gain changes) reported in this paper could in principle be attributed to cochlear mechanisms (e.g., nonlinear amplification at low levels, and compression, or even clipping, at high levels). Alternatively, it might be due to central neural processing mechanisms (like neural saturation, or neural tuning to a specific optimal sound level; e.g., Ryan and Miller, 1978), or to both mechanisms. On the basis of the current experiments it is not possible to dissociate these possibilities.

ACKNOWLEDGMENTS

This research was supported by the Netherlands Organization for Scientific Research (NWO—MaGW, Project No. 410.203.01, JV) and the University of Nijmegen (AJVO). The authors would like to thank Dr. Armin Kohlrausch, Dr. Leslie Bernstein, Dr. Wes Grantham, and one anonymous reviewer for their valuable comments on an earlier version of this paper. We would like to acknowledge the valuable technical assistance of T. van Dreumel and H. Kleijnen for building the LED-speaker hemisphere. We thank Floor Franssen and Joost Maier for their help in carrying out the first series of experiments. We are also grateful to Dr. H. Versnel for setting up the data acquisition software. Finally, we thank our subjects for their time and effort.

APPENDIX:

In the nonlinear regression model of the elevation responses, the gain, G_ε was taken as a function of stimulus duration, D, and sound level, L

$$\varepsilon_R = G_\varepsilon(L,D)\cdot\varepsilon_T + b_\varepsilon . \tag{A1}$$

Based on the stimulus-specific linear regressions, plotted in Fig. 2, and the resulting gains, plotted in Figs. 4–6, the following function was chosen to capture the observed effects:

$$G_\varepsilon(L,D) = \beta_1\cdot L + \beta_2\cdot L^2 + \beta_3\cdot(1-e^{-\beta_4\cdot D}),$$

$$b_\varepsilon = \beta_5 , \tag{A2}$$

with L stimulus level (in dB SPL), D duration (in ms). Fit parameters $\beta_1-\beta_5$ were obtained by minimizing the mean-squared error between model and data. The resulting regression parameters for each subject are listed in Table III.

Note that $-\beta_1/(2\beta_2)$ provides an estimate of the stimulus level that yields the highest elevation gain. For the subjects in this study, this optimal sound level was typically between 50–70 dB SPL. The value of $1/\beta_4$ determines at which stimulus duration the elevation gain is estimated to reach 63% of its maximum value: this yields values between 10–30 ms.

Blauert, J. (**1996**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Frens, M. A., and Van Opstal, A. J. (**1995**). ''A quantitative study of auditory-evoked saccadic eye movements in two dimensions,'' Exp. Brain Res. **107**, 103–117.

Goossens, H., and Van Opstal, A. (**1997**). ''Human eye–head coordination in two dimensions under different sensorimotor conditions,'' Exp. Brain Res. **114**, 542–560.

Hartmann, W. M., and Rakerd, B. (**1993**). ''Auditory spectral discrimination and the localization of clicks in the sagittal plane,'' J. Acoust. Soc. Am. **94**, 2083–2092.

Hofman, P. M., and Van Opstal, A. J. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.

Hofman, P. M., and Van Opstal, A. J. (**2002**). ''Bayesian reconstruction of sound localization cues from responses to random spectra,'' Biol. Cybern. **86**, 305–316.

Knudsen, E. I., and Konishi, M. (**1979**). ''Mechanisms of sound localization in the barn owl (*Tyto alba*),'' J. Comp. Physiol. **133**, 13–21.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Macpherson, E. A., and Middlebrooks, J. C. (**2000**). ''Localization of brief sounds: Effects of level and background noise,'' J. Acoust. Soc. Am. **108**, 1834–1849.

Middlebrooks, J. C. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.

Oldfield, S. R., and Parker, S. P. (**1984**). ''Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions,'' Perception **13**, 581–600.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C* (Cambridge University Press, Cambridge, MA).

Ryan, A., and Miller, J. (**1978**). ''Single unit responses in the inferior colliculus of the awake and performing rhesus monkey,'' Exp. Brain Res. **32**, 389–407.

Wightman, F. L., and Kistler, D. J. (**1989**). ''Headphone simulation of free field listening. I Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.

Infinite-impulse-response models of the head-related transfer function

Abhijit Kulkarni[a] and H. Steven Colburn[b]
Hearing Research Center and Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

(Received 24 September 2001; revised 19 December 2003; accepted 30 December 2003)

Head-related transfer functions (HRTFs) measured from human subjects were approximated using infinite-impulse-response (IIR) filter models. Models were restricted to rational transfer functions (plus simple delays) so that specific models are characterized by the locations of poles and zeros in the complex plane. The *all-pole* case (with no nontrivial zeros) is treated first using the theory of linear prediction. Then the general *pole-zero* model is derived using a weighted-least-squares (WLS) formulation of the modified least-squares problem proposed by Kalman (1958). Both estimation algorithms are based on solutions of sets of *linear* equations and result in efficient computational schemes to find low-order model HRTFs. The validity of each of these two low-order models was assessed in psychophysical experiments. Specifically, a four-interval, two-alternative, forced-choice paradigm was used to test the discriminability of virtual stimuli constructed from empirical and model HRTFs for corresponding locations. For these experiments, the stimuli were 80 ms, noise tokens generated from a wideband noise generator. Results show that sounds synthesized through model HRTFs were indistinguishable from sounds synthesized from original HRTF measurements for the majority of positions tested. The advantages of the techniques described here are the computational efficiencies achieved for low-order IIR models. Properties of the all-pole and pole-zero estimators are discussed in the context of low-order HRTF representations, and implications for basic and applied contexts are considered.
[DOI: 10.1121/1.1650332]

PACS numbers: 43.66.Qp, 43.64.Ha, 43.66.Pn [LRB] Pages: 1714–1728

I. INTRODUCTION

The acoustical transfer function from a free-field source to a listener's eardrum embodies all of the phase and amplitude effects associated with the location of both the source and receiver (e.g., the diffraction effects caused by the presence of the body, head, and pinna of the listener). This transformation, which is generally assumed to be linear, can be described either in the frequency domain, in which case it is referred to as the head-related transfer function (HRTF), or in the time domain as an impulse response, in which case it is called the head-related impulse response (HRIR). These alternative representations are related by the Fourier transform. The source-to-eardrum transformation, however it is represented, is a major determinant of perceived sound source location, and the efficient characterization of this transformation is the focus of this paper. Although it has been demonstrated that specific features in the HRTF carry information about sound source location, (Blauert, 1996; Middlebrooks, 1992), conclusive evidence on the exact nature of these features and their interpretation by the auditory system has not been reported; thus, it is not clear what the most efficient representation would be.

Most measurements of the HRTF consist of time–domain measurements of the HRIR, either directly as the response to an acoustic pulse or indirectly by deconvolving the average response to a wideband input such as a Golay code sequence or a maximum-length sequence. The result of these measurements is generally a finite length approximation to the HRIR. This can be thought of as a finite-impulse-response (FIR) filter response or as a finite-length approximation to the impulse response of an infinite-impulse-response (IIR) filter. In either case, the empirical HRTF is computed as the Fourier transform of the measured (finite-length) HRIR.

HRTFs, as empirically measured, have complicated characteristics with multiple peaks and notches in their magnitude spectra and several approaches have been suggested for efficient descriptions. The significance of specific characteristics of the HRTF for sound localization has been explored to some extent by combinations of HRTF modeling and experiments. For example, Martens (1987) used an application of principal-components analysis (PCA) to model HRTFs and reported a systematic variation of model parameters for azimuthal source positions. Kistler and Wightman (1993) extended the use of the PCA approach and, in an analysis that included the psychophysical validation of the model-reconstructed HRTFs, successfully demonstrated a technique for low-order HRTF representation. Kulkarni and his colleagues also explored the adequacy of simplifications in the magnitude spectrum (Kulkarni and Colburn, 1998) and in the phase spectrum (Kulkarni *et al.*, 1999) with psychophysical experiments. In a study more closely related to the

[a]Current address: Bose Corporation, MS 15D, The Mountain, Framingham, Massachusetts 01701.
[b]Correspondence to Steve Colburn at Department of Biomedical Engineering, 44 Cummington Street, Boston, Massachusetts 02251. Electronic mail: colburn@bu.edu

present one, Asano *et al.* (1990) used IIR models with different degrees of simplification to investigate the role of microscopic and macroscopic patterns in HRTFs for median plane localization. Elegant IIR modeling methods for HRTFs have also been reported by Blommer and Wakefield (1994). The current study further investigates the use of low-order IIR models to describe HRTFs with a focus on efficient (linear) computational methods.

One of the objectives of this study was to explore the effectiveness of low-order IIR models in preserving directional information. There are several motivations to our approach. First, a parsimonious description of the HRTF can have practical benefits in virtual auditory display applications. The use of HRTFs in the generation of virtual stimuli for presentation over headphones has been efficiently implemented by several laboratories (e.g., Wightman and Kistler, 1989a, 1989b). In these implementations, sound stimuli are convolved with the FIR impulse responses measured in the ear canal for a source at the position being simulated. Low-order IIR implementations of the HRTF have a significant computational advantage over their FIR equivalents. In spite of the evolution of signal processing hardware which makes FIR filtering a reasonable choice, this computational advantage cannot be ignored, particularly when the virtual display is used to render multiple sound sources and complicated, reverberant acoustical environments. Second, a perceptually appropriate low-order representation of the HRTF may provide insight into sound localization mechanisms. The usefulness of various cues embodied in the HRTF is incompletely understood, and identifying an appropriate simple representation of the HRTF can be used to study attributes in the ear-input signal that lead to directional perceptions. Finally, an appropriate low-order model can be used to study the physical mechanisms resulting in features in the HRTF (e.g., Algazi *et al.*, 2002). Because these features are idiosyncratic to individuals (e.g., Wightman and Kistler, 1989a), gaining this insight could result in computational methods for generating HRTFs that would not rely upon making empirical measurements from individuals. It should be noted that making such measurements is both tedious and requires specialized equipment and facility (e.g., Wightman and Kistler, 1989a).

In this paper, we demonstrate techniques that can be used to develop IIR models of any desired order. Further, we validate the models for two low-order conditions with psychophysical measurements. Our approach to validation is to investigate the discriminability of sounds synthesized through an empirically measured HRTF for a given position with that from a model of the HRTF. Note that the discrimination measure is a complete test of model implementation because it requires the model to render both the directional and nondirectional attributes of a source indistinguishably from its empirical characterization.

The remainder of this paper is organized as follows. In Sec. II we describe two IIR model architectures. Details of the model optimizations are provided in the Appendix. In Sec. III, the psychophysical evaluation of the modeled HRTFs is presented. The results presented here suggest that IIR models can provide efficient descriptions of HRTFs with a very small numbers of parameters.

II. IIR MODELS OF HRTFs

The problem of IIR filter modeling can be considered as a rational transfer function approximation problem, which has generated a class of solutions used to obtain parametric models of random processes (c.f. Kay, 1988). Within such a framework, the set of HRTFs for all possible directions are thought of as analogous to a random process. The sample wave forms of such a random process are functions of frequency and the underlying sample space is defined by the set of directions with respect to the head of the receiving listener. As a result of individual differences (e.g., in head dimensions and pinna geometry), this random process may vary across individuals and acoustical environments. The problem posed in this paper is for a single listener in an anechoic environment; however, the modeling techniques suggested are not restrictive and may be used for a general listener in any given acoustical environment.

A. Model architectures

A general review of discrete-time rational transfer function models is given in Appendix A. Two specific model architectures are investigated for determining an estimated $\hat{H}(e^{J\omega})$ for a given HRTF measured/empirical $H(e^{J\omega})$. The *all-pole* model has the form

$$\hat{H}(e^{J\omega}) = \frac{b[0]}{1+\sum_{k=1}^{p} a[k]e^{-J\omega k}}, \tag{1}$$

and the problem is to find the p coefficients $a[k]$ and the scale factor $b[0]$ that allow $\hat{H}(e^{J\omega})$ to best approximate $H(e^{J\omega})$.

The *pole-zero* model has the form

$$\hat{H}(e^{J\omega}) = \frac{b[0]+\sum_{k=1}^{q} b[k]e^{-J\omega k}}{1+\sum_{k=1}^{p} a[k]e^{-J\omega k}}, \tag{2}$$

and the problem is to find the $q+1$ coefficients $b[k]$ and p coefficients $a[k]$ to best approximate $H(e^{J\omega})$.

B. Model parameter estimation

The goal of the model fitting is to derive a fit to the perceptually salient features in the HRTF set using a minimal number of model coefficients. Ideally, therefore, it would be desirable to have a psychoacoustically motivated criterion to guide the model fitting algorithm. Additionally, the fitting algorithm should be computationally tractable, easy to implement, and robust in fits across various data. Given the incomplete understanding of human hearing abilities, an objective perception-based metric is hard to define and elusive. Consequently, our model fitting approaches were chosen to have well-behaved signal-processing and numerical properties and model evaluation was left to listening experiments. The question of ''optimal'' model order is not considered in the current study. We recognize, however, that depending on the application, this issue is clearly relevant.

The model parameter estimation methods used in this study are outlined below. Additional background and details of the methods are provided in Appendix A.

The coefficients $a[k]$ for the *all-pole* model were estimated using the autocorrelation method for linear prediction (Makhoul, 1975; Papoulis and Pillai, 2002). This method requires the solving of equations written in terms of the autocorrelation function $r_{hh}[k]$ of the m-point measured impulse response $h[n]$, where the autocorrelation function for the real-valued (and causal) $h[n]$ is defined by $r_{hh}[k]=\Sigma_{n=0}^{+\infty}h[n]h[n-k]$. Specifically, the following set of linear equations is solved for $a[k]$,

$$r_{hh}[k]=\begin{cases}-\sum_{l=1}^{p} a[l]r_{hh}[k-l]; & \text{for } k\geqslant 1;\\ -\sum_{l=1}^{p} a[l]r_{hh}[-l]+b[0]h[0] & \text{for } k=0.\end{cases} \tag{3}$$

Note that, in terms of the impulse response $h[n]$, the Yule–Walker equations are nonlinear, but in terms of the autocorrelation function computed from the measurements of $h[n]$, the equations are linear. For our approximation problem with $p<m$, the set of $m+1$ linear equations is overdetermined, and a simple solution is not possible. For the calculations below, the Levinson algorithm (Levinson, 1947; Kay, 1988) was used to solve this system of linear equations for several different model orders (i.e., for different values of p).

A weighted-least-squares approach is used to estimate the *pole-zero* model coefficients $b[k]$ and $a[k]$. As described in more detail in Appendix A.2, our approach was to minimize the error between the model $\hat{H}(e^{J\omega})$ and the original spectrum $H(e^{J\omega})$ on a decibel scale. This ensures that the solution is not biased toward spectral peaks (regions of high spectral energy) relative to valleys (regions with low energy). Given that it has been demonstrated that information in the form of both peaks (Blauert, 1996) and valleys (Musicant and Butler, 1985) is important for information about the sound source location, we consider the error metric on a decibel scale to be more relevant than an error metric utilizing absolute error on a linear scale. Specifically, we seek an iterative solution in terms of $A_i(e^{J\omega})$ and $B_i(e^{J\omega})$, which are defined as

$$A_i(e^{J\omega})=\sum_{k=0}^{p} a_i[k]e^{-J\omega k}$$

$$=\text{Fourier transform of the model coefficients } a_i[k],$$

$$B_i(e^{J\omega})=\sum_{k=0}^{q} b_i[k]e^{-J\omega k}$$

$$=\text{Fourier transform of the model coefficients } b_i[k],$$

and the model coefficients converge to the final IIR approximation. As outlined in Appendix A.2, the problem was reformulated to seek a solution that minimizes the following equation:

$$\hat{C}_i=\frac{1}{2\pi}\int_{-\pi}^{\pi}|W_i(e^{J\omega})|^2\left|\frac{H(e^{J\omega})A_i(e^{J\omega})}{A_{i-1}(e^{J\omega})}-\frac{B_i(e^{J\omega})}{A_{i-1}(e^{J\omega})}\right|^2 d\omega, \tag{4}$$

where the weighting functions $W_i(e^{J\omega})$ and the $A_{i-1}(e^{J\omega})$ are calculated from results of the previous iteration, as described below. [Note, as the iterations converge and $(A_i\rightarrow A_{i-1})$, the error becomes a weighted average of $H(e^{J\omega})-B_i(e^{J\omega})/A_i(e^{J\omega})$.] In this analysis, as explained more fully in Appendix A.2, the error metric above is defined in terms of the weighting function,

$$W_i(e^{J\omega})=\frac{\left[\log|H(e^{J\omega})|-\log\left|\frac{B_{i-1}(e^{J\omega})}{A_{i-1}(e^{J\omega})}\right|\right]^2}{|H(e^{J\omega})A_{i-1}(e^{J\omega})-B_{i-1}(e^{J\omega})|^2}. \tag{5}$$

(In addition to this weighting function, we restricted the frequencies included in the error analysis to frequencies relevant to human listeners. Specifically, only frequencies between 300 and 15 000 Hz were considered by applying a weighting function that had a value of unity in this frequency range and a value of zero outside this range.) As the iterations converge, the error metric converges to a measure of the decibel difference in the magnitude of the frequency response.

In order to achieve rapid solutions, the error equation is rewritten in the time domain, in a form that is equivalent to the regression equation, which allows for a fast and easy solution. First, we define the simplifying notation

$$Y_i(e^{J\omega})=W_i(e^{J\omega})\frac{H(e^{J\omega})}{A_{i-1}(e^{J\omega})} \tag{6}$$

and

$$X_i(e^{J\omega})=W_i(e^{J\omega})\frac{1}{A_{i-1}(e^{J\omega})}. \tag{7}$$

With this notation, the error can be rewritten as minimizing $\int|E(e^{J\omega})|^2 d\omega/2\pi$, where

$$E(e^{J\omega})=Y_i(e^{J\omega})A_i(e^{J\omega})-X_i(e^{J\omega})B_i(e^{J\omega}).$$

This equation can now be transformed to the time domain to give the following linear function, in terms of the $p+q+1$ variables $\{a_i[1],a_i[2],\ldots,a_i[p],b_i[0],b_i[1],\ldots,b_i[q]\}$, that determines the ith iteration of the filter,

$$e_i[j]=\sum_{k=1}^{p} a_i[k]y_i[j-k]+y_i[j]-\sum_{k=0}^{q} b_i[k]x_i[j-k]. \tag{8}$$

A minimum-squared-error is achieved by the minimization of $\Sigma_{j=0}^{N}(e_i[j])^2$, which is equivalent to optimizing the regression equation, and the solution for each iteration is very simple.

The number of poles and zeros (p,q) used for a given minimization was varied systematically to obtain different low-order model representations of a HRTF. At the start of each minimization procedure $(i=0)$, coefficients of $A(e^{J\omega})$ were initialized to the p all-pole coefficients obtained using the Levinson algorithm discussed in the previous section and

the coefficients of $B(e^{J\omega})$ were initialized to the first $q+1$ taps of the FIR filter being modeled. The procedure then typically terminated after four iterations of Eq. (4) to converge to a near-optimal set of coefficients for that given model order. In the final implementation, we used a total of six iterations to obtain the model coefficients.

C. HRTF data modeled

HRTFs measured from 47 human listeners were fit using the techniques described above. The human HRTF measurements were obtained from Dr. Fred Wightman at the University of Wisconsin and from the CIPIC database available from the University of California at Davis (Algazi *et al.*, 2001). The Wisconsin data used here correspond to that from the subject SDO reported in a previously published study (Wightman and Kistler, 1989a). The second set of Wisconsin data (SOW), also obtained at the University of Wisconsin anechoic chamber facility, was measured from one of the authors (AK) of this study.

The SDO dataset consists of 144 measurements from different source locations. Specifically there are six elevation measurements, (every 18° from +54 to −36°) at each of 24 azimuths (sampled every 15° starting from directly behind). The resulting impulse responses are 512-tap FIR filters at a sampling rate of 50 kHz. Details of the measurement protocol have been reported in Wightman and Kistler (1989a). The SOW dataset consists of 505 measurements. There are 14 elevation measurements (every 10° from +80 to −50°) at each of 36 azimuths (sampled every 10°, starting from directly behind), plus a measurement from directly above (+90°). The resulting impulse responses are 512-tap FIR filters measured at a sampling rate of 100 kHz. For the purposes of this study we downsampled these data, giving an equivalent 256-tap filter at 50 kHz and zero-padded it to 512 taps.

Finally, the CIPIC data consists of measurements from 45 listeners. There are 25 azimuths from −90 to +90°, where +90° is the location directly aligned with the left ear and −90° is the location directly aligned with the right ear. There are 50 elevations at each azimuth, from −45° (below in front) to 230° (below and behind) the listener. The measurements are available as 200-tap FIR filters at a sampling rate of 44.1 kHz. Details of the measurement protocol used to obtain these measurements have been reported in Algazi *et al.* (2001).

D. Model evaluation

Models for the HRTFs were obtained by assuming that they are minimum-phase functions. Given that any system function can be expressed as a combination of a minimum-phase function and an all-pass function (Oppenheim and Schafer, 1975),

$$H(e^{J\omega}) = H_{\min}(e^{J\omega}) H_{ap}(e^{J\omega}), \tag{9}$$

we assumed that the HRTF is a minimum-phase function and that the all-pass component is a position-dependent but frequency-independent time delay for that position. This minimum-phase assumption was found to be adequate in an independent study (Kulkarni *et al.*, 1999).

Forcing the IIR filter models to be minimum phase is advantageous because we only need to consider a fit to the HRTF magnitude function, the accompanying Hilbert phase being unique for a minimum-phase function. During the psychophysical testing of the models the appropriate interaural delay was introduced in the signals. This is discussed further in Sec. III C.

For each set of HRTFs, the mean log-magnitude functions of the measurements at the right and left ear, respectively, were computed. These mean functions include the direction-independent spectral features shared by all measurements at a given ear (e.g., the ear canal resonance at about 2.5 kHz and artifacts resulting from the $\frac{1}{4}$-wavelength notch in the measurement caused by standing waves, etc.). The mean functions were subtracted (in dB) from the log-magnitude function of each measurement at the appropriate ear. We refer to the residual log-magnitude functions as the *directional transfer functions* or DTFs after Kistler and Wightman (1991). As noted in that study, DTFs represent primarily direction-dependent spectral effects. The DTFs were then converted back to a linear scale before fitting them with the modeling protocols.

The primary results reported in this study are model fits to DTFs. As noted above, the HRTF is composed of the DTF and a common, direction-independent, transfer function. The models can, of course, be used to fit the whole HRTF and we do provide some results of fitting the complete HRTF when comparing our results with some previous studies. The DTFs are particularly interesting because they embody the direction-dependent features of the HRTF. Furthermore, for a virtual display implementation, it is best to implement the common transfer function as a fixed, position-invariant, filter in series with the variable DTF filter. The DTF, which would be a lower-order filter compared to a fit to the whole HRTF, can then be changed to render sound for particular directions. Having the variable filter be of low-order minimizes the storage requirements in virtual displays and also makes it easier to implement dynamic events like moving sound sources and switching source directions in general.

The goodness of fit can be quantitatively expressed in many ways. Almost all forms of this measure, however, are not completely satisfying from a psychoacoustical perspective and several forms were used in the preliminary evaluation of the pattern of errors from model approximations. All forms of error criteria we used resulted in very similar error distributions and we finally chose to express the error between the actual and fitted log-magnitude DTFs using the following expression:

$$E = 20 \left| \int_{-\pi}^{\pi} (\log_{10}|H(e^{J\omega})| - \log_{10}|\hat{H}(e^{J\omega})|)^2 \frac{d\omega}{2\pi} \right|^{1/2}. \tag{10}$$

In the above equation the notation $H(e^{J\omega})$ refers to the DTF and the notation $\hat{H}(e^{J\omega})$ refers to the model fit to the DTF. This error measure was used as a relative means to judge the accuracy of fits within and across HRTF datasets.

A large number of model orders, for both the all-pole and the pole-zero architectures, were simulated. We present

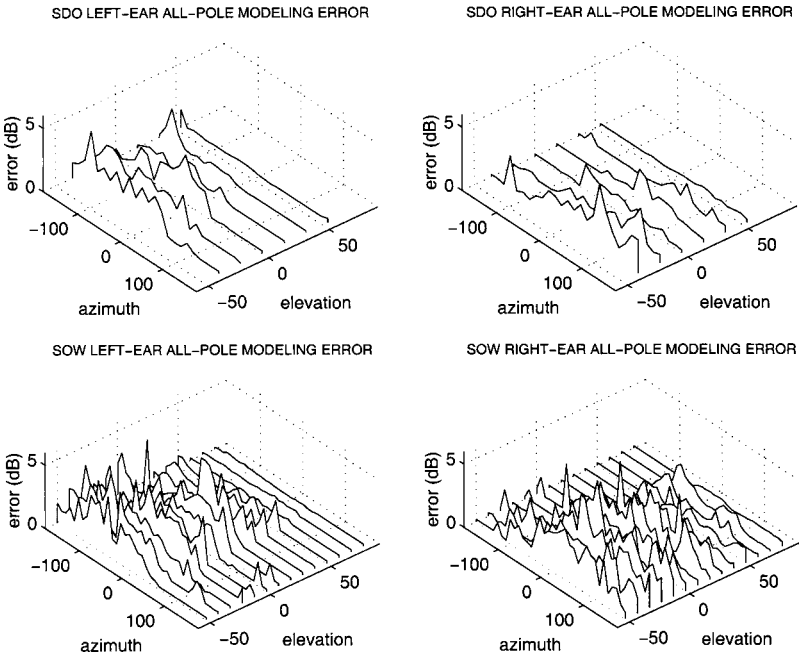


FIG. 1. Error distributions as a function of the source position for 25-pole DTF models for the SDO and SOW datasets (top and bottom panels), respectively. The left and right columns correspond to the left and right ear results. See the text for details of each dataset.

results only for the 25-pole (all-pole) model and the 6-pole, 6-zero (pole-zero) model. These cases were chosen as a reasonable compromise between the number of parameters and the resulting errors, both objectively calculated and subjectively perceived.

E. Results and discussion

The errors in fitting the human HRTF measurements were analyzed at each of the measured HRTF positions. Note that, following the minimum-phase assumption for HRTF phase, only fits to the magnitude spectra were analyzed. The distribution of errors across the position obtained from such an analysis, for the all-pole model with 25 poles, is presented in Fig. 1 for left and right ears of the SDO and SOW datasets. The corresponding results for a pole-zero model with six poles and six zeros are presented in Fig. 2. Note that a source located at $+90°$ azimuth is directly across from the left ear (i.e., the right ear is the shadowed ear) and a source located at $-90°$ is directly across from the right ear (i.e., the left ear is the shadowed ear). Example fits to DTFs for a set of azimuths at zero elevation from the SDO HRTF dataset, using both the all-pole model with 25 poles (dotted lines) and the pole-zero model with six poles and six zeros (dashed lines), are shown in Fig. 3. Fits to DTFs, using the same models, from six median plane elevations (zero degrees azimuth) from the same dataset are presented in Fig. 4. The calculated error in the model fits is indicated in each panel. Several observations are noteworthy. First, in general, the pole-zero model with six-poles and six-zeros provides good fits to DTFs. Predictably, fits arising from the all-pole model

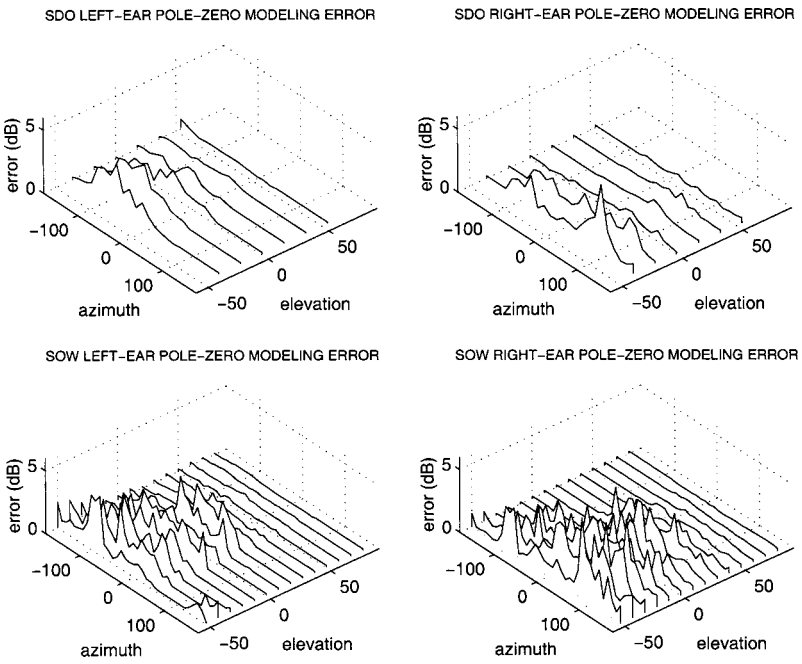


FIG. 2. Error distributions as a function of the source position for six-pole, six-zero DTF models for the SDO and SOW datasets (top and bottom panels), respectively. The left and right columns correspond to left and right ear results. See the text for details of each dataset.

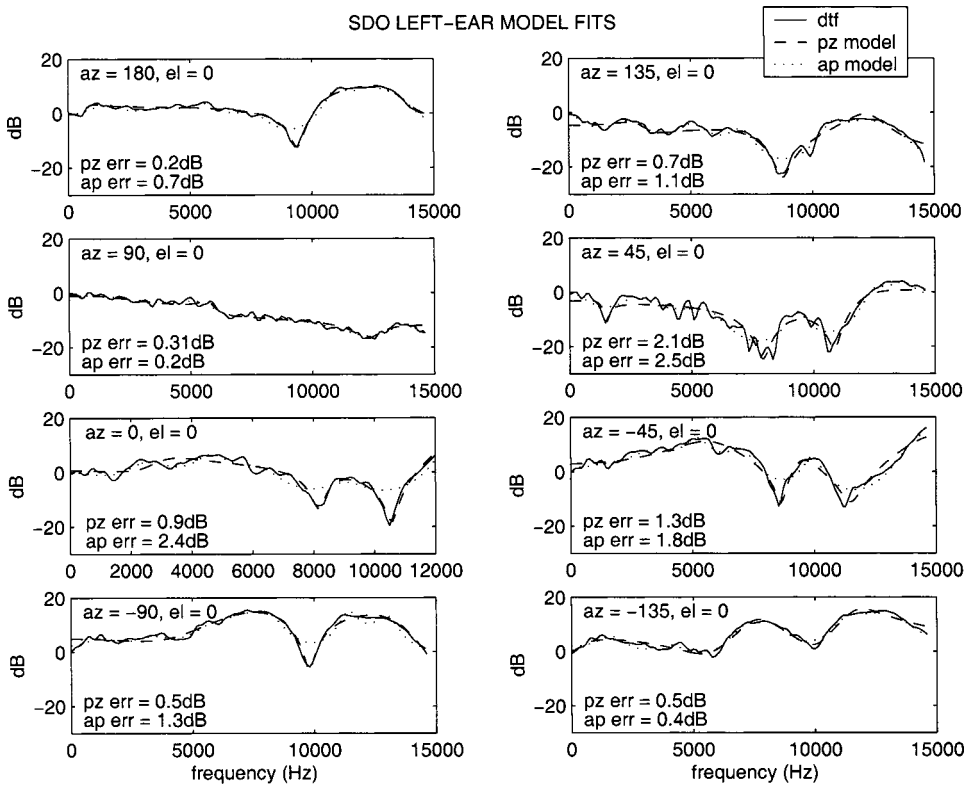


FIG. 3. Example fits to azimuthal (horizontal plane) DTFs (solid lines) from the SDO dataset using an all-pole (ap) model with 25 poles (dotted line) and a pole-zero (pz) model with six poles and six zeros (dashed lines). The empirical measurements are for every 45°, starting from directly behind. The error resulting from the fits is indicated in each panel for each model.

seem to better approximate peaks than troughs in the DTFs, whereas the pole-zero model reconstructions provide good approximations to both the peaks and the troughs in the DTFs. Second, comparing results shown in Figs. 1 and 2 reveals that errors in model reconstructions from both models are systematic, being maximum at the low source elevations (below 0°) for the ear farther away from the source and being minimum at all high source elevations. This may be because of a systematic trend in the complexity of the DTF, which arises from acoustical diffraction and scattering of the sound signal from the head and torso for different source locations. At low source elevations, at the ear shadowed from the source, the head and torso, in addition to the pinna, result in complex interactions in the sound waves reaching the far

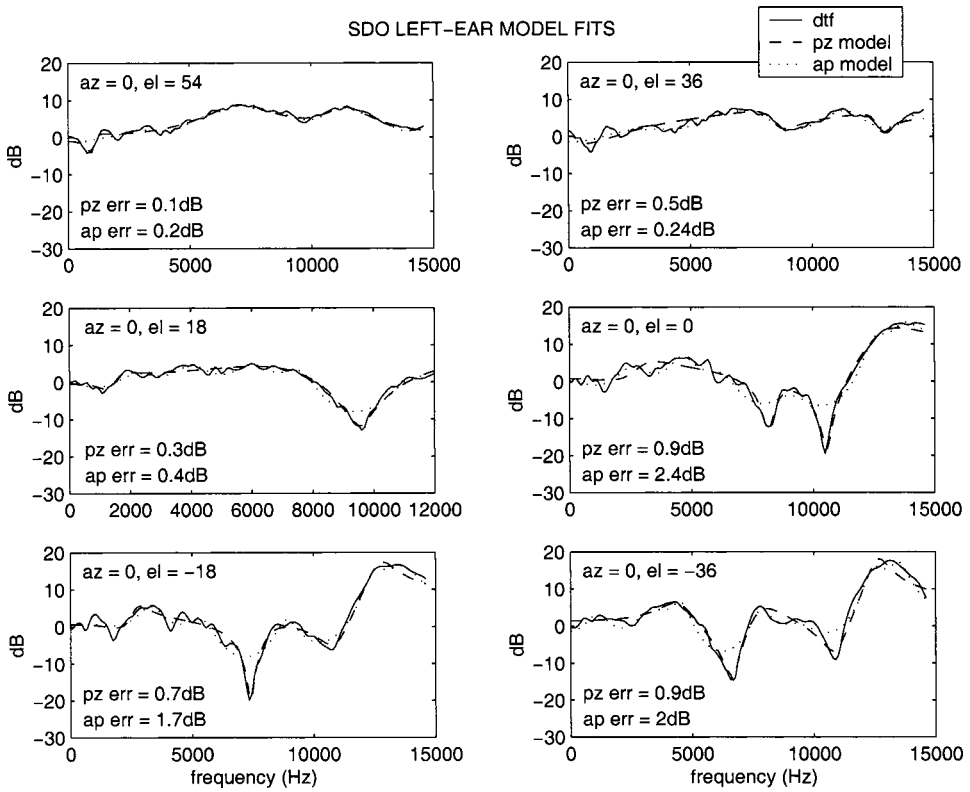


FIG. 4. Example fits to median plane DTFs (solid lines) from the SDO dataset using an all-pole (ap) model with 25 poles (dotted line) and a pole-zero (pz) model with six poles and six zeros (dashed lines). The empirical measurements are for every 18° between +54° and −36° in front. The error resulting from the fits is indicated in each panel for each model.

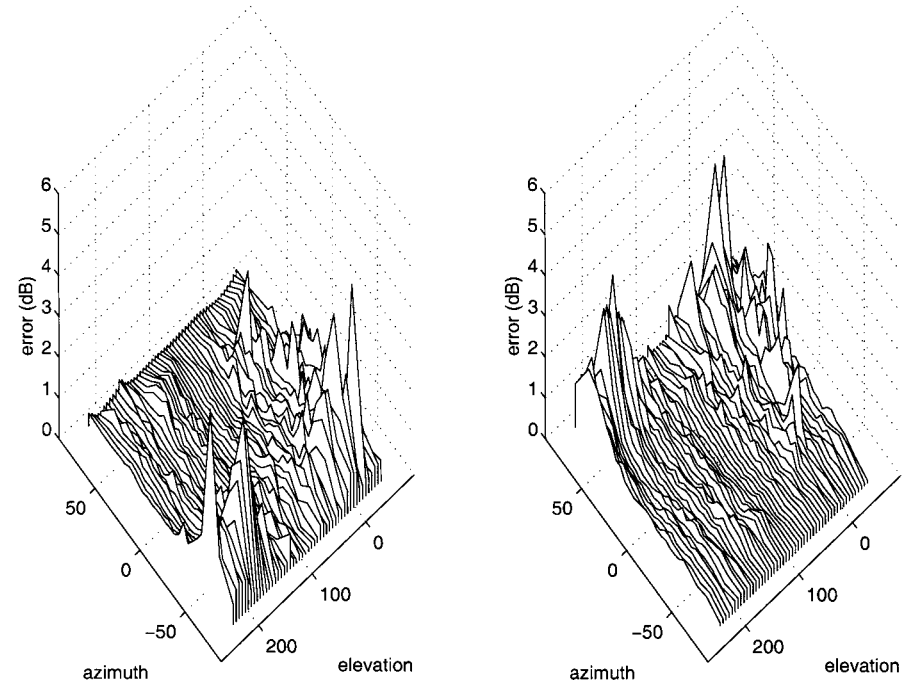


FIG. 5. Error distributions as a function of the source position for a six-pole, six-zero DTF model for SUBJECT 003 from the CIPIC HRTF database. The left and right columns correspond to the left and right ear results. See the text for details of each dataset.

ear. Also, the signal to noise ratio in the measurement is poorer at these locations. Hence, as may be noted from the error distributions, the error for the left and right ear fit is maximum for source locations where the ear is maximally shadowed. At high elevations, the acoustic path to either ear is more direct and includes primarily the pinna and the top portion of the head. The DTF at high elevations is smoother and better approximated by the models. This is consistent with there being a low (and fairly constant) error at the high elevations for all source azimuths.

The results from modeling data from the CIPIC database are very similar to those from the SDO and SOW datasets described above. As an example, the errors resulting from fitting a pole-zero model to subject 003 from the CIPIC database are shown in Fig. 5. The coordinate system for source location covers 25 azimuths from -90 to $+90°$, where $+90°$ is the location directly across the left ear and $-90°$ is the location directly across the right ear. There are 50 elevations at each azimuth, from $-45°$ (below in front) to 230° (below and behind) the listener. The location directly overhead has an elevation of 90° and the location directly in the back has an elevation of 180°. Note that, as with the SDO and SOW data, the errors in model reconstruction are worse at the low source elevations (below 0° and greater than 180°) for the ear maximally shadowed from the source. All the remaining 44 HRTF sets from the CIPIC database showed very similar error trends to the data shown in Fig. 5. The maximum error at any given location across all the CIPIC sets was 6.6 dB compared to 6.4 dB in the SOW set. The quality of fits across the CIPIC datasets was comparable to the fits shown in Figs. 3 and 4 for the all-pole and pole-zero models. Because of the large amounts of data involved, the corresponding results from the CIPIC database are not presented here.

The quality of the fits with the pole-zero model was further investigated by looking at the worst case fits in all the HRTF datasets. Again, because these observations can be generalized across all sets, only data for the SDO set are presented. The worst case fits with the pole-zero model for the SDO set are shown in the top panels of Fig. 6. The worst case fits occur at an azimuth of 105° and an elevation of $-36°$ for the left ear and at an azimuth of $-90°$ and an elevation of $-36°$ for the right ear. The errors corresponding to the fits are 3.5 and 5.3 dB, respectively.

Note that, consistent with the observations above, these worst-case fits occur for source locations where the ear is significantly shadowed from the source and the DTFs show considerable complexity and detail. While the model does a reasonable job of fitting the complicated spectrum for these source locations, it can be argued that it does not capture all the complexity in HRTF and hence might perceptually compromise the spatial quality of sounds rendered with the model. One choice would be to increase the numbers of poles and zeros in the model so that a better match could be obtained; however, a number of studies (Humanski and Butler, 1988; Morimoto, 2001; Macpherson and Middlebrooks, 2002; Wightman and Kistler, 1999) suggest that monaural spectral features in the far ear do not contribute significantly to spatial perception of the source. Specifically, Morimoto (2001) has pointed out that the contribution of the spectrum at the far ear to the localization of sources located more than 60° from the midline is not perceptually measurable. These studies also point out that the interaural level spectrum has little influence on spatial judgments. We believe, therefore, that the spectral detail seen in shadowed ear spectra (source azimuths exceeding 60° from the midline) are not perceptually relevant and that the smooth fit provided by the model (the top panels of Fig. 6) is psychoacoustically adequate. This question is explicitly addressed in Sec. III. With this in mind, we conducted another analysis of model fits, where the directions of interest were restricted so that far-ear azimuths of more than 60° from the midline were ignored. The worst-case fits obtained when the far-ear azimuths are constrained

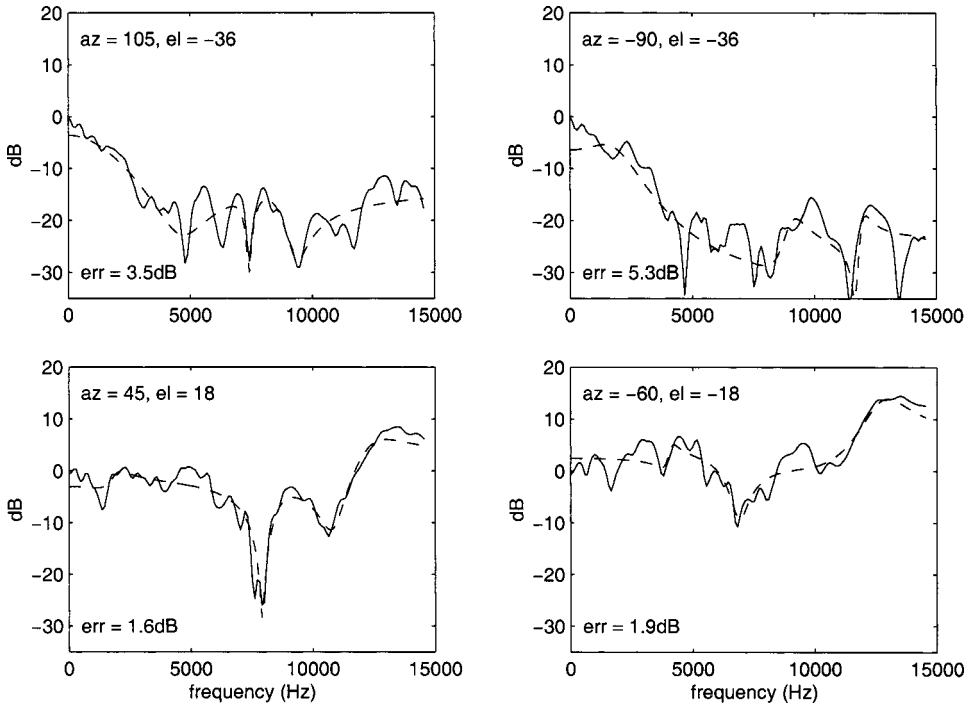


FIG. 6. Worst case fits to DTFs from the SDO data using a six-pole, six-zero model. The top panel shows the worst fits using the entire set and the bottom panel shows the worst fits when the set is constrained to exclude azimuths exceeding 60° from the midline for the shadowed ear. Left and right ear results are shown in the two columns and the modeling error is indicated in each panel.

to exclude azimuths exceeding 60° from the midline are shown in the lower panels of Fig. 6. The worst mean-square-error is now 1.6 and 1.9 dB for the left and right ear, respectively, and occurs for sources located at an azimuth of 45° and an elevation of 18° for the left ear and at an azimuth of −60° and an elevation of −18° for the right ear. Note that the model fits now appear to appropriately capture detail in the raw spectra in the worst cases in the restricted set.

The above analysis was carried out for all of the HRTF sets. The general nature of the complexity and the goodness of fit shown in Fig. 6 is consistent for all of the HRTF sets at which we looked. The exact location corresponding to the worst case fits was different for each set, although it always corresponded to a low-elevation from the range of azimuths that would correspond to the most shadowing of the source to the ear (−75° to −105° for the left ear and +75° to +105° for the right ear). The nature of errors seen when the HRTF set is meaningfully constrained to exclude some of the far ear locations is also consistent across the sets. That is, within the region where spectral detail is likely to be important, the model appears to capture the relevant detail in the DTF magnitude spectrum. Finally, the magnitude and range of fitting errors was comparable across the HRTF sets, so that we think that the fits shown in Figs. 3 and 4 are truly representative.

The model fits obtained above indicate that the model is capable of matching two or three significant features (i.e., peaks or troughs) in the DTF. In our experience with the numerous HRTF sets, excluding regions where the far ear is significantly shadowed, a model composed of six poles and six zeros adequately characterizes the DTF, although the model fits improve with an increasing order of the filters (the number of poles and zeros). A higher-order model can easily be chosen if the discrepancy in the model output is considered too large. In the absence of objective psychoacoustical criteria, the level of spectral detail in the HRTF that needs to be accurately matched by the models is difficult to assess. The appropriateness of the model fits are therefore further assessed via listening experiments below.

F. Interpretation of pole-zero placements

Results from the pole-zero model may be analyzed by imagining the positions of the poles and zeros in the complex z plane. Viewed in this way, it is apparent that the model has only six complex free parameters. Because we are dealing with a real system, complex poles and zeros occur in a conjugate-pairs pattern to provide a symmetry about the real axis. Hence, for any given position we need only to keep track of the real axis and any one half of the z plane. It may be possible to relate the peaks and valleys observed in the HRTF to specific pole-zero positions and understand their physical significance. Observing the trajectories of the poles and zeros with a changing source position could then provide insight into the physics of the filtering action performed by the pinna. We have not attempted to perform such an analysis. Note that the actual placement of the poles and zeros depends upon the order of the prescribed fit so that a complete analysis would require knowledge of the exact order of the physical system.

The filtering action of the torso, head, and pinna may be equivalently viewed as constituting a multielement resonance system. The direction-dependent resonators may be derived by decomposing the pole-zero representation of the DTF into second-order subsections. Using partial-fraction expansion techniques, the general pole-zero form of Eq. (2) may be expressed as

$$H(z)=\sum_{k=0}^{Nd} C_k z^{-k}+\sum_{k=0}^{Nr}\frac{b_{0k}+b_{1k}z^{-1}}{1-a_{1k}z^{-1}-a_{2k}z^{-2}}, \qquad (11)$$

where $z=e^{J\omega}$. This system has N_r+1 second-order resonances that depend on source location. These DTF resona-

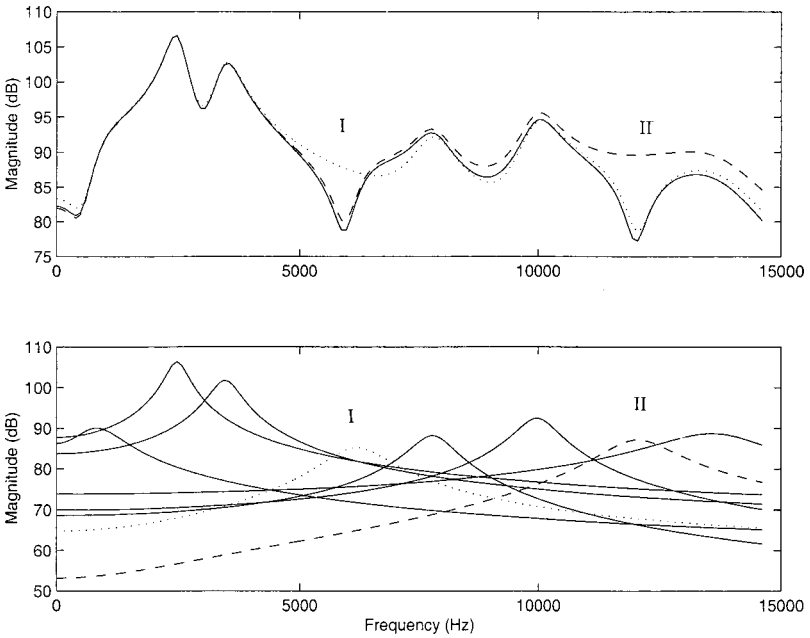


FIG. 7. An example eight-resonance fit (solid line) to a complete HRTF is plotted in the top panel and the corresponding eight resonances are shown in the bottom panel. The notches in the HRTF marked I and II in the top panel can be selectively removed (dotted and dashed lines, respectively) by not including the resonances marked I and II (and plotted with dotted and dashed lines) from the lower panel.

tors, combined with sections describing the acoustical effects independent of the source direction, describe the HRTF. The direction-independent effects, which include, for example, the transfer function of the ear canal and the concha cavum, are approximated by the mean log-magnitude function. The description of the system in Eq. (11) comprises $Nd+1$ simple scaled delay paths and $Nr+1$ second-order IIR resonators. Accordingly, the six-pole, six-zero model of the DTF results in a description comprising three second-order resonators ($Nr=2$) and a scaling factor ($Nd=0$) in parallel.

The parallel element architecture is intuitively appealing, as it may be used to describe the HRTF in terms of elemental filters. The multipath description of the system, in which each element is activated simultaneously and then combined to provide the final output, is an accurate description of the physical filtering process, although possible coupling between the elemental resonators is ignored. In such a formulation, spectral features of the HRTFs arise from the amplitudes and phases of the simple resonators and their interaction, as illustrated below. We note, however, that the problem of interpreting, exactly, the nature of the resonances is again confounded by the lack of knowledge of the order of the system. The set of resonances resulting from the decomposition described in Eq. (11) is unique for a given position only if the order of the system is determined. The resonances change if the prescribed order of the system is underestimated.

Note that an analysis of the HRTF in the form of discrete resonances was performed by Shaw (1968, 1974, 1996), who identified a group of eight resonant eigenfrequencies in his measurement of transfer function from human ears. Our description of parallel resonances is consonant with Shaw's description of the pinna filtering process and in fact provides an abstract description of the same process. A fit to a HRTF using eight resonances is shown in Fig. 7. In the top panel of the figure, the eight-resonance fit is plotted as the solid curve and the terms corresponding to the individual resonances are plotted in the lower panel. The notches in the HRTF marked I and II in the upper panel correspond to the resonances marked I (dotted line) and II (dashed line) in the lower panel. These resonances correlate well with similar descriptions obtained from Shaw's empirical studies. Note that the fit in Fig. 7 is a fit to the HRTF that is composed of a DTF and the common (direction independent) transfer function. While reasonable fits to the DTF were previously obtained using three resonances, additional resonances are required to fit the common component of the HRTF.

The digital implementation of a resonator element requires only two delay elements. From an implementation point of view, the parallel-resonator filter realization would have minimal propagation delay. This can have important consequences for the implementation of dynamic virtual-acoustic-space (VAS), where the speed of switching between two filters is bounded by the propagation delay in the filters.

III. PSYCHOPHYSICAL VALIDATION OF IIR FILTER MODELS

The modeling results reported above show that excellent fits (evaluated quantitatively and visually) to the HRTF magnitude spectrum can be obtained with low-order approximations. We realize, however, that the fitting criteria chosen were based on an overall match to energy, and as such may not be appropriate to use as a psychophysical metric to assess the goodness of model fits. On the other hand, owing to the multiple dimensions of perceptual space, an objective perceptual metric cannot be specified easily. The perceptual appropriateness of the models in rendering virtual sound stimuli was thus evaluated directly via psychophysical testing.

A. Psychophysical methods

The experiments measured the discriminability of sounds processed through measured HRTFs and sounds processed through model reconstructions of the HRTFs for the same location. In a four-interval, two-alternative forced-

choice (4I,2AFC) paradigm, subjects were presented with four sounds processed through empirical and model HRTFs. Stimuli in three intervals of a set of four were processed through empirical HRTF measurements for the position of interest. The stimulus in either the second or third interval of a trial (determined randomly for each trial) was processed through the model reconstruction of the HRTF. The task entailed subjects choosing this odd interval from the set of four. This paradigm has been reported to be particularly appropriate when binaural cues are complex (Bernstein and Trahiotis, 1982). The position was varied randomly from trial to trial (and was fixed for the four intervals of each trial) and correct answer feedback was provided on each trial.

Subjects were tested with HRTF reconstructions from various model orders. Preliminary experiments suggested that, whereas subjects could discriminate a 20-pole all-pole model reconstruction and a 4-pole, 4-zero pole-zero model from the original HRTF, they found the task to be very difficult when a 25-pole all-pole model or a 6-pole, 6-zero pole-zero model reconstruction was used. In the following sections, detailed results from testing the 25-pole all-pole model and the 6-pole, 6-zero pole-zero model are presented.

Finally, all our listening experiments were conducted with nonindividualized HRTFs. It has been reported previously (e.g., Wenzel *et al.*, 1993) that localization accuracy is degraded with nonindividualized HRTFs compared to using individualized HRTFs. Because obtaining individualized HRTFs is very specialized and tedious, several virtual acoustical space applications are constrained to use nonindividualized HRTF data. While we did expect a deterioration in absolute localization accuracy with our stimuli, feedback from our listeners suggests that our simulation did result in an externalized, well-localized, and compact sound source.

B. Subjects

Three subjects, one male and two females, participated as volunteers in the experiment; their ages were between 20 and 26. The first author is subject S1. All three subjects had normal hearing (as indicated by their pure tone audiogram), and had several hours of listening experience having previously participated in similar auditory experiments. Also, all subjects had several hours of formal and informal listening experience with virtual sound stimuli generated using the SDO dataset.

C. Stimuli

The basic stimulus constituted a 80 ms burst of wideband noise (5 ms cosine-squared onset–offset ramps) presented at a comfortable listening level of about 65 dB SPL. White noise samples derived from a random noise generator (Grayson-Stadler, Model 1382) were low-pass filtered via an analog low-pass filter having a cut-off frequency of 14 kHz and a 20 dB per octave roll-off. Stimuli were presented in the 4I,2AFC task with an interstimulus interval of 500 ms. A random rove in overall level (± 5 dB) was introduced in each interval. The rove was implemented to discourage subjects from using the level at any specific frequency as a cue to differentiate between original and model filters. Note, however, that the relations among spectral components across frequency was not modified by the rove and, if the spectral shapes were different for the original and model filters, this would still be available as a cue to listeners.

The stimuli were presented over headphones (Sony, MDR-V6) after being convolved with the desired HRTFs. The HRTFs used correspond to those from the SDO dataset. The positions at which HRTFs were tested were restricted to eight azimuths in the horizontal plane (every 45°, including straight ahead) and six elevations in the frontal median plane (every 18° $+54$ to -36°, including straight ahead at 0°). Hence a total of 13 positions were tested.

IIR filters for the all-pole and pole-zero models, characterizing the DTFs for the positions of interest, were implemented as follows. The all-pole filters were realized as a cascade of biquadratic equations realized in a *transposed direct form II* structure (Oppenheim and Schafer, 1975). One advantage of such an implementation, from a practical standpoint, is that real factors and complex conjugate pairs can be combined into second-order factors requiring minimum storage and enhanced computational efficiency. The pole-zero filters were realized as a parallel connection of elemental second-order resonators, as described in a previous section. Precomputed mean functions for the left and right ears representing the direction-independent spectral features common to all measurements were constructed as minimum-phase functions. These mean functions were modeled as a fixed 6-pole, 6-zero, minimum-phase filter. The entire HRTF was reconstructed by cascading (in series) the DTFs with the mean function appropriate for that ear.

Because we lacked a real-time IIR computational engine, the IIR filters corresponding to the DTFs and the fixed mean functions were implemented in the FIR form by computing their impulse responses. The model IIR filters are minimum-phase functions and therefore also have a minimum-phase impulse response. The impulse responses were chosen to have 512 taps and were shown, both numerically and by measurement, to match the magnitude and phase of the model IIR filters. A precomputed, position-dependent, frequency-independent time delay was introduced in the minimum-phase impulse response for the lagging ear in order to provide the listener with the appropriate interaural delay for the position of interest. More details on the construction of the stimuli are provided in Kulkarni *et al.* (1999). Stimuli using these impulse responses were presented using the Convolvotron. The Convolvotron, which is housed in an IBM PC compatible host computer, is a specialized digital-signal-processing device optimized for performing real-time convolutions. It presents, over headphones, the composite binaural signal derived from simultaneously convolving up to four sound sources with an equal number of HRTF pairs. For the purposes of this experiment, a single channel of source input corresponding to the noise stimulus was specified. For each presentation interval, a preprocessed HRTF pair was moved from a file on the IBM PC onto the memory of the Convolvotron that then performed the convolutions.

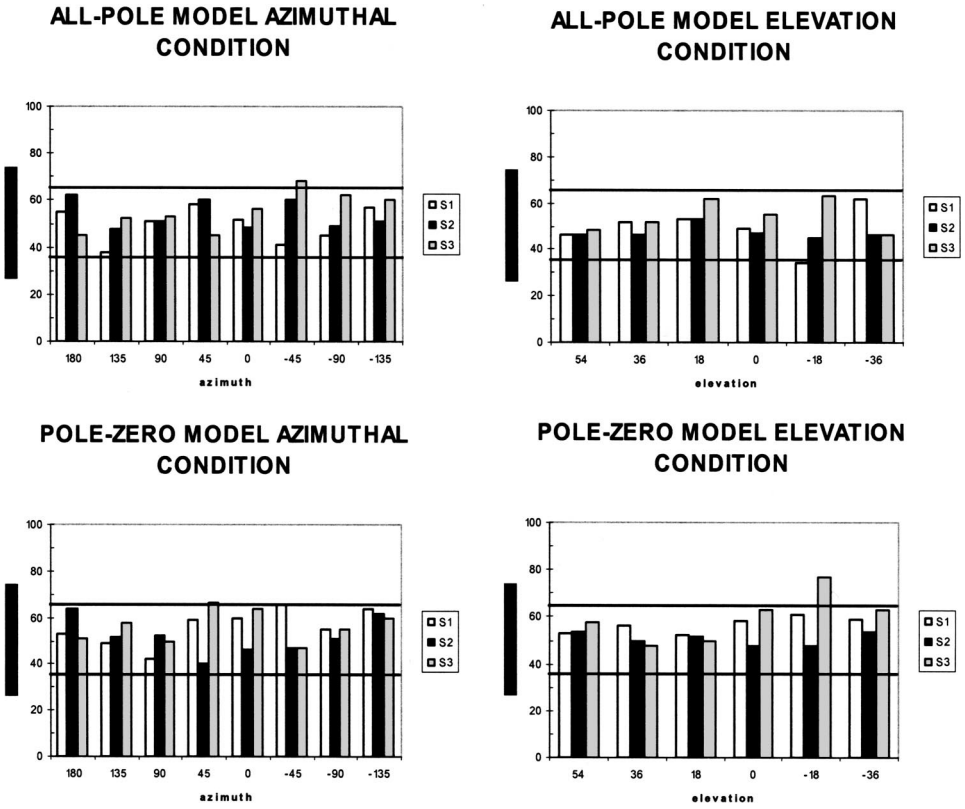


FIG. 8. The discrimination performance of subjects for 8 horizontal-plane and 6 median-plane location plane locations (every 45°, starting from directly behind) with the 25-pole all-pole model for the DTFs and with the 6-pole, 6-zero pole-zero model for the DTFs. The 95% confidence region at an $\alpha=0.05$ significance level is shown by the region bounded by the two solid lines.

D. Results

1. Subjective observations

All subjects reported perceiving compact and externalized sound images with all sound stimuli. Note that the absolute accuracy of the simulated source was not tested in our discrimination paradigm.

For the 25-pole (no zero) reconstruction and for the 6-pole, 6-zero reconstruction, subjects expressed great difficulty attempting to discriminate model HRTFs from the original measurements. There was no perceived loss of directional information in the model HRTFs. On most trials subjects reported the perception of four different sounding stimuli (independent noise samples at different levels) from the same direction.

2. Objective measurements

a. All-pole model. As noted above, for the 20-pole model, subjects had little difficulty discriminating model HRTFs from empirical HRTFs, and formal measurements were not made. For the 25-pole (no zero) model, the subject performance at the eight azimuths in the horizontal plane is presented in the top left panel of Fig. 8 and performance at the six elevations in the frontal median plane (including the 0° azimuth) is presented in the top right panel of Fig. 8. The individual performance for each location is determined from a total of 50 trials. The 95% confidence limits for chance were calculated from a binomial model and plotted as the two horizontal lines in the figure. For a significance level $\alpha=0.05$, the region bounded by the two solid lines on the graphs corresponds to a chance performance. We note from Fig. 8 that, in general, subjects are operating at chance: subject performance was outside the confidence limits in only two cases: subject S3 at the $-45°$ azimuth position and subject S1 at the $-18°$ elevation position. These deviations are not unexpected because a fraction of individual trials would be expected to exceed the confidence limits.

b. Pole-zero model. For the 4-pole, 4-zero model, subjects had little difficulty discriminating model HRTFs from empirical HRTFs and formal measurements were not made. For the 6-pole, 6-zero model, subject performance at the azimuthal locations and the median plane elevations is presented in the lower left and lower right panels of Fig. 8. As with the previous data, confidence bounds for a chance performance (at a significance level of $\alpha=0.05$) are indicated by the two solid lines on the graphs. Again, note that the subject performance rarely exceeds these bounds. Only subject S3 at the $+45°$ azimuth and the $-18°$ elevation positions and subject S1 at the $-45°$ azimuth position exceed the bounds, and even for these positions performance does not exceed the bounds by much.

IV. COMPARISON WITH PREVIOUS STUDIES

In a previous study, Asano *et al.* (1990) also adopted a pole-zero model to fit HRTFs measured in the median plane. They used the Kalman modification of the least-squares problem, presented in Appendix A, to obtain the model coefficients. They reported that very high model orders (40 poles and 40 zeros) were required to achieve psychophysical performance that was comparable to that achieved using the original (empirically obtained) HRTF set in a absolute localization task. We were able to demonstrate a good performance in our listening tasks using low model orders for DTFs (six poles and six zeros). We propose two explanations for this large difference in required model orders between the

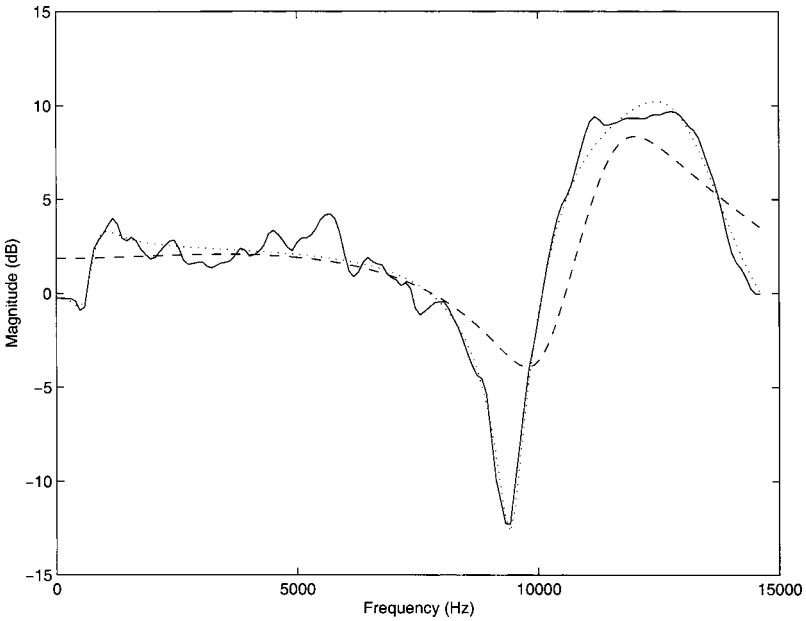


FIG. 9. Example DTF (solid line) and model fits resulting from the Kalman modification used by Asano and colleagues (dashed lines) and with the iterative weighting procedure (dotted line).

two studies. First, Asano and colleagues used the Kalman algorithm to fit their model, which by itself does not provide an optimal model fit. Several of the pole-zero pairs used by Asano and colleagues could therefore have been redundant and poorly utilized. In addition to the Kalman modification, we used the iterative weighting procedure described above. The results of fitting a DTF with a 6-pole, 6-zero model, with and without this weighting procedure, is presented in Fig. 9. For identical model orders, the superior fit (in the log-magnitude domain) resulting from the weighting function is obvious. Second, Asano and colleagues modeled the HRTF directly whereas DTFs and mean functions were modeled separately in the current study. By obtaining a mean function for all measurements, we can invest all pole-zero pairs to only fit the directional features of HRTFs. Because the mean functions contain features common to all measurements, we propose that it is most efficient for both storage and computation to process stimuli through desired DTFs and finally through a common filter stage characterizing the mean function. The high model orders required by Asano and colleagues could also result from requiring a fit of the nondirectional features, in addition to the directional ones. In a more recent study closely related to ours, Blommer *et al.* (1994) have reported on the pole-zero modeling of HRTFs. They propose a scheme that estimates the mean function and estimates the poles and zeros of the DTF using a gradient descent algorithm that minimizes the *log*-least-squares-error criterion. As in our method, Blommer *et al.* use the *log*-error criterion to overcome the shortcomings of the standard least-squares error metric of fitting spectral peaks better than valleys. Unlike our method (which sets up an approximate quadratic minimization), however, they solve the nonlinear problem algorithmically. There are merits to both approaches. Whereas the gradient descent approach is more direct, it is known to suffer from problems associated with local minima and is sensitive to the initial estimate of the pole-zero positions. It is also computationally very expensive. The weighted-least-squares (WLS) method, on the other hand, solves a linear equation and is therefore extremely fast and requires little computational resource. Another difference in our study is the interpretation of the low-order pole-zero model as describing a system of parallel elemental resonators. While we maintain that pole-zero patterns might provide some insight into the pinna filtering process, decomposing the DTF into elemental, *parallel* resonances is a more appropriate description of the physical phenomenon, and we consider it to be a promising direction for future study. Empirically, we have noted that the solution to the WLS formulation converges for all the HRTF data tested thus far. We hence consider it to be a very efficient algorithm that provides low-order HRTF solutions that are psychophysically correct.

V. CONCLUSIONS

In this study we explore the use of simple architectures for representing the directional information encoded by the HRTF. Results from rational-transfer function approximation strategies suggest computational strategies that are easy to implement in three-dimensional (3D) audio displays. A marked improvement in computational efficiency in rendering directional sound over using the empirical measurement is apparent. Also, the low-latency processing offered by the pole-zero IIR filter makes it particularly promising.

Note that although we have not yet addressed the issue of optimal model order to reconstruct HRTFs, this study supports the choice of low-order IIR models in general. The signal analysis and behavioral experiment results indicate that there is little loss of directional information in spatial sound synthesized through the low-order IIR filter sections. For the directions tested by us in the listening experiments, stimuli rendered through a 6-pole, 6-zero model of the DTF had inaudible differences from stimuli rendered through the actual DTF. While it is not certain that this result would hold across all listening positions and different HRTF sets, in-

creasing the order of the model (by adding a few extra poles and zeros) could provide acceptable fits for conditions where a discrepancy was noted.

The reduced number of parameters required to specify HRTFs makes the models amenable to physical interpretation. For example, as shown earlier, the pole-zero model can specify the DTF for a given position by six poles and six zeros in the complex z plane. Because poles and zeros characterize the resonances and antiresonances of a physical system, it may be possible to relate the poles and zeros of a DTF to features in the external ear geometry. Alternatively, the elemental resonance description of the DTFs sets up a meaningful description of the pinna filtering process. Among other applications, such an analysis could be applicable in determining a quantitative understanding of the variability of DTFs across a population of subjects. Future extensions of this work include both methods to determine optimal model orders and the relation of the chosen parameters to features in the HRTF used for sound localization.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health, NIDCD Grant No. RO1 DC00100.

APPENDIX A: RATIONAL TRANSFER FUNCTION MODELS

In the most general formulation of discrete-time rational transfer function models, an input driving function $u[n]$ and the output sequence $x[n]$ are related by the following linear difference equation:

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=0}^{q} b[k]u[n-k]. \qquad \text{(A1)}$$

This relation is simply represented in the frequency domain using discrete-time Fourier transforms (DTFT), as defined below. The DTFT of the filter impulse response is then the transfer function for this linear system. The transfer function $H(e^{J\omega})$ relating the output $x[n]$ to the input $u[n]$ for the general system described by Eq. (A1) is given by the rational function,

$$H(e^{J\omega}) = \frac{X(e^{J\omega})}{U(e^{J\omega})} = \frac{B(e^{J\omega})}{A(e^{J\omega})}, \qquad \text{(A2)}$$

where

$$X(e^{J\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-J\omega n}$$

$$= \text{Fourier transform of the output signal } x[n],$$

$$U(e^{J\omega}) = \sum_{n=-\infty}^{\infty} u[n]e^{-J\omega n}$$

$$= \text{Fourier transform of the input signal } u[n],$$

$$A(e^{J\omega}) = \sum_{k=0}^{p} a[k]e^{-J\omega k}$$

$$= \text{Fourier transform of the model coefficients } a[k],$$

$$B(e^{J\omega}) = \sum_{k=0}^{q} b[k]e^{-J\omega k}$$

$$= \text{Fourier transform of the model coefficients } b[k],$$

and $a[0]$ is constrained to be unity.

The denominator function $A(e^{J\omega})$ considered as a function of a complex variable $z = e^{J\omega}$, $A(z)$, should have all of its zeros inside the unit circle in the complex z plane in order to guarantee that the system described by Eq. (A1) is stable and causal. The zeros of $A(z)$ are called the poles of $H(z)$. They are important because they correspond to frequencies for which there can be an output with no input and thus correspond to the natural frequencies of the system. Similarly, the zeros of $B(z)$ are called the zeros of the system. Because the polynomials for $A(z)$ and $B(z)$ have orders p and q, respectively, there are p zeros and q poles [This nomenclature ignores the poles and zeros of $H(z)$ at the origin of the z plane. Note that the functions $A(z)$ is actually a pth-order polynomial in z^{-1}, so that $A(z)$ has p poles at $z = 0$ as well as p zeros just described. Similarly, $B(z)$ has q poles at $z = 0$. Thus, the system function $H(z)$ has a number of poles or zeros at $z = 0$ dependent on the difference between p and q. This is standard practice in linear system theory because these trivial poles and zeros are not specific to individual spectra when the numbers of poles and zeros, p and q, are given.] for the system described by Eq. (A1).

We note at this point that the problem of rational function approximation occurs in several fields, and different terminologies exist for equivalent models and operations. In signal processing, the general model is referred to as a *pole-zero* model, which specializes to an *all-pole* model when $q = 0$ or, equivalently, when all of the coefficients $b[k]$ except $b[0]$ are equal to zero. (As stated previously, the system function for the "all-pole" case of order p also includes p zeros at $z = 0$ that are ignored in the categorization of systems.) This all-pole case is also called a strictly autoregressive or AR process in the statistical literature. In this case, model Eq. (A1) reduces to

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + b[0]u[n]. \qquad \text{(A3)}$$

The dependence of the current value on past values of the output illustrates origins of the term autoregressive for this type of IIR system.

In the other extreme case, when the $a[k]$ coefficients are all zero (except $a[0]$, which is always unity), the general model reduces to an *all-zero* model of order q and the output can now be written as a weighted moving average of the input values, as shown by the equation

$$x[n]=\sum_{k=0}^{q} b[k]u[n-k]. \tag{A4}$$

(Again, the q poles at $z=0$ are ignored in this nomenclature.) A process described by these equations is called a moving-average or MA process in the statistical literature. In this all-zero case, the system impulse response $h[n]$ is given by the coefficients $b[n]$ because $x[n]=b[n]$ when $u[n]=\delta[n]$. (The function $h[n]$ is the response of the model system to the unit-sample or discrete-impulse function $\delta[n]$.) This all-zero case thus corresponds to a FIR filter description or a MA process.

Using these concepts of moving-average and autoregressive processes, the general formulation of the *pole-zero model* with p poles ($p\geqslant 1$) and q zeros ($q\geqslant 1$) is thus referred to in the statistical literature as an auto-regressive-moving average or ARMA (p,q) process because both AR and MA elements are included.

In this paper we focus on IIR filter models and thus includes the all-pole or strictly AR model and the general pole-zero or ARMA process models.

1. All-pole model for HRTFs

In the following paragraphs, a method is presented to determine the coefficients $a[k]$ for an all-pole filter that approximates a real system with a measured impulse response. Let $h[n]$ for $n=0,...,m$ represent the real-valued impulse response measured in a listener's ear-canal for a given source position and let $\hat{h}[n]$ be the impulse response of the approximating all-pole filter.

Because the approximating filter is all-pole, the IIR impulse response $\hat{h}[n]$ must satisfy Eq. (A3), resulting in the following equation with appropriate coefficients $a[k]$:

$$\hat{h}[n]=-\sum_{k=1}^{p} a[k]\hat{h}[n-k]+b[0]\delta[n], \tag{A5}$$

where $\delta[n]$ is the unit-sample input to the system and the coefficients $a[k]$ determine the IIR filter representation for the model system. It is assumed here that $p<m$ so that the number of coefficients of the approximating filter is less than the number of sample points specifying the empirical measurement to be approximated. The problem is to find coefficients $a[k]$ and $b[0]$ that result in a good correspondence between $\hat{h}[n]$ and $h[n]$. The general point of view is that, even though the measured $h[n]$ are given exactly by a FIR filter of order m, the actual HRIR filter may be more efficiently characterized by an IIR filter representation of lower order ($p<m$). Measurements of $h[n]$ for large n would be unreliable because of measurement noise so the measurements are terminated after a finite number of points.

In terms of Fourier transforms, Eq. (A5) becomes Eq. (1) in the text:

$$\hat{H}(e^{J\omega})=\frac{b[0]}{1+\Sigma_{k=1}^{p} a[k]e^{-J\omega k}}, \tag{A6}$$

and the problem is to find the coefficients $a[k]$ that allow $\hat{H}(e^{J\omega})$ to best approximate $H(e^{J\omega})$.

Although a variety of methods can be used to solve for the coefficients $a[k]$ (Makhoul, 1975; Burg, 1975; Kay, 1988), the autocorrelation method for linear prediction (Makhoul, 1975) is used here because the solution is computationally efficient, unbiased, and results in a filter that is stable. The outline of this method is described in the text.

2. Pole-zero model for HRTFs

Pole-zero model estimation is useful in extending the all-pole estimator to processes that cannot be modeled accurately using a finite number of poles, and observation noise in data nearly always makes the pole-zero model the appropriate one to use. However, as we demonstrate below, the solution of optimal pole-zero estimators is much more difficult because this requires the solution of a set of highly nonlinear equations.

As outlined in the text [cf. Eq. (2)], the pole-zero model is described by the corresponding transfer-function,

$$\hat{H}(e^{J\omega})=\frac{b[0]+\Sigma_{k=1}^{q} b[k]e^{-J\omega k}}{1+\Sigma_{k=1}^{p} a[k]e^{-J\omega k}}. \tag{A7}$$

If it is assumed that the impulse response is square summable (i.e., $\{h[n]\}\in l_2$), the canonical l_2 approximation problem may now be stated as follows. Given the impulse response sequence $\{h[0],h[1],\ldots,h[m]\}\in l_2$, find the recursive filter of the form in Eq. (A7) with an impulse response $\{\hat{h}[0],\hat{h}[1],\ldots\}$, which minimizes the following:

$$\begin{aligned}\|h-\hat{h}\|^2&=\sum_{k=0}^{m}(h[k]-\hat{h}[k])^2\\&=\frac{1}{2\pi}\int_{-\pi}^{\pi}|H(e^{J\omega})-\hat{H}(e^{J\omega})|^2\,d\omega\\&=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|H(e^{J\omega})-\frac{B(e^{J\omega})}{A(e^{J\omega})}\right|^2 d\omega=C.\end{aligned} \tag{A8}$$

Note that the value of the this cost function C depends only on the coefficients $\{a[k]\}$ and $\{b[k]\}$ or, equivalently, the Fourier transforms $A(e^{J\omega})$ and $B(e^{J\omega})$ when the measured impulse response, or equivalently $H(e^{J\omega})$, is held fixed.

Although the problem as just stated exhibits a natural choice of error, it is ill-behaved in other respects. First, because the IIR filter is recursive, the solution must be obtained by solving nonlinear equations. Second, the solution filter may be unstable, even though the data originated from a stable filter. Third, the solution filter provides a fit that minimizes the absolute error on a linear scale, which, as discussed in the text, may not be perceptually optimal.

We propose a modification of the approximation problem in the following three steps in order to overcome these shortcomings.

(1) We consider a cost function that is quadratic in the coefficients of the recursive filter in Eq. (A7). We do so by considering the modified problem proposed by Kalman (1958) that seeks coefficients that minimize the quadratic form

$$\hat{C}=\frac{1}{2\pi}\int_{-\pi}^{\pi}|H(e^{J\omega})A(e^{J\omega})-B(e^{J\omega})|^2\,d\omega. \quad \text{(A9)}$$

(2) We use the iterative procedure suggested by Steiglitz and McBride (1965) as an extension to the Kalman technique. This procedure entails iterative minimization of the following:

$$\hat{C}_i=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{H(e^{J\omega})A_i(e^{J\omega})}{A_{i-1}(e^{J\omega})}-\frac{B_i(e^{J\omega})}{A_{i-1}(e^{J\omega})}\right|^2 d\omega. \quad \text{(A10)}$$

It is easy to see that if convergence is obtained, i.e., if the coefficients of $A_i(e^{J\omega})$ converge for some value of i, the minimization of Eq. (A10) is equivalent to that of Eq. (A8).

(3) Finally, we seek a solution on a decibel scale by introducing a weighting function in the integrand of Eq. (A10),

$$\hat{C}_i=\frac{1}{2\pi}\int_{-\pi}^{\pi}|W_i(e^{J\omega})|^2\left|\frac{H(e^{J\omega})A_i(e^{J\omega})}{A_{i-1}(e^{J\omega})}-\frac{B_i(e^{J\omega})}{A_{i-1}(e^{J\omega})}\right|^2 d\omega, \quad \text{(A11)}$$

where

$$W_i(e^{J\omega})=\frac{\left[\log|H(e^{J\omega})|-\log\left|\frac{B_{i-1}(e^{J\omega})}{A_{i-1}(e^{J\omega})}\right|\right]^2}{|H(e^{J\omega})A_{i-1}(e^{J\omega})-B_{i-1}(e^{J\omega})|^2}. \quad \text{(A12)}$$

This solution overcomes the shortcomings mentioned above: the solution filter is obtained by solving a set of linear equations, it is stable, and the solution minimizes the error on a decibel scale. The final representation of the problem in Eq. (A12) can be used for any choice of orders p and q.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**), "The CIPIC HRTF Database," *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, NY, 21–24 October, 2001, pp. 99–102.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). "Role of spectral cues in median plane localization," J. Acoust. Soc. Am. **88**, 159–168.

Bernstein, L. R., and Trahiotis, C. (**1982**). "Detection of interaural delay in high-frequency noise," J. Acoust. Soc. Am. **71**, 147–152.

Blauert, J. (**1996**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).

Blommer, M. A., and Wakefield, G. H. (**1994**). "On the design of pole-zero approximations using a logarithmic error measure," IEEE Trans. Signal Process. **42**, 3245–3248.

Burg, J. (**1975**). "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford University.

Fan, H. (**1988**). "Application of Benveniste's convergence results in the study of adaptive IIR filtering algorithms," IEEE Trans. Inf. Theory **34**, 692–709.

Humanski, R. A., and Butler, R. A. (**1988**). "The contribution of the near and far ear towards localization of sound in the saggital plane," J. Acoust. Soc. Am. **83**, 2300–2310.

Kalman, R. E. (**1958**). "Design of a self-optimizing control system," Trans. ASME **80**, 468–478.

Kay, S. M. (**1988**). *Modern Spectral Estimation* (Prentice–Hall, Englewood Cliffs, NJ).

Kistler, D. J., and Wightman, F. L. (**1991**). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647.

Kulkarni, A., and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature (London) **396**, 747–749.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1999**). "Sensitivity of human subjects to head-related-transfer-function phase spectra," J. Acoust. Soc. Am. **105**, 2821–2840.

Levinson, N. (**1947**). "The Wiener RMS (Root Mean Square) error criterion in filter design and prediction," J. Math. Phys. **25**, 261–278.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The Duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Makhoul, J. (**1975**). "Linear prediction: A tutorial review," Proc. IEEE **63**, 561–580.

Martens, W. L. (**1987**). "Principal components analysis and resynthesis of spectral cues to perceived direction," in *Proceedings of the International Computer Music Conference*, edited by J. Beauchamp (International Computer Music Association, San Francisco, CA), pp. 274–281.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.

Morimoto, M. (**2001**). "The contribution of two ears to the perception of vertical angle in saggital planes," J. Acoust. Soc. Am. **109**, 1596–1603.

Musicant, A. D., and Butler, R. A. (**1985**). "Influence of monaural spectral cues on binaural localization," J. Acoust. Soc. Am. **77**, 202–208.

Oppenheim, A. V., and Schafer, R. W. (**1975**), *Digital Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ).

Papoulis A., and Pillai, S. U. (**2002**). *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York).

Shaw, E. A. G., and Teranishsi, R. (**1968**). "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240–249.

Shaw, E. A. G. (**1974**). "The external ear," in *Handbook of Sensory Physiology* (Springer-Verlag, Berlin, Berlin), Vol. 5, Part 1.

Shaw, E. A. G. (**1996**). "Acoustical features of the human external ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Lawrence Erlbaum, NJ).

Steiglitz, K., and McBride, L. E. (**1965**). "A technique for the identification of linear systems," IEEE Trans. Autom. Control **AC-10**, 461–464.

Wenzel, E. M., Arruda, M., and Kistler, D. J. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone stimulation of free-field listening I: stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone stimulation of free-field listening II: psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Wightman, F. L., and Kistler, D. J. (**1999**). "Sound localization with unilaterally degraded spectral cues," J. Acoust. Soc. Am. **105**, 1162.

Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing

Christopher W. Turner
Dept. of Speech Pathology and Audiology, Dept. of Otolaryngology Head and Neck Surgery, University of Iowa, Iowa City, Iowa 52242

Bruce J. Gantz
Dept. of Otolaryngology Head and Neck Surgery, University of Iowa, Iowa City, Iowa 52242

Corina Vidal
Dept. of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52242

Amy Behrens
Dept. of Speech Pathology and Audiology, Dept. of Otolaryngology Head and Neck Surgery, University of Iowa, Iowa City, Iowa 52242

Belinda A. Henry
Dept. of Communicative Disorder, University of Wisconsin-Madison, Madison, Wisconsin 53706

(Received 17 July 2003; revised 8 January 2004; accepted 26 January 2004)

The purpose of this study was to explore the potential advantages, both theoretical and applied, of preserving low-frequency acoustic hearing in cochlear implant patients. Several hypotheses are presented that predict that residual low-frequency acoustic hearing along with electric stimulation for high frequencies will provide an advantage over traditional long-electrode cochlear implants for the recognition of speech in competing backgrounds. A simulation experiment in normal-hearing subjects demonstrated a clear advantage for preserving low-frequency residual acoustic hearing for speech recognition in a background of other talkers, but not in steady noise. Three subjects with an implanted ''short-electrode'' cochlear implant and preserved low-frequency acoustic hearing were also tested on speech recognition in the same competing backgrounds and compared to a larger group of traditional cochlear implant users. Each of the three short-electrode subjects performed better than any of the traditional long-electrode implant subjects for speech recognition in a background of other talkers, but not in steady noise, in general agreement with the simulation studies. When compared to a subgroup of traditional implant users matched according to speech recognition ability in quiet, the short-electrode patients showed a 9-dB advantage in the multitalker background. These experiments provide strong preliminary support for retaining residual low-frequency acoustic hearing in cochlear implant patients. The results are consistent with the idea that better perception of voice pitch, which can aid in separating voices in a background of other talkers, was responsible for this advantage.
[DOI: 10.1121/1.1687425]

PACS numbers: 43.66.Sr, 43.66.Ts, 43.71.Ky [PFA] Pages: 1729–1735

I. INTRODUCTION

For many people with severe and profound hearing losses, cochlear implants have restored speech understanding to remarkable performance levels that acoustic amplification via hearing aids was unable to provide. However, the decision to undergo implantation surgery involves some tradeoffs, as the patients' residual acoustic hearing is no longer usable, and only electric stimulation is available. For example, many users of cochlear implants report that the perception of sound becomes ''mechanical'' or ''raspy'' when compared to their memories of acoustic hearing, and that many of the aesthetic qualities of sound are diminished. This loss of aesthetic quality of sound is most likely related to a decrease in the ability to perceive the pitches of sounds (Gfeller *et al.*, 2002). The loss of pitch perception is primarily a consequence of the limited spectral resolution of current cochlear implants, which does not appear to be a limitation for understanding speech in quiet for the most successful implant users (Fishman *et al.*, 1998). However, understanding speech in background noise requires spectral resolution even finer than that required to understand speech in quiet (Fu *et al.*, 1998). Even the most successful implant users only realize perhaps 6–8 channels of distinct ''place–frequency'' information across the entire spectral range, and this deficit in spectral resolution has a direct negative consequence on the implant patients' ability to understand speech in background noise (Friesen *et al.*, 2001).

A recent development in cochlear implants has been to implant an electrode only partially into the cochlea, in order to preserve the residual acoustic hearing that many patients still have for low frequencies (Von Ilberg *et al.*, 1999; Gantz and Turner, 2003). In these patients, usable acoustic hearing is usually present up to frequencies of 500 or 750 Hz, and the electrical stimulation provides the patient with high-

frequency speech information. Thus, these patients perceive sound via a ''combined acoustic and electric'' (A+E) mode. In addition to the possibility that preserving residual acoustic hearing may have for the aesthetic qualities of sound, it is also possible that preserving residual hearing may contribute to better speech recognition in background noise.

There are several mechanisms by which the preserved residual low-frequency hearing might improve speech understanding in noise as compared to the traditional full-length (long) cochlear implant. The low-frequency residual acoustic hearing presumably has better spectral resolution than the low-frequency portion of a traditional cochlear implant. Henry and Turner (2003a) showed that normal-hearing listeners could resolve spectral ripples nearly an order of magnitude more closely spaced in frequency than cochlear implant users. Although the presence of sensorineural hearing loss typically might decrease spectral resolution compared to normal hearing, patients with sensorineural hearing loss still had better spectral resolution than that provided by a typical long-electrode cochlear implant (Henry and Turner, 2003b). This advantage in spectral resolution might provide a relative benefit in perceiving the spectral features of speech sounds, particularly when presented in noise. On the other hand, many of the features of speech that depend upon spectral resolution (i.e., place of articulation) are located in the higher frequency regions of the spectrum, and low-frequency residual hearing therefore may not be of much assistance.

Another way in which residual acoustic hearing might be helpful to the implant listener would be when speech recognition is tested in a background of multiple talkers. Whereas most normal-hearing listeners can often perform as well or better when listening in fluctuating backgrounds of other talkers as compared to steady noises, implant users usually perform more poorly under these circumstances. Nelson *et al.* (2003) found that cochlear implant users have considerable difficulty in recognizing speech in modulated-noise maskers. Their study also demonstrated, by presenting spectrally limited speech to normal-hearing listeners (to simulate cochlear implant processing), that the reduced spectral resolution was responsible for the problems that implant users experience in fluctuating backgrounds. Qin and Oxenham (2003) demonstrated that even with 24 channels of frequency resolution provided to normal-hearing listeners in a simulation of cochlear implant speech, performance was poorer than for unprocessed speech in a background of a competing talker. Stickney *et al.* (2003) reported that traditional cochlear implant users showed no advantage in recognizing speech presented with a competing talker as compared to steady noise. These studies attribute their findings to the fact that the cochlear implant listeners could not gain an advantage (as normal-hearing listeners did) by perceiving the different pitches of the talkers. Dorman *et al.* (1996) as well as Gfeller *et al.* (2002) have shown that cochlear implant users have great difficulty in distinguishing the pitches of tones, with frequency difference limens for low-frequency tones approaching 100 Hz in some cases. Thus, preserving low-frequency acoustic hearing for cochlear implant patients might, in such cases, lead to an advantage in speech understanding in a background of other talkers, as compared to traditional cochlear implants. Some support for this concept has been demonstrated by Kong *et al.* (2003), who found that cochlear implant users showed improved speech recognition in a competing-talker background when they were allowed to use their low-frequency acoustic hearing in the contralateral ear, even though the contralateral ear by itself was not capable of any speech recognition.

If the improved pitch perception of residual low-frequency hearing could be used by the listener to ''separate'' various voices via fundamental frequency, then the patient may experience improved speech understanding in multi-talker backgrounds. Different fundamental frequencies assist the listener to ''group'' the various upper-frequency components of speech and therefore improve recognition of the target voice (Assmann, 1999). Brokx and Nooteboom (1982), Assmann and Summerfield (1990), Culling and Darwin, (1993) and Bird and Darwin (1999) have shown the importance of the fundamental frequency cue for the separation of simultaneous voices in normal-hearing listeners. However, several studies have indicated that traditional cochlear implant users have difficulty in perceiving the fundamental frequency of signals for frequencies greater than 200 Hz. This is due to the fact that place–frequency cues for the fundamental are generally poor (due to poor spectral resolution), and envelope (temporal) cues for the fundamental are only salient at the lower frequencies (Geurts and Wouters, 2001; Green *et al.*, 2002).

The present experiments investigate the possibility that residual low-frequency acoustic hearing can provide benefits for speech understanding in background noises. Two different background conditions were employed, speech-shaped steady noise and competing talkers, in order to distinguish between the several hypothesized advantages of preserving residual hearing. A simple improvement due to increased spectral resolution of speech features should occur equally in both noise and competing-talker backgrounds, whereas an improvement that is due specifically to an advantage in the perception of the voice pitch would be expected to appear most strongly in the multiple-talker background. The first experiment employs simulations of cochlear implant processing (both traditional or long-electrode, and the ''combined acoustic and electric'' or A+E approach). The second experiment uses the same measures of speech understanding in backgrounds for two groups of actual patients using either traditional long-electrode cochlear implants or the combined acoustic plus electric implants (i.e., A+E).

II. EXPERIMENT 1: SIMULATIONS IN NORMAL-HEARING LISTENERS

A. Subjects

15 young-adult listeners participated in this experiment. All had hearing within 20 dB of the normal standards at octave audiometric frequencies (0.25–8.0 kHz) and were native speakers of American English.

B. Stimuli and procedures

The task for the listeners was to identify a spondee (two-syllable) word spoken by a female talker in the presence of a

background sound. The 12 spondee items were homogeneous in difficulty and were digitized from a commercial recording (Harris, 1991). The fundamental frequency of the spondee items ranged from 212–250 Hz. The spondees ranged in duration from 1.12 to 1.63 s. For each presentation, the spondee was chosen randomly from the set of 12. Following each presentation, the listener responded on a touch screen with the spondee that they thought had been presented. The listeners were required to respond on each trial, and instructed to guess if they were not sure of the correct answer. The nontest ear was plugged during the testing.

Two different backgrounds were employed. The competing-talker condition consisted of two simultaneously presented sentences originally recorded as items on the SPIN test (Bilger, 1984). One background talker was a male (fundamental frequency range (81–106 Hz) and the other a female (fundamental frequency range 149–277 Hz). This female talker was not the same talker who produced the spondee. The two background voices were mixed together at equal rms amplitudes. The same mixed-sentence background was presented on each trial. The other background condition was a steady-state white noise that had been low-pass filtered at −12 dB/octave above 400 Hz, to generally simulate the long-term speech spectrum. The same sample of noise background was presented on each trial. The spectra of the competing-talker background and the steady noise were not matched; the competing-talker spectrum contained considerably more spectral peaks and valleys than the steady noise. The competing background signal durations (both sentences and noise) were 2.5 s, and the onset of the target spondee was 500 ms following the onset of the background signal.

The spondees and the backgrounds were presented in three conditions. The first was an unprocessed condition which consisted of the unprocessed speech spondee and the unprocessed background. The second condition was a simulation of a 16-channel cochlear implant, implemented by using the temporal speech envelope within each frequency channel to modulate a corresponding narrow frequency band of noise. Both the target spondees and the background noises were processed. This general technique has been used to simulate cochlear implant speech in numerous studies (i.e., Shannon *et al.*, 1995) and has been shown to provide a good approximation of the theoretical maximum performance of cochlear implant patients for a given degree of frequency resolution (Fishman *et al.*, 1998). The current procedure was implemented using routines written in MATLAB, and the specifics for this 16-channel simulation are described in detail in Henry and Turner (2003a). The third condition was designed to simulate the short-electrode "acoustic plus electric" (A+E) situation. The unprocessed spondees and backgrounds were each low-pass filtered at 500 Hz using a −24-dB/octave digital filtering algorithm. The 16-channel simulations of the spondees and backgrounds were high-pass filtered at 500 Hz, using a similar digital filtering algorithm at −24 dB/octave. These low-pass unprocessed and corresponding high-pass implant simulations were then combined to yield the A+E condition, which had the same relative balance between the low- and high-frequency portions of the spectrum as the unprocessed speech. These A+E stimuli therefore consisted of the entire upper 13 channels (and part of the 14th channel) of the electric simulation mixed with the acoustic signal below 500 Hz. The background signals (noise and competing talker) were processed separately from the spondees, and were then combined following the appropriate attenuation values to obtain the desired signal-to-noise (S/N) ratio, expressed in the rms average value of the spondee and the background.

Prior to any speech in noise testing, each subject participated in one or more practice runs to familiarize them with the spondees and the responses. In this practice run, the spondees were presented without any background noise. All subjects were able to recognize the spondees at 100% accuracy following these practice sessions.

All signals were presented via a loudspeaker in sound field, and the spondees were presented at an average level of 68 dB SPL. Both target spondees and backgrounds were stored on a Macintosh G4 computer and output through separate channels of a DigiDesign 16-bit digital-to-analog converter. The level of the background was controlled by a TDT programmable attenuator. An adaptive procedure was then used to determine the 50%-correct point (in terms of S/N ratio) for recognition of the spondees in noise (SRT). The spondees were initially presented at a signal-to-noise ratio of either +10 or +20 dB (depending upon the condition); this allowed the listener to easily identify the target voice and recognize the spondees of the first few trials. For each correct response the S/N ratio was decreased by 2 dB and for each incorrect response the S/N ratio was increased by 2 dB. For a single run this procedure continued until 14 reversals had occurred and the final value for that run was taken as the average of the final ten reversals. Each subject completed four runs in each condition, and their final data for that condition were taken as the average of the last three runs. Each subject completed all four runs of a condition before progressing to another condition. The order of conditions was randomized across subjects.

C. Results and discussion

Figure 1 displays the results averaged across subjects of experiment 1. The SRT in noise (in dB S/N ratio) is plotted as a function of the three processing conditions. It is clear that there are large differences between three processing conditions when the background is composed of competing talkers, whereas the differences between processing conditions are smaller or nonexistent for the noise background. The general finding of improved speech recognition in a background of voices for unprocessed speech over spectrally limited speech is in agreement with past results (Qin and Oxenham, 2003). The present results differ slightly from that of Qin and Oxenham (2003) for the case of a steady noise background. In their study, unprocessed speech yielded SRTs that were 5.5 dB better than 24-channel processed speech, whereas in our study the improvement for unprocessed speech over 16-channel speech was 2.2 dB (which was not significant). Perhaps differences in the specific speech materials and maskers account for this discrepancy.

The comparison of particular interest for this study was to determine if supplementing the cochlear implant speech

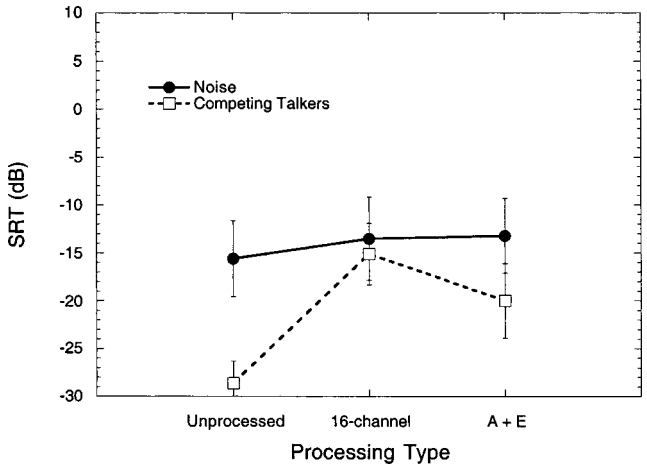


FIG. 1. Group mean SRT values for the acoustic simulations presented to normal-hearing listeners in the two types of background stimuli. The error bars represent the standard deviations across subjects for each condition.

with natural low-frequency acoustic hearing (A+E) could be used to improve some of the advantage in competing talkers that is lost to traditional cochlear implant users. The SRT in competing talkers for unprocessed speech was −28.6. dB, as compared to −15.1 dB for the 16-channel simulation; this was a 13.5-dB disadvantage for the simulated implant speech. The addition of low-frequency unprocessed speech to the simulation reduced this disadvantage to 8.6 dB. In steady noise, the differences between processing conditions were less than 2.5 dB for all comparisons. A two-way ANOVA was performed on the data for the A+E and 16-channel conditions for both noise and competing-talker backgrounds. Both main effects were significant (condition; $F=5.40$, df=1,56; $p=0.0024$; background, $F=17.75$, df =1,56, $p=0.0001$). Of interest is the significant interaction between the two main effects ($F=6.58$, df=1,56 $p=0.013$), which indicates that an advantage was seen in competing talkers over noise for the A+E condition as compared to the 16-channel condition, but not for the noise background condition. The lack of advantage for maintaining low-frequency acoustic hearing in noise suggests that presumably improved spectral resolution for acoustic low frequencies (as compared to 16-channel processed speech) does not result in an improvement in speech recognition in general, consistent with the idea that low-frequency speech cues are not particularly dependent upon fine spectral resolution. One possibility is that the improved spectral resolution in the low frequencies presumably leads to the ability to use pitch information to separate talkers in a multiple speaker situation. Thus, the simulation experiments provide evidence that residual low-frequency acoustic hearing can provide an advantage for speech recognition in a background of other talkers.

III. EXPERIMENT 2

A. Subjects

The subjects for experiment 2 were adult users of cochlear implants. The traditional ''long-electrode'' group consisted of 20 patients, each using the Nucleus 24 cochlear implant and its associated speech processor. They were tested using their own speech-processor maps and strategies (12 used the ACE strategy, 3 used the CIS strategy, and 5 used the SPEAK strategy). Each had been using an implant for at least 24 months.

The A+E group consisted of three patients implanted with the Iowa/Nucleus Hybrid 10-mm short-electrode device (Gantz and Turner, 2003). These patients were the first three patients to receive the 10-mm electrode and each had been wearing the device for at least 12 months prior to the data collection, and their data for speech recognition in quiet and in noise were no longer improving over time. Two of these three A+E subjects wore hearing aids in their test ear that were fit to amplify the low-frequency portion of the spectrum (unaided thresholds of the two subjects with hearing aids for frequencies of 500 Hz and below were 60–65-dB HL and their aided thresholds were 40-dB HL or better). The third A+E subject did not require a hearing aid to amplify low-frequency hearing (pure-tone thresholds of 20–25-dB HL for 500 Hz and below). The short-electrode cochlear implant stimulated 6 channels in the basal end of the cochlea, using a CIS processing strategy. The cochlear implant frequency maps that these patients found most beneficial in everyday life were also used in this study. For two of the subjects (the ones who used hearing aids) the frequency range assigned to these electrodes was 1062–7937 Hz. For the third subject, the frequencies assigned to the implant were 687–5187 Hz. The hybrid system improved consonant recognition for this group approximately 40% over the hearing-aid-only condition (Gantz and Turner, 2003).

B. Stimuli and procedures

The stimuli for experiment 2 were the same natural (unprocessed) spondees and backgrounds as used in one of the conditions in experiment 1. The nonimplant ear for all subjects was plugged during testing. Cochlear implant users listened to the spondees presented in background signals through their everyday speech processor. The A+E subjects listened to the stimuli using their cochlear implant speech processor and their acoustic hearing (which for two of them included the use of an in-the-ear hearing aid in the test ear). The spondees were presented at 68 dB SPL. During the practice sessions the subjects were allowed to adjust the output levels of their devices. The practice sessions revealed that all implant users, except for the two poorest-performing traditional electrode subjects, could identify 100% of the spondees in quiet. The two poorest-performing long-electrode implant users could only identify approximately 80% of the spondees in quiet. All implant users completed at least four runs of the adaptive SRT procedure in each of the two background conditions, and data were collected until at least three runs showed no improvement over time. The final result was taken as the average of the final three runs.

C. Results and discussion

The mean data for the two groups (long-electrode vs A+E) across the two background conditions are displayed in Fig. 2. The most obvious difference between both types of

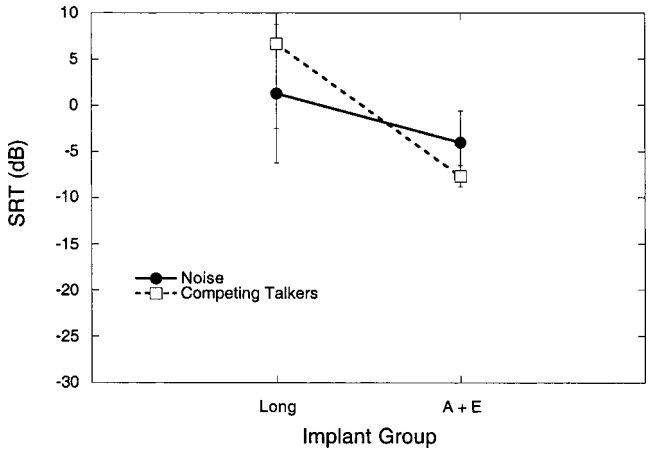


FIG. 2. Group mean SRT values for the two groups of cochlear implant listeners in the two types of background stimuli. The long-electrode group consists of all 20 long-electrode subjects. The error bars represent the standard deviations across subjects for each condition.

cochlear implant users of Fig. 2 and the normal-hearing listeners of Fig. 1 (listening to the same unprocessed stimuli) is that the implant users perform much more poorly than the normals. In steady noise, normal-hearing subjects' SRT's are approximately −15 dB SNR, whereas the implant users are approximately 15 dB poorer. In the competing-talker background, the difference is more striking, with normal-hearing listeners outperforming the traditional implant users by more than 30 dB, and the A+E users by 20 dB for unprocessed stimuli. Even the 16-channel cochlear implant simulation group mean data from the normal-hearing subjects (Fig. 1) are approximately 15 dB better than those of the actual implant users for both steady noise and competing-talker backgrounds.

There are at least several factors contributing to this deficit. First is the general inability of implant users to perform well in noise backgrounds, as shown by Fu *et al.* (1998) and Friesen *et al.* (2001). Typical cochlear implant users do not possess the spectral resolution required to accurately identify speech in noise, and even the 16-channel simulation condition in the present experiment overestimates the spectral resolution of probably all cochlear implant users. A second reason is the particular disadvantage that cochlear implant users show in understanding speech in a competing-talker background, as shown by Nelson *et al.* (2003) and Stickney *et al.* (2003). A third reason is that implant patients typically do not have a full population of surviving auditory nerves, and this can result in a general disadvantage in speech recognition (even in quiet) for electric stimulation as compared to normal-hearing listeners (Fishman *et al.*, 1998).

The A+E patients also showed a deficit compared to the normal-hearing subjects of experiment 1, and several additional factors most likely contributed to this difference. The A+E subjects received only 6 channels of electrical stimulation for the high frequencies, whereas the normal-hearing subjects had much better spectral resolution (even in the simulation which had 13–14 channels). An additional factor may be that the electric stimulation for the A+E patients is directed to a position in the cochlea that is considerably more basal than normal, due to the 10-mm insertion depth of the electrode array. This frequency–place mismatch has been shown to negatively affect speech recognition in combined acoustic and electric hearing (Gantz and Turner, 2003; Brill *et al.*, 2001)

A question of considerable clinical utility is whether the A+E approach offers an advantage over the traditional implant, as suggested by the simulation study of experiment 1. As seen in Fig. 2, the mean SRTs for the A+E subjects were lower than that of the traditional implant user, for both noise and competing-talker backgrounds. Statistical analysis of these data using a mixed-mode ANOVA, with background as a within-subjects factor and implant type as a between-subjects factor, showed a significant interaction between type of implant and background condition ($F=18.85$,df$=1,21$;$p<0.001$). Follow-up *t*-tests indicated that the differences between groups occurred only for the competing-talker condition [$t(21)=2.63$,$p<0.01$] and not for the steady noise [$t(21)=1.18$;$p>0.10$]. These results were in agreement with the outcome of the simulation experiment. The variability across subjects is displayed in detail in Figs. 4 and 5 and discussed below.

The group of 20 long-implant users included a wide range of speech recognition abilities, as is typical for a cochlear implant subject pool. Recognition scores on a test of consonant /aCa/ materials presented in quiet (Turner *et al.*, 1995, Fu *et al.*, 1998) ranged from 13% to 74%, with a group mean of 47%. The three A+E subjects had a mean score on this same consonant test of 63% correct (range 53%–71%). It therefore appears that the long-electrode cochlear implant patients in the previous comparison were not only poorer than the A+E patients for speech in background noises, but also poorer for speech recognition in general. This discrepancy could confound the across-subjects comparisons of Fig. 2, if one is looking for real-patient evidence to support the theoretical concept that preserving residual low-frequency acoustic hearing is advantageous. Therefore, the long-implant patients were subdivided to form a smaller subgroup of subjects that had, on average, the same speech scores in quiet as the A+E subjects. Beginning with the top-performing long-implant user on the /aCa/ test and moving downward in ability, additional subjects were added to form a ''matched subgroup'' until the mean value for the long-implant group was within 1 percentage point of the mean for the A+E group (63%). This matched group contained 10 of the original 20 subjects. The group mean results of this comparison are shown in Fig. 3. As in Fig. 2, the mean values for the A+E group are better than the ''matched'' long-implant group for the competing-talker condition (9-dB advantage). This group comparison was in the same pattern as the previous all-subjects comparison. Using a mixed-mode ANOVA, with background as a within-subjects factor and implant type as a between-subjects factor, a significant interaction between type of implant and background condition was observed ($F=20.76$,df$=1,11$,$p=0.001$). Follow-up *t*-tests indicated that the two groups were not different for steady noise [$t(11)=0.89$;$p>0.5$], but were different for the competing-talker background [$t(11)=1.84$;$p<0.05$]. Thus, even when differences in speech recognition in quiet are accounted for, the A+E approach appears to offer a significant

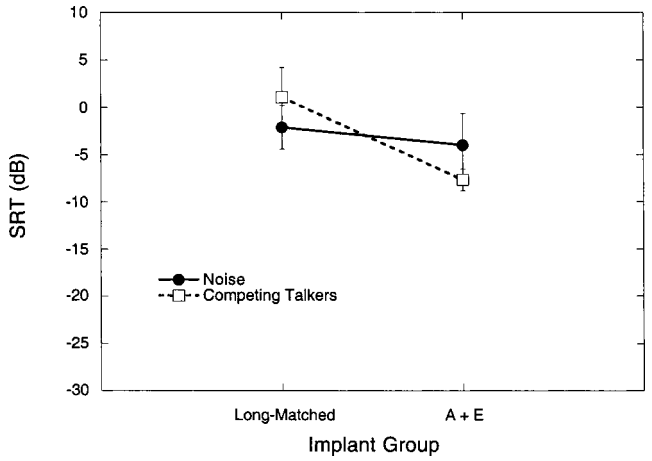


FIG. 3. Group mean SRT values for the two groups of cochlear implant listeners in the two types of background stimuli. The long-electrode group consists of the ten long-electrode subjects matched to the short-electrode subjects in terms of consonant recognition in quiet abilities. The error bars represent the standard deviations across subjects for each condition.

advantage over the long-electrode cochlear implant in a multitalker background.

Figures 4 and 5 display in histogram format the individual data for the traditional implant users and the A+E subjects for the noise and competing-talker backgrounds, respectively. The A+E subjects are indicated by the dark solid bars, the members of the matched group of long-electrode implant users by the hatched bars, and the remaining long-electrode implant users by the open bars. In Fig. 4, the SRT scores in the steady noise are shown. The A+E subjects' data are at the upper end of the entire distribution; however, when compared only to the matched group, their scores are not distinguished. In Fig. 5, the data for speech in the competing-talker background are plotted. In this case, not only are the A+E scores at the upper end of the entire distribution, they are also better than any of the matched group's scores. These raw data also provide strong preliminary support to the idea that preserving acoustic hearing in cochlear implant patients can provide an advantage for understanding speech in a background of other talkers, but not in steady noise.

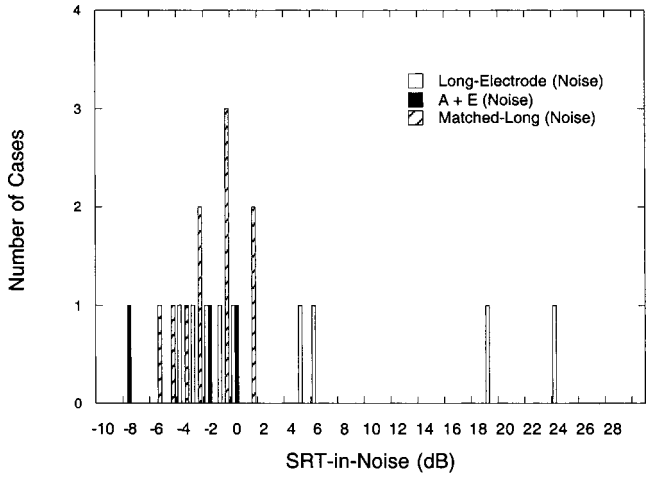


FIG. 4. Individual cochlear implant listeners' SRT values for the condition where the background was steady noise.

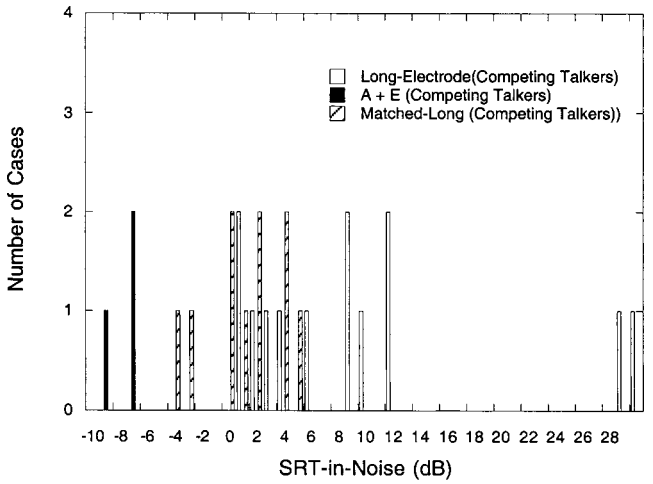


FIG. 5. Individual cochlear implant listener' SRT values for the condition where the background was competing talkers.

IV. GENERAL DISCUSSION

Theoretical advantages of preserving low-frequency acoustic hearing in cochlear implant patients for understanding speech in background noises were presented. A simulation experiment using normal-hearing subjects provided clear evidence that providing unprocessed low-frequency acoustic speech information yielded an advantage for the condition that the background is composed of competing speech. For the present speech and masking stimuli, only a small and nonsignificant advantage was observed for steady noise. These results are in agreement with the idea that the low-frequency acoustic hearing allows the listener to perceive the fundamental frequencies of the talkers and assists in separating the target speech from a background of other talkers. The same task was employed in a group of traditional long-electrode cochlear implant users, as well as three subjects using the acoustic plus electric approach that employs a ''short-electrode'' cochlear implant, which preserves low-frequency acoustic hearing. The acoustic plus electric approach shows significant advantages over the long-electrode cochlear implant for the recognition of speech in multitalker backgrounds, but not in steady noise, similar to the simulation study. While the recognition of speech presented in a background of competing talkers for both groups of cochlear implant patients was certainly poorer than that observed for normal-hearing listeners, the preservation of low-frequency acoustic hearing using the acoustic plus electric device can reduce at least some of the deficit seen for traditional long-electrode cochlear implant users.

ACKNOWLEDGMENTS

Supported (in part) by research grants 1 R01 DC 000377 and 2 P50 DC00242 from the National Institutes on Deafness and Other Communication Disorders, National Institutes of Health; Grant RR00059 from the General Clinical Research Centers Program, NCRR, National Institutes of Health; the Iowa Lions Sight and Hearing Foundation; and the Cochlear

Corporation for developing a cochlear implant to our specifications and providing devices at no cost. Special thanks are also given to Greg Flamme for helpful discussions on this project.

Assmann, P. F. (**1999**). ''Fundamental frequency and the intelligibility of competing voices,'' Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1–7 Aug. 1999, pp. 179–182.

Assmann, P., and Summerfield, Q. A. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Bilger, R. C. (**1984**). ''Speech recognition test development,'' ASHA Reports #14, 2–15.

Bird, J., and Darwin, C. J. (**1999**). ''Effects of a difference in fundamental frequency in separating two sentences,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A., Palmer, A., Rees, Q., Summerfield, and R. Meddis, (Whurr, London).

Brill, S., Lawson, D., Wolford, L., Wilson, B., and Schetzer, R. (**2001**). ''Speech Processors for Auditory Protheses,'' 11th Quarterly Progress Report, NIH Project N01-DC-8-2105.

Brokx, J., and Nooteboom, S. (**1982**). ''Intonation and the perception of simultaneous voices,'' J. Phonetics **10**, 23–26.

Culling, J. T., and Darwin, C. J. (**1993**). ''Perceptual separation of simultaneous vowels: Within and across-formant grouping,'' J. Acoust. Soc. Am. **98**, 3454–3467.

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 processor,'' J. Acoust. Soc. Am. **99**, 1174–1184.

Fishman, K., Shannon, R. V., and Slattery, W. (**1998**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant strategy,'' J. Speech Hear. Res. **40**, 1201–1215.

Friesen, L., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Gantz, B. J., and Turner, C. W. (**2003**). ''Combining acoustic and electric hearing,'' Laryngoscope **113**, 1726–1730.

Geurts, L., and Wouters, J. (**2001**). ''Coding of fundamental frequency in continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.

Gfeller, K., Turner, C. W., Woodworth, G., Mehr, M., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (**2002**). ''Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults,'' Cochlear Implants International, Volume 3(1), pp. 31–55.

Green, T., Faulkner, A., and Rosen, S. (**2002**). ''Spectral and temporal cues to pitch in noise-excited vocoder simulation of continuous interleaved sampling cochlear implants,'' J. Acoust. Soc. Am. **112**, 2155–2164.

Harris, R. W. (**1991**). Speech Audiometry Materials Compact Disk, Brigham Young University, Provo Utah.

Henry, B. A., and Turner, C. W. (**2003a**). ''The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners,'' J. Acoust. Soc. Am. **113**, 2861–2873.

Henry, B. A., and Turner, C. W. (**2003b**). ''Spectral shape perception and speech recognition in normal hearing, hearing impaired, and cochlear implant listeners,'' presented at the Association for Research in Otolaryngology, Palm Beach, FL.

Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (**2003**). ''Speech and melody recognition in binaurally combined acoustic and electric hearing,'' J. Acoust. Soc. Am. (submitted).

Nelson, P., Jin, S.-H., Carney, A., and Nelson, D. (**2003**). ''Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners,'' J. Acoust. Soc. Am. **113**, 961–968.

Qin, M. K., and Oxenham, A. J. (**2003**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,'' J. Acoust. Soc. Am. **114**, 446–454.

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

Stickney, G. S., Zeng, F. G., Litovsky, R. V., and Assmann, P. F. (**2003**). ''Cochlear implant speech recognition with speech maskers,'' J. Acoust. Soc. Am. (submitted).

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

Von Ilberg, C., Keifer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecher, E., and Klinke, R. (**1999**). ''Electro-acoustic stimulation of the auditory system,'' ORL **61**, 334–340.

Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation

John M. Deeks and Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

(Received 5 September 2003; revised 14 January 2004; accepted 23 January 2004)

Two experiments used simulations of cochlear implant hearing to investigate the use of temporal codes in speech segregation. Sentences were filtered into six bands, and their envelopes used to modulate filtered alternating-phase harmonic complexes with rates of 80 or 140 pps. Experiment 1 showed that identification of single sentences was better for the higher rate. In experiment 2, maskers (time-reversed concatenated sentences) were scaled by −9 dB relative to a target sentence, which was added with an offset of 1.2 s. When the target and masker were each processed on all six channels, and then summed, processing the masker on a different rate to the target improved performance only when the target rate was 140 pps. When the target sentence was processed on the odd-numbered channels and the masker on the even-numbered channels, or vice versa, performance was worse overall, but showed similar effects of pulse rate. The results, combined with recent psychophysical evidence, suggest that differences in pulse rate are unlikely to prove useful for concurrent sound segregation. [DOI: 10.1121/1.1675814]

PACS numbers: 43.66.Ts, 43.71.Ky [GK] Pages: 1736–1746

I. INTRODUCTION

Normally hearing listeners can use differences in fundamental frequency ($\Delta F0$s) between concurrent speech sounds to segregate them perceptually (Brokx and Nooteboom, 1982; Scheffers, 1983; Assmann and Summerfield, 1990; Culling and Darwin, 1993). For example, introducing a $\Delta F0$ between the members of a pair of concurrent vowels can lead to a significant improvement in the ability of listeners to identify both members of the pair (Scheffers, 1983). Subjectively, the improvement is often accompanied by an impression of two voices producing different vowels on different pitches (Assmann and Summerfield, 1990). Assmann and Summerfield (1990) presented pairs of concurrent vowels with $\Delta F0$s between them to normal-hearing listeners. They showed that the pattern of vowel identification could be well fitted by a model that combined the effects of filtering in the auditory periphery with an analysis of the temporal pattern of activity predicted to occur in the auditory nerve. A similar spectro-temporal approach has been adopted by Meddis and Hewitt (1992) and by Brown and Cooke (1994). Darwin (1992) has shown that frequency components that are resolved by the peripheral auditory system contribute most to the segregation process. In the normal auditory system, the individual frequencies of these resolved harmonics are encoded both by their place of excitation, and by the temporal responses of subsets of auditory-nerve fibers tuned to those frequencies.

Encoding of $F0$ in a cochlear implant differs from that found in normal acoustic hearing (Moore and Carlyon, 2004). The incoming signal is passed through a bank of filters that are too wide to resolve individual harmonics, and, in most modern strategies, used to modulate a pulse train having a fixed rate. Hence, there is little opportunity for subsets of stimulated fibers to encode the temporal structure of an individual, resolved, harmonic, as occurs in a normal ear. However, for a periodic sound presented in quiet, $F0$ can be represented as temporal fluctuations in the envelope resulting from harmonics interacting within the passband of each channel. Geurts and Wouters (2001) measured just-noticeable differences in synthetic vowel $F0$, by implant users whose device implemented the continuous interleaved sampling (CIS) processing strategy (Wilson *et al.*, 1991). For a vowel with an $F0$ of 150 Hz, their four subjects could hear differences of between 4.0% and 13.3%. However, their task involved the sequential presentation of $\Delta F0$, and it is likely that different processes are involved in the use of $\Delta F0$ to segregate concurrent voices. Indeed, unlike normally hearing listeners, implant users are unable to exploit differences between the gender of a target speaker and that of an interfering speaker, consistent with them being unable to use $F0$ differences for concurrent sound segregation. For these listeners, competing speech can impair performance even at a favorable signal-to-noise ratio of +16 dB (Nelson and Jin, 2002).

Further insight into the difficulties implant users face when attempting to use pitch cues to separate competing sounds comes from acoustic simulations of cochlear implant speech processors (Shannon *et al.*, 1995). Typically, these extract the temporal envelope in each of several frequency regions, and use this envelope to modulate a band of noise or sinusoidal carrier. When listening to such simulations, normally hearing subjects are also able to detect $F0$ differences of a few percent between sequentially presented sounds, provided that these have a reasonably low $F0$ and are not presented in a reverberant environment (Qin and Oxenham, 2003a). However, also consistent with the cochlear implant literature, they are more susceptible to interfering noise than when they listen to unprocessed speech (Qin and Oxenham, 2003b).

One reason for this deficit may lie in the inability of the auditory system to use purely temporal cues to extract the $F0$s of two periodic stimuli that excite the same region of the

 0001-4966/2004/115(4)/1736/11/$20.00

cochlea. This evidence comes from studies in which mixtures of two periodic pulse trains were either applied to a single cochlear implant electrode, or presented to normally hearing listeners after filtering so as to remove low harmonic numbers. In both cases, listeners fail to hear the two underlying pitches but instead report a single pitch corresponding roughly to that of the higher-rate pulse train (Carlyon, 1996; Carlyon *et al.*, 2002). Indeed, contrary to the effects seen with resolved harmonics, listeners do not consistently report pairs of such pulse trains that have widely different rates as sounding less fused than pairs whose rates are more similar (Carlyon, 1996). This suggests that, when two voices interact within a single channel, there should be little potential for implant users to exploit $F0$ differences for concurrent sound segregation.

The situation is slightly more encouraging when pairs of unresolved complex tones are filtered into *separate* frequency regions (Carlyon, 1994). Such a situation might arise when two speakers utter voiced sounds with different $F0$s, and which, fortuitously, contain formants that occupy distinct and well-separated frequency regions. Listeners can detect $F0$ differences between such complexes (Carlyon, 1994), and Darwin (1992) has shown that across-frequency differences in $F0$ can cause a formant to ''pop out'' from the remainder of a voiced syllable, even when the harmonics of that formant is unresolved. Hence, source segregation via $\Delta F0$ may be possible provided the unresolved harmonics excites different regions of the basilar membrane. However, it should be noted that in Darwin's study the popping out of a formant consisting of unresolved harmonics required a much larger $\Delta F0$ than when the harmonics was resolved, and, importantly, there was no evidence that this popping out affected the phonetic identity of the remainder of the syllable.

In the present study we investigated the extent to which, in the absence of resolved harmonics, normally hearing listeners can use pitch cues to segregate concurrent speech sounds. Specifically, we investigated the use of these ''purely temporal'' pitch cues under conditions where two competing sources occupied either completely overlapping or nonoverlapping, interleaved frequency ranges. To do so, we modified the noise-vocoding simulation of cochlear implant speech processors by replacing the noise-band carriers with bands of unresolved harmonics. This resulted in a waveform in each frequency channel that resembled a modulated pulse train. Spectral overlap of the two input waveforms was then manipulated by either processing both sentences on all six channels of the simulation and mixing them together, or presenting one sentence on the odd-numbered channels alone and the other on the even-numbered channels. The temporal pattern of stimulation was manipulated by using carrier harmonic complexes with either the same or different $F0$s for the two speech sounds. In the simulation, the pitch cues were provided by the carrier $F0$s used for each source, which were constant during a given sentence. This differs from previous approaches in which the ''true'' $F0$ of a single voice was tracked on a moment-by-moment basis and reflected in the rate of trains of pulses or noise bursts (Blamey *et al.*, 1984; Faulkner *et al.*, 2000). By holding pulse rate constant for a given source, we aimed to maximize the chances of observing any effects of a rate difference between sources on concurrent sound segregation (Carlyon *et al.*, 2000). This also allowed us greater control of the resolvability of the pulse train, and hence of the validity of the simulation as a model of electric hearing.

The present study had two aims. First, we wished to determine the extent to which the psychophysical findings previously obtained with rather simple stimuli generalized to the perceptual segregation of competing sentences. Specific predictions arising from these findings are described in Sec. IV A, which also describes the conditions of our main experiment in more detail. Second, we wished to investigate whether, and how, implant users might be able to segregate concurrent sounds. The processing scheme we used took as its input two sources that have already been separated, and so could not, by itself, be implemented in a real-world device. Section V C briefly discusses one way in which automatic source segregation might be achieved. More generally, by gaining control of the temporal representation of competing speech sounds in the auditory periphery, the present experiments probe the temporal-processing limitations of the auditory system when resolved harmonics are absent.

II. DESCRIPTION AND VALIDATION OF SIGNAL-PROCESSING TECHNIQUE

A. Overview

The aim of our new signal-processing technique was to introduce a pitch cue that was encoded by purely temporal means, analogous to that produced by the pulse rate in a cochlear implant (Carlyon *et al.*, 2002). In particular, we wished to exclude any ''place of excitation'' cues to pitch. To do so, we modified the popular ''noise vocoder'' simulation of cochlear implant speech processors (Shannon *et al.*, 1995) by replacing the noise carrier with a harmonic complex (Fig. 1). Although the raw waveform of harmonic complex tones can, depending on the phase spectrum, resemble a pulse train, filtering by the normal cochlea can result in the place of excitation cues that we wished to avoid. To overcome this, a number of further modifications had to be made.

Place of excitation cues can be minimized by passing a complex with a low fundamental frequency ($F0$) through bandpass filters having relatively high center frequencies, at which auditory filters are broadest (Shackleton and Carlyon, 1994). Psychophysical experiments using this approach have yielded results that are similar to those obtained with single-channel cochlear implant simulations (Carlyon, 1996; McKay and Carlyon, 1999; Carlyon and Deeks, 2002; Carlyon *et al.*, 2002). We therefore used $F0$s of (in different conditions) 40 and 70 Hz, and did not use any analysis channels having center frequencies (CFs) below 1089 Hz (Table I). Each of these manipulations had a potentially undesirable consequence.

First, the use of low $F0$s meant that the ''pulse rate'' in the carrier signal might be too low to adequately sample the signal envelope. To alleviate this, harmonics were summed in alternating phase (Patterson, 1976; Shackleton and Carlyon, 1994), leading to pulse rates of 80 and 140 pps (double

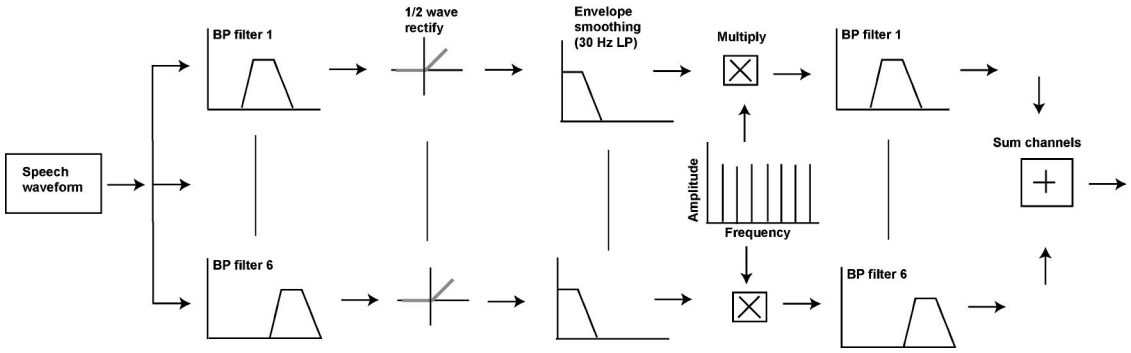


FIG. 1. Processing of ASL stimuli into experimental stimuli. The sentence waveform is analyzed using three or six bandpass filters. The output of each of these is half-wave rectified and smoothed using a 30-Hz low-pass filter. The resulting waveform is then multiplied with the carrier (a harmonic source with rate=80 or 140 pps) before passing through the bandpass synthesis filters. These outputs are then summed.

the respective $F0$s). The signal envelope in each band was low-pass filtered at 30 Hz (4th-order Butterworths, giving 24-dB/octave attenuation rate), to reduce any aliasing effects in sampling the envelope even at the lower pulse rate. This also had the effect of eliminating any envelope fluctuations at rates equal to the $F0$ of the input voice, so the dominant ''low'' pitch of the complex was determined by the carrier pulse rate. The bandpass filters used for each channel were sixth-order Butterworths, with attenuation rates outside the passband of 18 dB/octave.

Second, eliminating analysis filters below 1089 Hz meant that we either had to discard speech information below that frequency, or spectrally shift the speech by introducing a mismatch between the analysis and synthesis bandpass filter in each channel. We chose the former option, as informal listening revealed that it produced a much smaller decrement in performance (cf. Fu and Shannon, 1999; Baskent and Shannon, 2003).[1] The root-mean-square levels in each band of the processed stimuli were set to the same level, thereby ''whitening'' the spectrum, and the overall level of the processed speech was 57.3 dB SPL. Finally, we added a continuous low-pass noise in order to mask distortion products having a frequency equal to the pulse rate (Pressnitzer and Patterson, 2001). It was generated by passing a white noise through a low-pass filter (Kemo VBF25.03; attenuation 48 dB/octave outside passband) having a corner frequency of 400 Hz and a spectrum level, within its passband, of 27.9 dB SPL.

To illustrate the temporal pitch code introduced by the scheme, a white noise was processed in the same way as the speech stimuli in the main experiment, with a pulse rate of 80 pps. The output was then passed through the peripheral stages of the Auditory Image Model (Patterson *et al.*, 1995). The resulting basilar-membrane motion (BMM), which demonstrates the effects of gammatone filtering, is shown in Fig. 2(a). It can be seen that all channels show a periodicity at a rate of 80 pps. Figure 2(b) shows the BMM in response to a condition, described in Sec. III B, in which the even-numbered channels of the processor output were deleted.

TABLE I. Frequency properties for each channel used in the simulations (kHz). The last column shows the number of components in the passband of the channel when $F0=70$ Hz.

Band	High-pass	Center	Low-pass	Bandwidth	Comps in BW
1	0.937	1.089	1.261	0.324	4
2	1.261	1.457	1.680	0.419	5
3	1.677	1.933	2.221	0.544	7
4	2.221	2.549	2.922	0.701	10
5	2.922	3.346	3.828	0.906	12
6	3.828	4.376	5.000	1.172	16

B. Pitch preference judgments for sine- and alternating-phase single-channel stimuli

To check that the pitch conveyed by the pulse trains produced by the signal processing was purely temporal in origin, a preliminary control experiment was performed. All stimuli in this preliminary experiment were processed in the same way as the speech stimuli in the main experiment, except only the lowest-frequency channel, where resolved harmonics are most likely to occur, was used (see Table I). In addition, the harmonics used for the carrier could be summed in either alternating (ALT) or sine (SIN) phase. The input was always 0.5 s of white noise, with 20-ms raised-cosine onset and offset ramps. Each trial involved the presentation of three stimuli, the first of which was always an ALT-phase stimulus with $F0$ of either 40 or 70 Hz. The other two stimuli were always in SIN phase; one had an $F0$ equal to that of the first sound, and one had an $F0$ an octave higher. Shackleton and Carlyon (1994) have shown that, if the harmonics are unresolved, subjects should select this latter stimulus as having a pitch more like the first; in contrast, if the harmonics are resolved, the SIN-phase stimulus having an $F0$ equal to that of the first sound should be judged as more similar.

Table II shows the percentage of trials on which each subject reported that the ALT-phase complex had a pitch more like that of the SIN-phase complex having an $F0$ one octave higher. The results clearly show that the pitch arising from the lowest channel of the processed stimuli is one octave above the $F0$ of the complex, corresponding to a repetition rate of 80 or 140 pulses per second. In the remainder of this article we describe the carrier in terms of its pulse rate, rather than its $F0$.

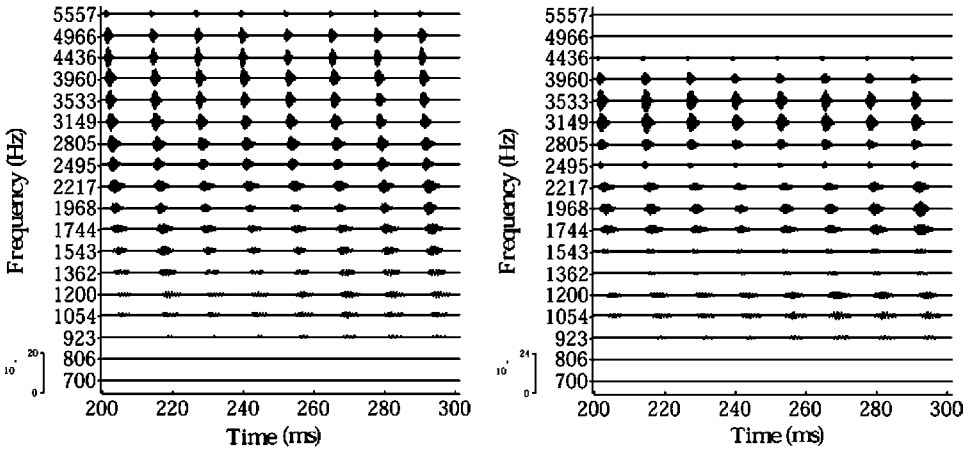


FIG. 2. Simulation of basilar-membrane motion (Patterson *et al.*, 1995) in response to white noise. (a) with all channels active. (b) with only odd-numbered channels active.

III. EXPERIMENT 1. PERFORMANCE WITH A SINGLE SENTENCE AS A FUNCTION OF STIMULATION RATE AND NUMBER OF CHANNELS

A. Rationale

The purpose of experiment 1 was to investigate the effects of two manipulations—pulse rate and number of channels—on the perception of speech processed using the new scheme. These data were then used when interpreting the effects of experiment 2, which involved mixing target and masker sentences with the same or different pulse rates, and under conditions where the number of target channels differed.

B. Method

1. Subjects

Sixteen normally hearing listeners (11 females and 5 males) took part. Thresholds in quiet for pure tones at 0.5, 1, 2, and 4 kHz were less than 20 dB HL in both ears for all subjects, except for two subjects with thresholds of 27.1 and 28.1 dB HL at 2 kHz in one ear. All subjects had English as their first language. Ages ranged from 18 to 36 years. All had some prior experience with listening to distorted speech sounds, but not with the type involved in this experiment.

2. Stimuli and procedure

The speech material was taken from the MRC Institute of Hearing Research Adaptive Sentence Lists (ASL) (MacLeod and Summerfield, 1990). These sentences were based on the BKB sentences (Bench and Bamford, 1979) and produced by a male speaker of southern British English. Each sentence was scored for three keywords, using the ''loose'' scoring technique (Bench and Bamford, 1979). Each subject was presented with one sentence list per condition. As there were 15 sentences in each list, and 3 keywords per sentence, this yielded a score out of 45 for each condition.

The task for all subjects was to listen to each processed sentence and report (via keyboard entry) as many words as they could. Subjects clicked a button to move on to the next trial. Conditions were tested in blocks of 15 sentences, with short breaks taken between blocks. Testing took place in a double-walled sound-attenuating booth containing headphones, a mouse, and keyboard, and within sight of a large monitor.

The experiment measured speech reception performance with respect to combinations of two factors: speech processed into three or six channels, and with stimulation rates of 80 or 140 pps. Subjects were tested in a repeated-measures design across the four conditions, and were randomly assigned to one of four groups. The order of testing across these groups was counterbalanced. For three-channel conditions, half the subjects in each group were tested on odd-channel simulations, and half on even-channel simulations. The sentence lists used were also counterbalanced across conditions and groups. Hence, each group experienced a different sentence list for each condition, but, averaged across groups, each sentence was used an equal number of times for every condition. This was done to ensure that any differences in performance across condition were due to the processing rather than to any coincidental differences in difficulty between sentence lists. (Although the ASL lists are equated for difficulty across sentences, this equating was not performed with the type of processing used here.) The order of sentence presentation was randomized in each block for every subject and condition.

TABLE II. Percentage of trials in which the ALT-phase $F0$ complex was judged to have a pitch more like the SIN-phase $2F0$ complex than that of the SIN-phase $F0$ complex.

	Reference $F0$ (Hz)	
Subject	40	70
1	96.67	98.33
2	96.67	100.00
3	100.00	100.00

C. Results

Figure 3 shows the group mean percent correct for each condition. For the three-channel conditions, data from the odd- and even-channel subgroups were averaged, as they produced very similar results (odd vs even=35% vs 32% at 80 pps and 48% vs 43% at 140 pps).

Overall, performance was better for the six-channel conditions than for the three-channel conditions, and better at a rate of 140 pps than of 80 pps. These trends were confirmed by a repeated-measures ANOVA, having within-subject fac-

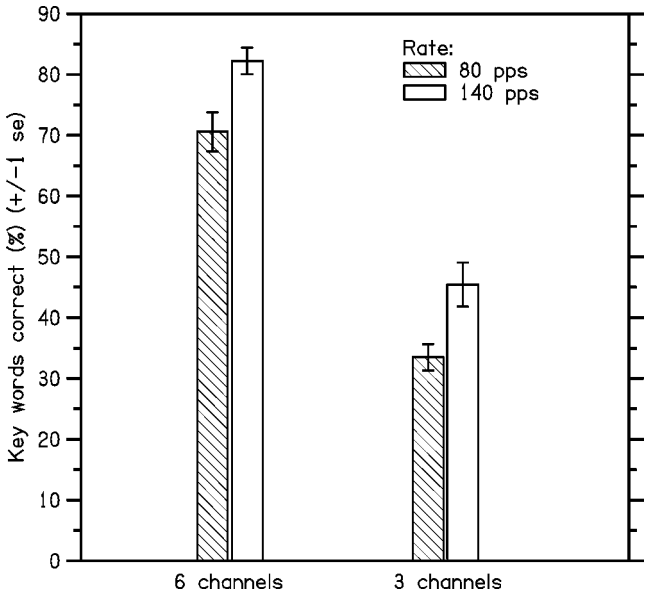


FIG. 3. Group mean scores (%) for single sentence conditions (±1 standard error).

tors of channel (three or six channels) and rate (80 or 140 pps), and a between-subject factor of group (odd or even channel subgroups for the three-channel conditions). Main effects of channel ($F_{(1,14)}=246.2$, $p<0.001$) and rate ($F_{(1,14)}=25.0$, $p<0.001$) were significant. There were no significant interactions and no effect of group.

D. Discussion

The experiment showed that higher carrier rates led to significantly higher speech reception scores. There have been no previous studies investigating the effect of pulse rate with normally hearing listeners and acoustic simulations, but a number of researchers have investigated the effect in cochlear implant listeners. Although the majority of these studied used pulse rates that exceeded the highest value of 140 pps employed here, Fu and Shannon (2000) investigated the range between 50 and 500 pps in four users of the Nucleus 22 implant. They implemented a four-channel CIS strategy, and, consistent with our results, found that phoneme recognition improved as stimulation rate was increased from 50 to 150 pps. Performance did not improve with further increases in rate, although Loizou *et al.* (2000) found that consonant and word recognition improved with increases in rate from 400 to 2100 pps.

Experiment 1 also showed that performance deteriorated as the number of channels was reduced from six to three. This was implemented by dropping alternate channels [Fig. 4(a)], unlike previous studies, in which a reduction in channel number was produced by broadening the analysis and carrier filters [Fig. 4(b)]. This effectively leads to ''holes'' in the speech spectrum, and the drop in performance can be attributed to the resulting loss of information, rather than to a loss of spectral resolution. In addition, it may be that some central limitation prevented subjects from effectively combining information across spectral bands separated by gaps. This possibility is suggested by the finding that introducing noise into (albeit much wider) spectral gaps can improve sentence recognition scores (Warren *et al.*, 1997).

IV. EXPERIMENT 2: SENTENCE SEGREGATION USING PULSE-RATE AND CHANNEL DIFFERENCES

A. Rationale and overview

Experiment 2 studied the ability of listeners to use pulse-rate and/or channel differences in segregating concurrent sentences. It used ''target'' sentences similar to those of experiment 1, and added them to a masker which started 1.2 s before the target. The target and masker were processed on

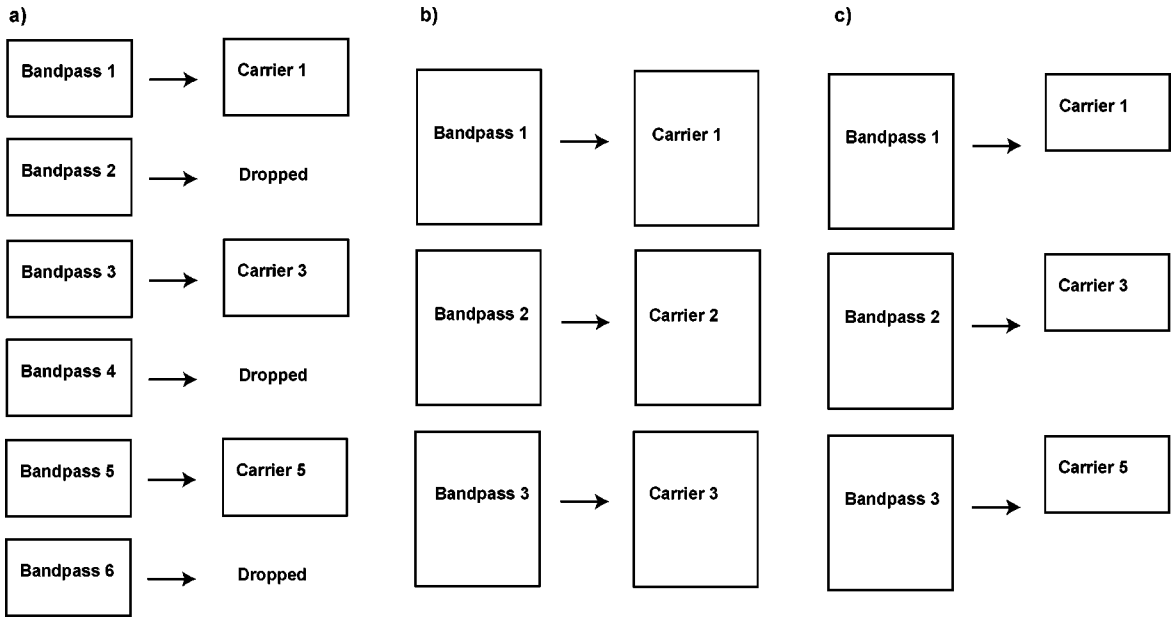


FIG. 4. Relation between analysis filters and carriers in three processing schemes. (a) Method used in this experiment, whereby alternate channels are dropped. (b) Conventional method, whereby fewer channels are accompanied by wider analysis filters. (c) Alternative method, whereby the entire frequency range is analyzed but still only narrow carrier bands are used (resulting in spectral mismatch).

all six channels and then mixed, or presented on three interleaved channels each. Orthogonal to this manipulation, the target and masker could be processed on either the same or different pulse rates.

As noted in the Introduction, previous psychophysical research suggests two ways in which rate differences could affect performance. First, when two equal-level pulse trains are mixed into the same channel, listeners hear a pitch equal to that of the higher-rate train (Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003). Combined with the fact that the target sentence was added to the masker after some delay, this leads to an interesting prediction. When the masker is processed at a rate of 80 pps, introducing a 140-pps target should cause a change in pitch, perhaps aiding segregation of the two sources. In contrast, when the masker is processed at 140 pps, introducing an 80-pps target may leave the pitch unchanged, at least when the masker and target have the same level. This suggests that, when the target and masker levels are equal, introducing a rate difference between them should only help when the target is processed at the higher rate. In fact, as discussed below, the masker and target levels varied over time, and the target was on average more intense than the masker. Hence, as discussed further in Sec. V B, we might expect some advantage of a rate difference when the target is processed at the lower rate, but we would still expect the effects of a rate difference to be smaller than when the target is on the higher rate.

A second possible effect of a pulse rate difference occurs when the target and masker are processed on different channels. As discussed in the Introduction, it is possible that listeners could exploit a rate difference by comparing the temporal pattern of stimulation across different regions of the basilar membrane (Darwin, 1992; Carlyon, 1994). We would expect segregation based on such a cue to be symmetric, in that it should not depend on whether the target is processed at the higher or at the lower rate.

B. Method

1. Stimuli

The sentence material was again taken from the ASLs. Target sentences were taken from different lists than the previous experiment, and were mixed with maskers. The maskers were constructed from three concatenated ASL sentences, which were time reversed and truncated to a duration of 3.5 s, with 20-ms Hanning onset and offset ramps. The type of masker was chosen so as to have many of the spectral and temporal characteristics of interfering speech, while itself being unintelligible. This latter characteristic avoided problems associated with scoring responses that corresponded either to whole words in a competing sentence, or to composites of words from a target and competing sentence. None of the sentences used to construct the maskers had been used in the first part of the experiment, and each target sentence was combined with a masker constructed from a unique set of three ASL sentences.

Both the target and masker were processed in the same way as described in the previous experiment. They were each separately processed into three or six channels with the desired carrier rate (depending on the condition). The masker was then attenuated by 9 dB relative to the target sentence, and the target was added to the masker with an offset of 1.2 s. Because this value is an integer multiple of the periods of the 80-pps and 140-pps pulse trains, the pulses from each source were simultaneous when the target and masker were processed on the same rate. This gave rise to a single pulse train in each frequency region, that was modulated by an envelope derived from the sum of the filtered masker and target waveforms. All stimuli were generated at a sample rate of 22 050 Hz with 16-bit resolution.

TABLE III. Parameters used in conditions 1 to 4 of experiment 2.

	Target sentence		Masker sentence	
Condition	Rate	Chans	Rate	Chans
1a	80	All	80	All
1b	140	All	140	All
2a	80	All	140	All
2b	140	All	80	All
3a	80	Odd	80	Even
3b	80	Even	80	Odd
3c	140	Odd	140	Even
3d	140	Even	140	Odd
4a	80	Odd	140	Even
4b	80	Even	140	Odd
4c	140	Odd	80	Even
4d	140	Even	80	Odd

Stimulus presentation was through the same headphones as for experiment 1. Subjects were again seated in the double-walled sound-attenuating chamber containing the headphones, keyboard, and mouse, within sight of the computer monitor. They were instructed to report back as much of the target sentence as they could. They were informed that the target sentence would start later than the interfering sound, and that it might be accompanied by an increase in loudness.

2. Subjects, design, and procedure

The experiment investigated factors of number of channels (all six, and odd- or even-channels only) and rate (80 and 140 pps) with respect to target and masker. The same subjects who participated in experiment 1 took part.

Table III shows the structure of conditions. In condition 1, both target and masker were processed into six channels, and both had the same carrier rate (either 80 or 140 pps). In condition 2, both target and masker were processed into six channels, but this time had different carrier rates (target=80 pps; masker=140 pps, or vice versa). Condition 3 involved simulations with targets and maskers processed on odd or even channels only (target=odd, masker=even, or vice versa). Both target and masker had the same carrier rate, which was either 80 or 140 pps. Condition 4 involved simulations with targets and maskers having different channels and carrier rates.

Subjects were randomly assigned to one of four groups. For each group, a different set of sentences was used for every condition. However, by counterbalancing the alloca-

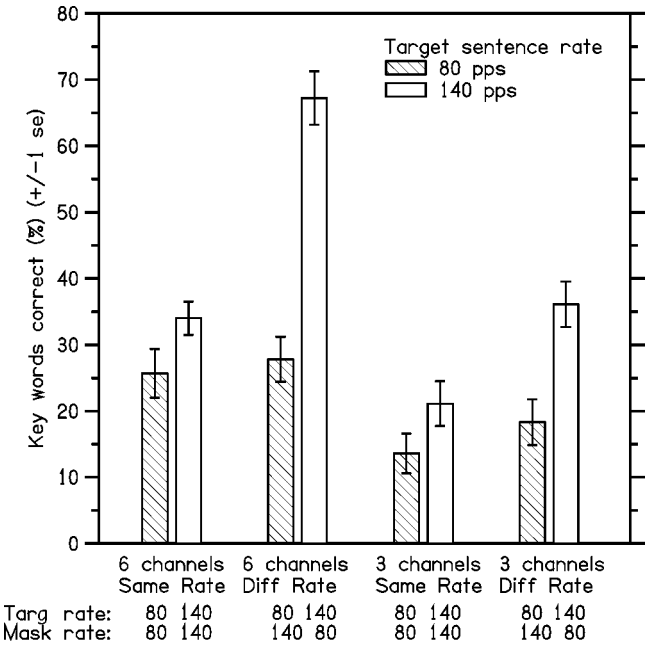


FIG. 5. Group mean scores (%) for target and interferer sentence conditions (±1 standard error).

tion of sentence lists to condition across groups, each sentence was used an equal number of times (although for different subjects) for each condition. The order in which the different conditions were run was also fully counterbalanced across groups. Subjects were tested in a total of eight conditions. All 16 subjects performed conditions 1a, 1b, 2a, and 2b. For conditions 3 and 4, subjects were further divided into two subgroups, depending on whether odd- or even-numbered channels were used for the target. Hence eight subjects did conditions 3a, 3c, 4a, 4c, and eight did 3b, 3d, 4b, and 4d.

C. Results

As with experiment 1, differences between performance in odd- and even-channel subgroups in the three-channel conditions were small, and data were averaged across the odd- and even-channel sub-groups. Figure 5 shows the group mean and standard error for each condition. Overall, performance is worse than that found for single-sentence tasks (Fig. 3). The remaining trends will be discussed with reference to the results of a repeated measures ANOVA applied to the data. The ANOVA had within-subject factors of channels (three vs six), target rate (80 vs 140 pps), and rate difference (target rate same or different to masker rate). It also used a between-subjects factor of group (odd or even channels for three-channel conditions), which was found not to be significant.

As expected from the results of experiment 1, performance was better at a target rate of 140 pps than of 80 pps (open vs filled bars: $F_{(1,14)}=97.0$, $p<0.001$). It was also better for six than for three channels ($F_{(1,14)}=60.3$, $p<0.001$), indicating that any effect of spectral separation between the target and masker channels was swamped by the reduction in conveyed information produced by deleting alternate channels. This latter reduction was substantial, producing an average decrease in performance from 37% to 23% for masked speech (experiment 2), compared with 76% to 39% correct for speech in quiet (experiment 1).

Performance was also better overall when the target and masker were processed on different rates ($F_{(1,14)}=39.3$, $p<0.001$). This difference occurred only when the target was processed on the higher rate (open bars), leading to a interaction between target rate and rate difference ($F_{(1,14)}=26.74$, $p<0.001$).[2] As was discussed in Sec. IV A, this is consistent with the psychophysical evidence that, when two pulse trains interact in the same channel, the pitch is determined by the higher of the two pulse rates. Two further points are worth making about this interaction. First, it was not the case that introducing a rate difference *decreased* performance when the target was processed on 80 pps, as would have occurred if the 140-pps stimulus simply produced more excitation in the auditory periphery. Hence, by introducing a rate difference, it is possible to improve performance when the target is processed on the higher rate, without any cost when processed on the lower rate—at least for conditions like those used here, where the target was more intense than the masker. Second, it occurred not only when the target and masker were presented on the same six channels, but also when they were processed on three interleaved channels each. Specifically, a rate difference failed to improve performance in the three-channel condition when the target was processed on the lower rate. Hence, there is no evidence for the symmetric improvement caused by introducing a rate difference that one might expect if an across-frequency rate comparison aided the perceptual segregation of concurrent speech sounds (Sec. IV A).

The effect of a rate difference when the target and masker were presented on separate channels was qualitatively similar to that observed when they shared all six channels. This may have been due to spread of excitation on the basilar membrane, causing the responses to target and masker to overlap partially. This is illustrated in Fig. 2(b), which shows the output of a bank of auditory filters in response to a white noise processed only on the three odd-numbered channels. It can be seen that although there was no energy in the even-numbered channels of the stimulus, there is some activity in the outputs of auditory filters having CFs within the passbands of those channels. For example, auditory filters centered on 2805 and 3060 Hz both fall within the passband of even-numbered channels (Table I), yet show a nonzero output. This issue is discussed further in Sec. V C.

Finally, we should note that the effect of a rate difference was actually *smaller* in the three- than in the six-channel conditions—as revealed by a significant interaction between the effects of a rate difference and of the number of channels ($F_{(1,14)}=25.84$, $p<0.001$). We can think of two reasons for this interaction. First, it may be an artifact of the finding that performance was lower overall in the three-channel conditions. It could be that the improvement from 17% to 27% (averaged across rate) in the three-channel condition actually reflected a similar change in underlying sensitivity as the numerically larger increase (from 30% to 48%) in the six-channel condition. This would occur, for example, if low levels of performance were partially affected by a

floor effect. If this were true, we might expect across-subject standard deviations to be smaller over this range. To test this, we calculated the regression between standard deviation and percent correct for all subjects and all conditions of experiments 1 and 2. The resulting regression had a slope of only -0.004 ($r^2=0.02$), which was not significantly different from zero—arguing against a floor effect. An alternative explanation is that, when the competing stimuli were filtered into interleaved channels, there was less interaction between them. Because the dominance of the pitch of higher-rate pulse trains has only been observed when the components fall within a single channel (Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003), this effect would be expected to depend on the extent of such within-channel interactions. Where such interactions are reduced, whatever segregation does occur might be expected to depend on spectral separation between the masker and target, rather than on differences in pulse rate.

V. DISCUSSION

A. Overview

This article presents results from a new signal-processing strategy that reproduces several features of the peripheral pattern of stimulation produced by cochlear implants. In common with previous simulations (Shannon *et al.*, 1995) it extracts the envelope in a number of fairly broad frequency bands, and uses each envelope to modulate a carrier signal having an appropriate frequency content. The novel aspect lies in being able to modify the temporal aspects of the carrier, and, as our preliminary experiment showed, to do so without introducing resolved frequency components. This allowed us to study the effects of varying the ''pulse rate'' in each channel, and of introducing a difference between the pulse rates applied to competing sources. It resulted in a number of new findings, some of which have implications for concurrent sound segregation in general, and, in particular, for how this may be achieved by cochlear implant users. Furthermore, some aspects of the data on the effects of pulse rate differences between the target and masker were interpreted in terms of recent evidence on the basic psychophysics of ''purely temporal'' pitch perception.

B. Effects of pitch differences and neural refractoriness

1. Comparison to previous data on temporal pitch perception

One of the most important findings from the present study is that processing a masker on a different rate from the target aids performance only when the target rate is higher than the masker rate. As discussed in Sec. IV A, this is roughly consistent with psychophysical evidence that, when two pulse trains of rates $R1$ and $R2$ are mixed on the same channel in acoustic or electric hearing, subjects hear a single pitch roughly equal to $R2$ (Carlyon, 1996; Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003). This should result in a large increase in pitch when a 140-pps target is added to an 80-pps masker (condition ''T140/M80:'' new pitch about 140 pps), but a much smaller (or no) change when an 80-pps target is added to a 140-pps masker (T80/M140: pitch stays close to 140 pps). One caveat is worth mentioning when relating the two sets of findings: The pairs of pulse trains that were mixed together in the psychophysical experiments had equal levels, whereas here the level of the target, averaged over its total duration, was 9 dB more intense than that of the masker. Presumably, there will be some SNR at which the target will dominate the pitch even when it is processed at a lower rate than the masker. However, it is worth noting that the levels of both target and masker varied throughout the utterances, and so there will have been instances where the SNR was lower than 9 dB (and some where it was higher). Hence, the psychophysical data would predict a trend in the direction observed here, although it is perhaps a little surprising that a rate difference had no beneficial effect at all when the target was processed on the lower rate.

C. Implications for concurrent sound segregation

1. Overview

As pointed out in the Introduction, cochlear implant users currently experience particular difficulty in understanding speech when competing sounds are present. Part of this problem may well be due to the incomplete survival of peripheral processes. However, because similar limitations are observed with acoustic simulations of cochlear implant speech-processing algorithms, it seems likely that these algorithms do not faithfully transmit cues that are important for concurrent sound segregation. As discussed in the Introduction, pitch cues are likely to fall into this category.

There are three broad classes of potential solution to this problem. First, it is possible that, without explicitly segregating the sources *a priori*, monaural cues to concurrent sound segregation could be introduced in new speech-processing strategies, perhaps combined with new electrode designs that restrict the spread of excitation along the neural array. For example, one could attempt to reintroduce the combined place-of-excitation and timing coding of resolved harmonics. However, as pointed out by Moore and Carlyon (2004), this seems unlikely to be achieved in the foreseeable future. A more promising and immediate solution is provided by bilateral implants. Although this approach does not currently allow significant use of interaural timing cues, substantial advantages can be produced by the head shadow effect (e.g., van Hoesel and Tyler, 2003). One drawback, however, is that bilateral implantation is unlikely to be economically justifiable in the medium term, at least in countries whose healthcare system is primarily publicly funded (Summerfield *et al.*, 2002). As van Hoesel and Tyler (2003) have pointed out, the head-shadow effect raises the (cheaper) possibility of using two microphones to route two separate sources to a single implant. They also pointed out that such a strategy, unlike bilateral implants, would not allow the listener to select each source at will. Indeed, this is a problem for any scheme which performs automatic source segregation: for the listener to take advantage of it, the two sources must be encoded in a way which then allows them to be perceptually segregated, and for the listener to switch between sources as required. The present experiments implemented an acoustic analog of

one such encoding scheme, in which different sources are applied to different subsets of electrodes, and with different carrier rates applied to electrodes encoding different sources.

2. Across-frequency rate differences

We believe that the results of the present study should lead one to be cautiously pessimistic about the use of across-channel differences in pulse rate to allow perceptual segregation of different sound sources. The only significant advantage gained by presenting the target and masker on different rates occurred when the target was on the higher of the two possible rates. Performance was substantially better when the target was processed on 140 pps and the masker on 80 pps (T140/M80) than in condition T140/M140 (both processed on 140 pps). However, in a real-life device, where the target and masker are not specified beforehand, the choice is likely to be between processing the two sources on different rates or processing them both on the higher of the two rates. Our results indicate that the former option could lead to worse performance than the latter, as listeners were worse in condition T80/M140 than in condition T140/M140 (compare the third and second bars from the left in Fig. 5).

Experiment 2 showed only a small effect of rate differences when the target and masker were processed on separate channels. Furthermore, this advantage showed the same asymmetry as occurred when they were both processed on all six channels, consistent with it being due to spread of excitation between adjacent channels. The center frequencies of adjacent channels shown in Table I were separated by an average of about 23%. Hence, an auditory filter centered on one frequency band would respond to components in the middle of a neighboring band with an attenuation of approximately 30 dB (Patterson *et al.*, 1982). However, there would also have been auditory filters centered on the boundary between two channels, and, when the masker and target were interleaved, these would have responded about equally to the two sources.

Although we cannot rule out the possibility that a wider channel separation would have allowed subjects to use across-channel rate differences more effectively, we think it unlikely that such a process could aid concurrent sound segregation in existing cochlear implants. The rate differences used here were large and constant, and yet the pattern of results was consistent with a purely within-channel effect. Unless an implant were able to produce substantially more effective between-channel attenuation than the 30 dB in the present study, it seems unlikely that, even if across-channel rate differences were usable, this could effectively be realized. This stands in marked contrast to the situation in normal acoustic hearing, where resolved harmonics are present, in which case such across-channel processes can have a marked effect (Broadbent and Ladefoged, 1957; Darwin, 1992).

3. Assigning sources to different channels

An additional finding of experiment 2 was that performance was consistently worse when the target and masker were each presented to three (separate) channels than when they were mixed into the same six channels. As noted in Sec. IV C, this indicates that any advantage of spectrally separating the two sources was swamped by the deleterious effects of reducing the number of channels per source. One might be tempted to conclude that any attempt to separate two sources on a channel-by-channel basis is doomed to failure, because it would reduce the number of channels for each source. However, there are two reasons why this is not necessarily so. First, the effects of halving the number of channels should decrease, the more channels there are to start with (Friesen *et al.*, 2001). Second, the effects of halving the number of channels may be reduced by adopting a slightly different method of splitting the two sources than the one used here [Fig. 4(a)]. Note that the ''standard'' method, of doubling the bandwidths of both the analysis and carrier channels [Fig. 4(b): Shannon *et al.*, 1995], would not be appropriate, as alternate channels would be needed for separate sources. However, one could double the bandwidth of the analysis channels, while using the same (narrow) carrier channels as in the present scheme—as shown in Fig. 4(c). We intend to explore this issue in a future study.

4. Conclusions

In summary, the absence of an advantage produced by a difference in pulse rate when the target is presented on the lower of two rates suggests that rate differences are unlikely to provide a consistent cue for segregation produced by a real-world device. However, although the present study showed no advantage of spectrally separating the two sources, it is possible that such an advantage could be observed when more channels are available. For implant users to take full advantage of such a strategy, it is likely that they would have to exploit cues to concurrent sound segregation other than differences in pulse rate used here. Such cues might include the natural across-frequency covariations in amplitude present in each source.

VI. SUMMARY

(i) We have described a new simulation of cochlear implant hearing in which the envelope in each frequency band modulates a bandpass filtered pulse train. Because the frequency components of the carrier are unresolved in the auditory periphery, this allows one to simulate the effects of varying pulse rate in electric hearing. Pitch can be manipulated independently of the $F0$ of the input.

(ii) Performance is better when the target is processed at a rate of 140 pps than at 80 pps, and when processed on six than on three channels. Both of these findings occurred for speech in quiet and in the presence of a masker.

(iii) Presenting the target on only the odd-numbered channels and the masker on even-numbered channels (or vice versa) produced worse performance than processing them both on all six channels. Hence, for the particular stimuli used here, any benefits of spectral separation were outweighed by the reduction in number of channels conveying the target.

(iv) There was no evidence that listeners can exploit across-channel differences in pulse rate to identify a target sentence in the presence of a masker.

(v) When the target and masker are processed on the same channels, and the target level is higher, performance can be improved by processing the masker on a lower rate than the 140-pps target (compared to when they are processed on the same rate). However, processing the masker on a higher rate than the 80-pps target neither helps or hinders, again compared to the case where both are processed on the same rate.

ACKNOWLEDGMENTS

The signal-processing scheme used here was based on a MATLAB implementation of the noise-vocoder algorithm, based on code written by Philipos Loizou and subsequently modified by Stuart Rosen. We thank Johannes Lyzenga for further modification of the code, and for advice on its implementation. We also thank Johannes Lyzenga, Christopher Long, and Fan-Gang Zeng for helpful comments on an earlier draft of this manuscript.

[1]Recent evidence has shown that speech perception in noise can be improved by presenting *unprocessed* low-frequency sound to implant users having low-frequency hearing in either the implanted or contralateral ear (Kong *et al.*, 2003; Turner *et al.,* 2003). However, performance is poor when the low-frequency information is presented via an implant or an acoustic simulation thereof (Nelson and Jin, 2002; Kong *et al.*, 2003), presumably due to the lack of resolved harmonics (see the Introduction). It therefore seems unlikely that low-frequency *information* plays an especially important role in implants or in experiments such as ours, in which resolved harmonics are absent.

[2]A two-way ANOVA performed only on the data with the 80-pps target revealed an effect of number of channels, but no effect of having the target on the same vs a different rate ($F_{(1,15)}=1.39$, $p=0.26$). When a similar analysis was performed on the data with a 140-pps target, the effect of masker rate was highly significant ($F_{(1,15)}=71.56$, $p<0.001$).

Assmann, P., and Summerfield, Q. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Baskent, D., and Shannon, R. V. (**2003**). ''Speech recognition under conditions of frequency–place compression and expansion,'' J. Acoust. Soc. Am. **113**, 2064–2076.

Bench, J., and Bamford, J. (**1979**). *Speech-hearing Tests and the Spoken Language of Hearing Impaired Children* (Academic, London).

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., and Clark, G. (**1984**). ''Speech processing strategies using an acoustic model of a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **76**, 104–110.

Broadbent, D. E., and Ladefoged, P. (**1957**). ''On the fusion of sounds reaching different sense organs,'' J. Acoust. Soc. Am. **29**, 708–710.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perceptual separation of simultaneous voices,'' J. Phonetics **10**, 23–36.

Brown, G. J., and Cooke, M. (**1994**). ''Computational auditory scene analysis,'' Comput. Speech Lang. **8**, 297–336.

Carlyon, R. P. (**1994**). ''Detecting pitch–pulse asynchronies and differences in fundamental frequency,'' J. Acoust. Soc. Am. **95**, 968–979.

Carlyon, R. P. (**1996**). ''Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker,'' J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P., and Deeks, J. M. (**2002**). ''Limitations on rate discrimination,'' J. Acoust. Soc. Am. **112**, 1009–1025.

Carlyon, R. P., Moore, B. C. J., and Micheyl, C. (**2000**). ''The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones,'' J. Acoust. Soc. Am. **108**, 304–315.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). ''Temporal pitch mechanisms in acoustic and electric hearing,'' J. Acoust. Soc. Am. **112**, 621–633.

Culling, J. F., and Darwin, C. J. (**1993**). ''Perceptual separation of simultaneous vowels—Within and across-formant grouping by *F*0,'' J. Acoust. Soc. Am. **93**, 3454–3467.

Darwin, C. J. (**1992**). ''Listening to two things at once,'' in *Audition, Speech, and Language*, edited by B. Schouten (Mouton-De Gruyter, Berlin), pp. 133–148.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). ''Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants,'' J. Acoust. Soc. Am. **108**, 1877–1887.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparisons of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. J., and Shannon, R. V. (**1999**). ''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. (**2000**). ''Effect of stimulation rate on phoneme recognition by Nucleus-22 cochear implant listeners,'' J. Acoust. Soc. Am. **107**, 589–597.

Geurts, L., and Wouters, J. (**2001**). ''Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.

Kong, Y.-Y., Stickney, G. S., and Zeng, F.-G. (**2003**). ''Combined acoustic and electric hearing for speech in noise and melody recognition,'' 2003 Conference on Implantable Auditory Prostheses; Asilomar Conference Grounds, Pacific Grove, CA.

Loizou, P. C., Poroy, O., and Dorman, M. (**2000**). ''The effect of parametric variations of cochlear implant processors on speech understanding,'' J. Acoust. Soc. Am. **108**, 790–802.

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.

McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.

Meddis, R., and Hewitt, M. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.

Moore, B. C. J., and Carlyon, R. P. (**2004**). ''Perception of pitch by people with cochlear hearing loss and by cochlear implant users,'' in *Springer Handbook of Auditory Research: Pitch Perception*, edited by C. J. Plack and A. J. Oxenham (Springer, Berlin, in press).

Nelson, P. B., and Jin, S.-H. (**2002**). ''Understanding speech in single-talker interference: Normal-hearing listeners and cochlear implant users,'' J. Acoust. Soc. Am. **111**, 2429.

Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.

Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.

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. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 97–104.

Qin, M. K., and Oxenham, A. J. (**2003a**). ''The effects of simulated cochlear-implant processing on *F*0 discrimination,'' J. Acoust. Soc. Am. **113**, 2224.

Qin, M. K., and Oxenham, A. J. (**2003b**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,'' J. Acoust. Soc. Am. **114**, 446–454.

Scheffers, M. T. M. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' Doctoral dissertation, University of Groningen, Netherlands.

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.

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

Summerfield, A. Q., Marshall, D. H., Barton, G. R., and Bloor, K. E. (**2002**). ''A cost-utility scenario analysis of bilateral cochlear implantation,'' Arch. Otolaryngol. Head Neck Surg. **128**, 1255–1262.

Turner, C., Gantz, B., Lowder, M., and Gfeller, K. (**2003**). ''Combining acoustic and electric hearing: Clinical studies,'' 2003 Conference on Implantable Auditory Prostheses; Asilomar Conference Grounds, Pacific Grove, CA.

van Hoesel, R. J. M., and Tyler, R. S. (**2003**). ''Speech perception, localization, and lateralization with bilateral cochlear implants,'' J. Acoust. Soc. Am. **113**, 1617–1630.

van Wieringen, A., Carlyon, R. P., Long, C. J., and Wouters, J. ''Pitch of amplitude-modulated irregular-rate stimuli in electric and acoustic hearing,'' J. Acoust. Soc. Am. **114**, 1516–1528.

Warren, R. M., Riener Hainsworth, K., Brubaker, B. S., Bashford, J. A., and Healy, E. W. (**1997**). ''Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps,'' Percept. Psychophys. **59**, 275–283.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

A three-dimensional model of vocal fold abduction/adduction

Eric J. Hunter[a)]
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204

Ingo R. Titze
National Center for Voice and Speech, Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242, and National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204

Fariborz Alipour
National Center for Voice and Speech, Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

(Received 26 June 2002; revised 18 December 2003; accepted 27 December 2003)

A three-dimensional biomechanical model of tissue deformation was developed to simulate dynamic vocal fold abduction and adduction. The model was made of 1721 nearly incompressible finite elements. The cricoarytenoid joint was modeled as a rocking–sliding motion, similar to two concentric cylinders. The vocal ligament and the thyroarytenoid muscle's fiber characteristics were implemented as a fiber–gel composite made of an isotropic ground substance imbedded with fibers. These fibers had contractile and/or passive nonlinear stress–strain characteristics. The verification of the model was made by comparing the range and speed of motion to published vocal fold kinematic data. The model simulated abduction to a maximum glottal angle of about 31°. Using the posterior-cricoarytenoid muscle, the model produced an angular abduction speed of 405° per second. The system mechanics seemed to favor abduction over adduction in both peak speed and response time, even when all intrinsic muscle properties were kept identical. The model also verified the notion that the vocalis and muscularis portions of the thyroarytenoid muscle play significantly different roles in posturing, with the muscularis portion having the larger effect on arytenoid movement. Other insights into the mechanisms of abduction/adduction were given. [DOI: 10.1121/1.1652033]

PACS numbers: 43.70.Bk [AL] Pages: 1747–1759

I. INTRODUCTION

Vocal fold dynamics can be treated in two parts: (1) large and relatively slow deformations occurring when the vocal folds abduct/adduct or elongate as a result of moving boundaries or internal muscle forces, and (2) small and relatively fast deformations occurring when the tissue is driven into oscillation. These two parts are referred to as, respectively, vocal fold posturing and vocal fold vibration. In speaking and singing, vocal fold posturing occurs in a nonperiodic (but ultimately always cyclic) fashion at frequencies of less than 10 Hz. Flow-induced vocal fold vibration, on the other hand, usually occurs at 100 Hz or more. Although posturing and vibration can be treated separately, many aspects of vibration (e.g., pitch, loudness, and voice quality) are dependent on posturing.

Understanding vocal fold posturing is important to both speech scientists and clinicians. For example, vocal fold posturing has been shown to affect the intensity of the voice (Titze and Sundberg, 1992; Murry *et al.*, 1998), the pitch (Hirano *et al.*, 1970; Honda, 1983), devoicing (Yoshioka, 1981), and ventilation or glottal aspiration (Tomori *et al.*, 1998). Vocal fold abdutory/adductory posturing is also important for swallowing and effort closure of the airway (e.g., for lifting). Adduction of the vocal folds has been hypothesized to be the key to optimizing airflow–tissue interaction in vocal onset and, thus, self-sustained oscillation (Titze, 1988; Cooke *et al.*, 1997).

The study of vocal fold posturing is also essential for improved phonosurgery. Unilateral vocal fold paralysis, frequently the consequence of injury to or infection of the recurrent laryngeal nerve (or the vegus nerve from which it branches), can significantly impair an individual's speech, swallowing, and ability to protect the airway. A common phonosurgical treatment of unilateral vocal fold paralysis is vocal fold medialization, an artificial posturing of a vocal fold to set the shape and stiffness so that phonation is improved (Neuman *et al.*, 1994). With an increased understanding of the mechanics of abduction/adduction, especially in terms of the local and global stiffness properties of the modified tissue, it is expected that phonosurgical procedures like medialization could be enhanced.

Previous studies of vocal posturing mechanics focused on quantifying the range and direction of arytenoid cartilage motion around the cricoarytenoid joint (CAJ) (Broad, 1968; Kotby *et al.*, 1992). Although such descriptive kinematics provided a logical foundation for understanding CAJ-based posturing, a study of the forces caused by laryngeal connec-

a)Corresponding author: Eric Hunter, Ph.D., National Center for Voice and Speech, The Denver Center for the Performing Arts, 1245 Champa Street, Denver, Colorado 80204. Telephone: 303-446-4839; electronic mail: ehunter@dcpa.org

tive tissues is needed to establish causal relations between vocal posturing and laryngeal muscle forces. This point was noted by Sanders *et al.* (1994), who deemed this relationship essential to understanding the full range of vocal fold movements seen during phonation, respiration, and effort closure. Sanders *et al.* illustrated this point by using an *in vivo* canine larynx to demonstrate the effects of contracting different portions of one laryngeal muscle, the posterior cricoarytenoid muscle (PCA), on vocal fold posturing. Further, he suggested that there may be as many as four functionally different compartments of the thyroarytenoid (TA) affecting vocal fold mechanics. Therefore, describing orientations and distributions of the portions of the different laryngeal muscles, such as the TA, is a necessary step in understanding CAJ-based vocal fold posturing.

Creating a biomechanical model of CAJ-based vocal posturing is important because it can quantify the effects of individual and coordinated muscle contractions on arytenoid cartilage movements and vocalization. Farley (1996) created this type of model to study the dependence of the fundamental frequency of phonation on glottal width and closure. The posturing portion of his model was based on rigid body mechanics, from which he calculated pitch for a given arytenoid posture. Although he was able to use his model to predict glottal width from muscle activation, there were two areas that he did not address. First, Farley reported only static results; he did not address the dynamics of vocal posturing, such as the speed of glottal width change. Second, because his model was based on rigid body mechanics, he did not address tissue deformation. These two limitations restricted the type of vocal pathologies for which Farley's model could be applied; for example, it could not simulate the effect of a polyp or scar on fundamental frequency. In contrast, a dynamic model based on deformable body theory and solved by the finite element method (FEM) could include localized changes in the vocal fold geometry, providing the foundation for simulating a wide range of disorders.

FEM has previously been deployed to simulate other speech organs, such as the tongue (Wilhelms-Tricarico, 1995), the velum (Berry *et al.*, 1999), and the vocal folds in oscillation (Gunter, 2003; Alipour *et al.*, 2000; Dang and Honda, 1998). It is noteworthy that only Wilhelms-Tricarico's model of the tongue included large deformation theory with tissue properties based on fiber directions. In particular, FEM models of the vocal folds have thus far been based only on small deformation theory, which are not valid for the large movements seen in vocal fold posturing without making additional assumptions (see Sec. II A). Thus, the approach of Wilhelms-Tricarico's motivated this study.

The goal of this research was to advance the modeling of vocal fold posturing in three dimensions; this study specifically focuses on large deformation abduction/adduction posturing with the fibrous tissues intrinsic to the vocal folds playing a key role. In light of this goal, the specific research question was as follows: To what degree can a three-dimensional biomechanical model of vocal fold abduction/adduction match the dynamics of human vocal fold abduction/adduction (in terms of glottal angle), with percent muscle activation as the driving inputs? A model was created that focused on the motion of the arytenoid cartilages and included vocal fold deformation and intrinsic fibrous tissues. Fundamental to the model's development was the creation of a fiber–gel composite, which included the contribution of one-dimensional muscle fibers superimposed on a three-dimensional isotropic gel. The mathematical steps used in the creation of this fiber-gel were detailed, and the model's output was then compared to the existing data on the glottal angle and the speed of glottal closure.

II. METHODS

Following general FEM procedures, there were six steps in the model's creation: (1) definition of dynamic finite element equations; (2) identification of the target geometry (anatomical properties); (3) determination of constitutive equations and material properties (e.g., tissue elasticity, density, and incompressibility); (4) specification of the boundary conditions and constraints; (5) implementation of the finite elements (discretization of the domain and assembly of element equations); and (6) solution of the model with numerical results. Quantities such as fiber length, area, and orientation were classified as anatomical properties, while passive and contractile characteristics of the intrinsic fibers were classified as material properties.

A. Finite element equation definitions

The FEM equations had to adequately represent the large deformations (or finite strains) that occur in vocal posturing. It was predetermined that the equations would be based on eight-node hexahedral elements and hexahedral-reduced forms (prism and pyramid shapes). Some of the critical equations are provided below (as an overview), along with added details of a large-deformation extension; the specific FEM equations used, based on more general derivations found in intermediate FEM texts (e.g., Cook *et al.*, 1989; Bathe, 1996), can be found in Hunter (2001).

The derivation of these FEM equations commences with the potential energy Π of the target geometry,

$$\mathbf{\Pi} = U - R, \tag{1}$$

where U is the strain energy and R is the virtual work of the system in deformation. The strain energy was computed as a volume integral of the strain energy density U_0,

$$U = \int\int\int U_0\, dV, \quad where \;\; U_0 = \frac{1}{2}[\boldsymbol{\epsilon}]^T[\mathbf{S}][\boldsymbol{\epsilon}], \tag{2}$$

which can be thought of as a three-dimensional extension of the elastic energy in a spring obeying Hooke's Law: $U_0 = 1/2\, kx^2$ (the stiffness matrix $\mathbf{S}$ is comparable to k, while strain vector $\boldsymbol{\epsilon}$ is similar to x).

The entire system of equations is ultimately represented by a matrix version of the general equation of motion,

$$\mathbf{Ku} + \mathbf{D\dot{u}} + \mathbf{M\ddot{u}} - \mathbf{R} = 0, \tag{3}$$

which is derived from a constitutive equation for tissue (see Sec. II C), a small-strain vector $\boldsymbol{\epsilon}$, a trial solution, and boundary conditions (described in Sec. II D). Terms in this equation are a symmetric stiffness matrix $\mathbf{K}$, displacement vector $\mathbf{u}$, damping matrix $\mathbf{D}$, mass matrix $\mathbf{M}$, and vector $\mathbf{R}$ (repre-

senting the boundary conditions and applied forces). The dimensions of these matrices are dependent on the number of nodes (solution points) and the number of degrees of freedom. The damping matrix **D** comes from a proportion of the mass and stiffness matrices,

$$\mathbf{D}=\alpha\mathbf{M}+\beta\mathbf{K}, \tag{4}$$

where α and β are, respectively, Rayleigh mass and Rayleigh stiffness damping. The Rayleigh terms are frequency dependent and related to a common damping ratio ζ (e.g., Bathe, 1996, p. 797),

$$\zeta=\frac{\alpha}{4\pi f}+\beta\pi f. \tag{5}$$

Since ζ is a directly measurable quantity, α and β can be calculated from two measures of the damping, ζ_1 and ζ_2, at two different frequencies, f_1 and f_2.

However, because Eq. (3) is based on small strain assumptions, it cannot adequately represent the large deformations that occur in vocal fold posturing, nor can it represent nonlinear material behavior, which is also found in the vocal folds. In this study, the Newton–Raphson method (e.g., Bathe, 1996, p. 755) was used to allow for large deformation and nonlinear material behavior. This specific method used an iteration technique to solve the system during static situations, or at a single time step in a dynamic solution,

$$\mathbf{Ku}-\mathbf{R}=0. \tag{6}$$

The iteration technique begins with the calculation of the out-of-balance load vector $\Delta\mathbf{R}$, or the difference between the applied loads and the restoring force **F**,

$$\mathbf{R}-\mathbf{F}=\Delta\mathbf{R}. \tag{7}$$

Given $\Delta\mathbf{R}$, an incremental displacement $\Delta\mathbf{u}$ is calculated from Eq. (6) and convergence is checked. If $\Delta\mathbf{u}$ is not sufficiently small, $\Delta\mathbf{R}$ is reevaluated and the stiffness matrix is updated,

$$\mathbf{K}_{\mathbf{new}}\Delta\mathbf{u}-\Delta\mathbf{R}=0. \tag{8}$$

At this point the process repeats itself with a new stiffness matrix. Iterations continue as needed until a preset tolerance is met.

By using the iteration technique in finite element integration, large deformations could be accounted for in a single time step. Solving Eq. (3) with the time dependence was accomplished using the FEM difference technique (Alipour *et al.*, 2000). A smaller time step in the finite difference solution would reduce the number of iterations needed in the Newton–Raphson technique. However, with large deformations, the need for a stiffness matrix correction would never be negated regardless of how small the time steps became; only the frequency of needed corrections would be reduced.

B. Target geometry (anatomical properties)

The targeted geometry of the posturing model was a human male larynx, with the reference frame being the cricoid cartilage (Fig. 1). The primary anatomical focus was the tissue between the glottis and the thyroid cartilage, including

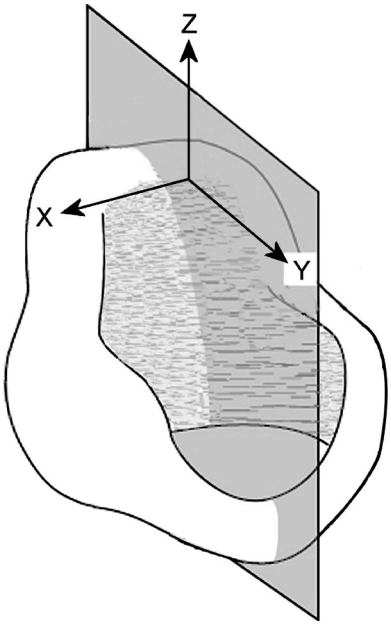


FIG. 1. Coordinate system from vantage point of cricoid cartilage (after Selbie *et al.*, 1998).

the arytenoid cartilage. The model's *y* axis was anterio-posterior (front to back) with $y=0$ at the tip of the cadaveric vocal process; the *x* axis was medio-lateral (left to right), while the *z* axis was inferio-superior (down to up). The unstressed (neutral) tissue state was cadaveric. With geometric symmetry across the midplane assumed, the model consisted of only the right vocal fold, similar to a hemilarynx (Fig. 2).

The thyroid cartilage was selected to be the outer edge of the model. The shape of the thyroid cartilage was derived from average measurements reported by Eckel and Sittel (1995). The shape and position of the vocal fold anterior commissure (Fig. 2, point D) and the vocal process (point E) were averaged from Eckel and Sittel (1995), Eckel *et al.* (1994), and Hirano and Sato (1993). The posterior glottis was determined solely from photographs of histological slices (Hirano and Sato, 1993).

The midmembranous medial curvature (Fig. 3), which was based on the canine vocal fold (Tayama *et al.*, 2002), extended from the vocal process (Fig. 2, point E, 1.5 mm from the midline) to near the anterior commissure (points D, superior and inferior points shown), with a membranous length of approximately 13 mm. A smooth transition at both ends of the fold was applied via a nonplanar Coons patch (a surface patch between four arbitrary boundary curves). The inferior border of the model was defined at the base of the TA muscle (Fig. 3) (as defined by Tayama *et al.*, 2002). The model's entire superior surface was initially simulated to be flat.

In the modeled arytenoid cartilage, the distance between the vocal process and muscular process was taken from nine human male cartilages (Tayama *et al.*, 2001). The arytenoid cartilage height was trimmed at the vocal fold superior surface so that it would not protrude above the vocal fold. To simulate the effect of the CAJ, the arytenoid's CAJ facet was extended inferiorly (4 mm) into the space where the cricoid

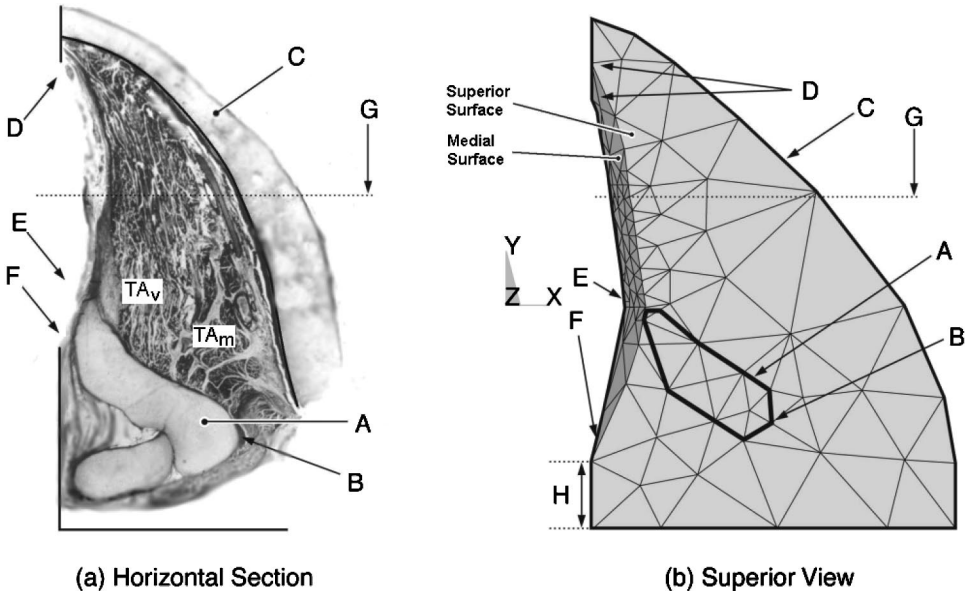


FIG. 2. The right vocal fold: (a) horizontal slice at the level of the vocal process; and (b) the view of the model from above. Vocal fold landmarks: A—arytenoid cartilage; B—muscular process; C—thyroid cartilage; D—anterior commissure; E—vocal process; F—posterior commissure; G—midmembranous vocal fold line; and H—posterior border.

cartilage would normally be (further details provided in Sec. II D).

The only ligament modeled was the vocal ligament. Muscle tissue included the four intrinsic laryngeal muscles attached to the arytenoid: the TA, the PCA, the interarytenoid (IA), and the lateral cricoarytenoid (LCA). The two bellies of the TA muscle, the thyrovocalis (TA_V), and the thyromuscularis (TA_M), were treated separately. TA fiber bundles (with measured insertion points, direction cosines, lengths, and areas taken from six human cadaver larynges; Cox *et al.*, 1999) were grouped into five macrobundles (which were shared between the TA_M and TA_V); individual direction cosine averages for each of the five macrobundles only approximated the rule $l^2+m^2+n^2=1$ (± 0.06). The lengths, areas, and direction cosines of the PCA, LCA, and IA were taken from eight canine larynges (Mineck *et al.*, 2000). The average insertion points of the PCA and LCA were attached to the muscular process, while the IA was attached at the most posterior arytenoid position, which represents the likely point at which the IA acts on the arytenoid (given the sling-like wrapping of the IA around the arytenoid).

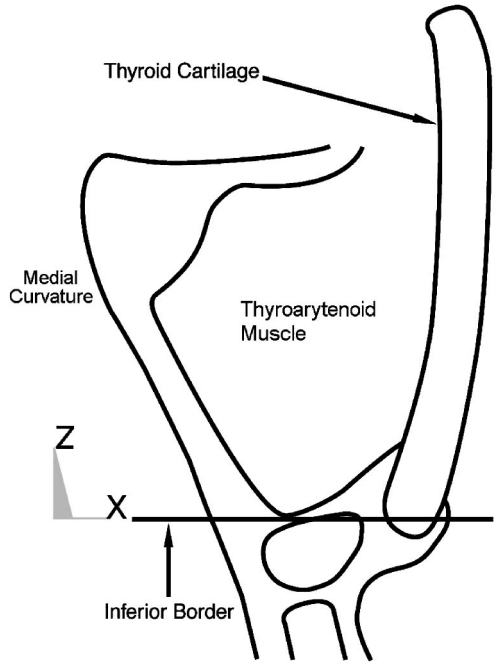


FIG. 3. Schematic drawing of midmembranous coronal section of vocal fold with the lower border of the model's target area labeled.

C. Constitutive equations and material (tissue) properties

Tissues in the model were classified as either fibrous or nonfibrous. Fibrous tissues included the vocal ligament and intrinsic vocal fold muscles. Nonfibrous tissues included the cartilages, mucosa, superficial layer of the lamina propria (excluding the vocal ligament), and other miscellaneous tissues. To quantify the mechanics of fibrous and nonfibrous tissue deformations, an individual constitutive equation (CE) was defined for each. In the model, the various tissues' material properties (material-specific parameters in the CEs) were obtained from the literature.

1. Constitutive equations

The relation between the stress and strain of a tissue was represented by a CE. Three assumptions about vocal tissues were made in defining the CEs used in the model: (1) all tissue was nearly incompressible; (2) all tissue was isotropic; and (3) fibrous tissues were a superposition of contractile or other fibrous characteristics onto isotropic tissues. Isotropic tissue was described by the three-dimensional Hooke's law. In the familiar one-dimensional linear system, Hooke's law is $\sigma=E\epsilon$, where σ and ϵ represent the stress and strain, respectively, and E is the tangent Young's modulus (a measure of a tissue's stiffness). The three-dimensional Hooke's law (which can be found in many variations in intermediate mechanics texts) takes into account both shear and normal components of stress and strain. For the symmetric shears in an isotropic material, Hooke's law can be written in matrix notation, which is ideal for FEM (Cook *et al.*, 1989),

$$\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{pmatrix} = \begin{pmatrix} [^3s] & [^30] \\ [^30] & [^3s'] \end{pmatrix} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ \epsilon_{xy} \\ \epsilon_{yz} \\ \epsilon_{zx} \end{pmatrix}, \tag{9}$$

where the following submatrices are defined,

$$[^3s] = \frac{2\mu}{1-2\nu} \begin{pmatrix} 1-\nu & \nu & \nu \\ \nu & 1-\nu & \nu \\ \nu & \nu & 1-\nu \end{pmatrix}, \tag{10}$$

$$[^3s'] = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \mu \end{pmatrix}. \tag{11}$$

In these equations, μ is the shear modulus, which is related to E by the Poisson's ratio ν (a measure of a material's compressibility),

$$\mu = \frac{E}{2+2\nu}. \tag{12}$$

After the substitution of Eqs. (10)–(12), the material parameters in Eq. (9) are E and ν.

Fibrous tissues intrinsic to the vocal folds contain both elastin and collagen fibers in their extracellular matrix; they also have a nonlinear stress–strain relationship; hence a nonlinear E. Titze (1996) defined the axial passive fiber stress σ_p for a one-dimensional laryngeal fiber as a piecewise static stress–strain relationship composed of a linear and an exponential portion. The CE is continuous, as is its first derivative,

$$\begin{aligned} \sigma_p &= -\frac{\sigma_0}{\epsilon_1}(\epsilon_y - \epsilon_1) - B\sigma_2(\epsilon_y - \epsilon_2) \\ &\quad + \sigma_2[e^{B(\epsilon_y - \epsilon_2)} - 1], \quad \epsilon_y > \epsilon_2, \\ \sigma_p &= -\frac{\sigma_0}{\epsilon_1}(\epsilon_y - \epsilon_1), \quad \epsilon_y \leqslant \epsilon_2, \end{aligned} \tag{13}$$

where σ_0 is the stress at zero strain (representing the prestrained state of the laryngeal fibers), σ_2 is a scale factor for the exponential function, B is an exponential strain constant, ϵ_1 is the strain at zero stress, ϵ_2 is the strain at which the nonlinear exponential function begins, and ϵ_y is the axial strain (independent variable). The nonlinear tangent Young's modulus is the derivative of Eq. (13) with respect to strain,

$$E = \frac{d\sigma_p}{d\epsilon_y}. \tag{14}$$

In the dynamic case, Eqs. (13) and (14) are used in a differential equation that also includes viscous losses. Titze (1996), using the classic Kelvin model [e.g., the lower part of Fig. 4(a)], defined a differential equation to represent an arbitrary tissue with both an active (contractile) and a passive component with parallel stiffness k_p, series stiffness k_s, and internal viscous damping d. Titze defined the dynamic differential relationship between ϵ_y and total axial stress σ_T as

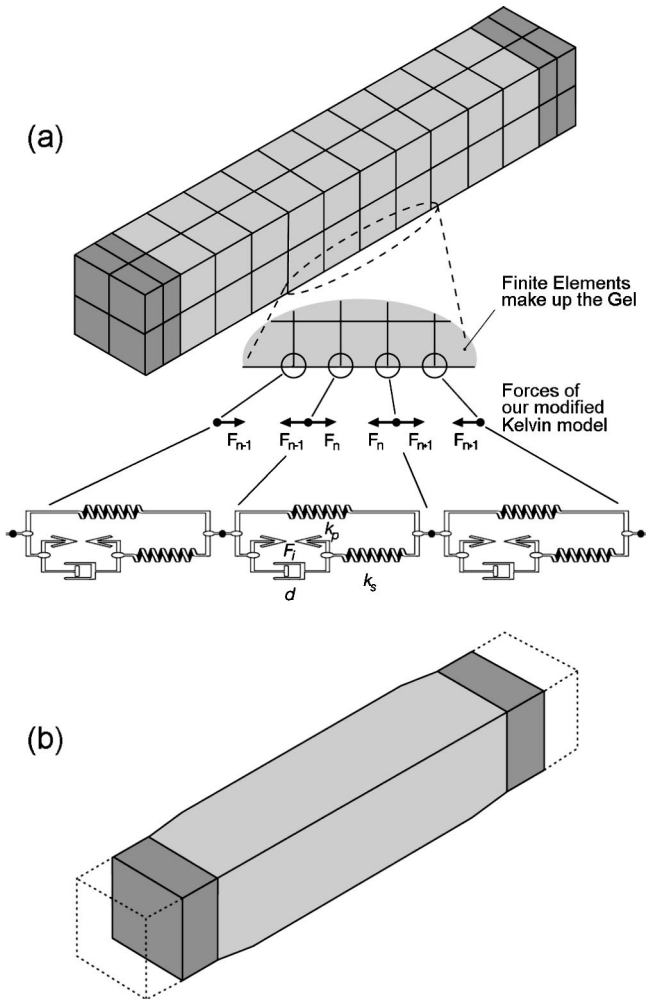


FIG. 4. A model of the thyroarytenoid muscle bundle with cartilage on both ends (shaded): (a) forces from our modified Kelvin fiber model superimposed on specific nodes in series to add viscous losses and contraction forces; and (b) contraction of the modeled thyroarytenoid muscle bundle with the dotted line depicting the initial position.

$$t_s\dot{\sigma}_T + \sigma_T = \sigma_p(\epsilon_y) + Et_p\dot{\epsilon}_y, \tag{15}$$

where t_s is a series time constant, t_p is a parallel time constant, σ_p is the passive stress [Eq. (13)], and $\dot{\epsilon}$ is the axial strain rate. Work cycle losses in Eq. (15) are represented by product terms, which include the time constants. The two time constants are defined in terms of stiffness and damping components.

$$t_s = d/k_s, \tag{16}$$

$$t_p = d\frac{k_p + k_s}{k_p k_s}. \tag{17}$$

Unlike passive-only tissue, muscle tissue can internally generate stress, which then acts on the surrounding structure. The active stress σ_a in a muscle is created from the force generated by overlapping proteins (actin and myosin); it is dependent not only on strain, but also strain rate and activation (a in percent: 0 to 100). Wilhelms-Tricarico (1995) specified the active stress CE in product form,

$$\sigma_a = a\sigma_m g(\dot{\epsilon}_y) f(\epsilon_y). \tag{18}$$

TABLE I. General properties as needed by the finite element model.

Parameter description	Symbol	Cartilage	Misc. tissue	Ligament	TA
Young's modulus	E	30 MPa	20.7 kPa	*See Table II*	*See Table II*
Poisson's ratio	ν	0.47	0.47	0.47	0.47
Rayleigh mass	α	32.3	32.3	32.3	32.3
Rayleigh stiffness	β	6.87×10^{-4}	6.87×10^{-4}	6.87×10^{-4}	6.87×10^{-4}
Density	ρ	1.043 g/cm^3	1.043 g/cm^3	1.043 g/cm^3	1.043 g/cm^3
Cross-sectional area	A	NA	NA	5 mm^2	63.8 mm^2

In Eq. (18), σ_m is a constant defining the maximum contractile stress a muscle can produce. Functions $g(\dot{\epsilon}_y)$ and $f(\epsilon_y)$ depend on the strain rate and strain, respectively, and define the nature of the active components of the muscle; the piecewise function $g(\dot{\epsilon}_y)$, adopted from Wilhelms-Tricarico (1995), is dependent on the strain rate and maximum strain rate $\dot{\epsilon}_m$,

$$g(\dot{\epsilon}_y)=\mathrm{Max}\left[0,\frac{\dot{\epsilon}_y/\dot{\epsilon}_m+1}{1-3\dot{\epsilon}_y/\dot{\epsilon}_m}\right],\quad \dot{\epsilon}_y\leqslant 0$$
$$=\frac{9\dot{\epsilon}_y/\dot{\epsilon}_m+1}{5\dot{\epsilon}_y/\dot{\epsilon}_m+1},\quad \dot{\epsilon}_y>0. \tag{19}$$

The function $f(\epsilon_y)$, a stress–strain factor, is modeled as a polynomial from experimental data (Titze, 1996),

$$f(\epsilon_y)=\mathrm{Max}[0,1-b(\epsilon_y-\epsilon_m)^2], \tag{20}$$

where ϵ_m is the strain at maximum contractile stress, and b is a coefficient.

Internal contractile stress is related to active stress through a differential equation that contains an empirically derived internal contraction time t_i,

$$t_i\dot{\sigma}_i+\sigma_i=\sigma_a. \tag{21}$$

The total stress within a fiber is now the internal active stress, σ_i, added to the passive stress from Eq. (15),

$$t_s\dot{\sigma}_T+\sigma_T=\sigma_p+Et_p\dot{\epsilon}_y+\sigma_i. \tag{22}$$

Equations (21) and (22) then represent our complete modified Kelvin fiber model, which includes a muscle fiber's active and passive stress. This fiber model, used in conjunction with the three-dimensional isotropic CE above, is the basis for the three-dimensional fibrous tissues in the FEM posturing model.

2. Material properties

Material properties (tissue-specific parameters from the CEs) were listed in Tables I, II, and III. All vocal tissues in the model were nearly incompressible, $\nu=0.47$ (0.5 being theoretically correct for incompressible materials). All tissues were also assigned a density ρ of 1.043 grams per cubic centimeter, as measured by Perlman (1985) for laryngeal tissue. Using Eq. (5), Rayleigh mass and stiffness quantities ($\alpha=32.3$ and $\beta=6.87\times10^{-4}$) were calculated from the reported median of two ζ ranges ($\zeta=0.15$ at 30 Hz and $\zeta=0.3$ at 140 Hz) (Chan and Titze, 1998).

Cartilage material was specified with a linear E of 30 MPa (Li, 1995). All other nonfibrous tissues were assigned an E value of 20.7 kPa, as measured from canine laryngeal muscle tissue at low strains (Alipour-Haghighi and Titze, 1991). The fibrous TA muscle's nonlinear E [Eq. (14)] and the time constants [Eqs. (16) and (17)] were obtained by fitting the equation parameters to canine stress–strain data (Fig. 5) (Table II, from Titze, 1996). The vocal ligament's nonlinear E was obtained from human ligament data in a similar fashion (Min *et al.*, 1995). The optimization methodology for obtaining the fiber model parameters for all laryngeal muscles from laboratory data is the subject of an ongoing study.

Active material properties were defined as those tissue-specific parameters in Eqs. (18)–(22). To obtain the TA muscle parameters listed in Table III (Titze, 1996), three types of muscle data were needed: (1) a muscle's twitch response and tetanic contraction; (2) the contraction and relaxation times; and (3) the force–velocity and length–tension relationships. Studies have focused on PCA contraction force and timing (Cooper *et al.*, 1994) as well as IA, LCA, and PCA contraction times (Alipour *et al.*, in preparation). Nevertheless, of the five intrinsic muscles of the larynx, only two (the CT and TA) have been sufficiently measured and quantified so as to fully populate all of the parameters of the fiber model (Titze, 1996; Alipour and Titze, 1999). Therefore, the LCA, IA, and PCA were assumed to have the same material property parameter values as the TA but different anatomical properties (e.g., length and orientation). The fiber model parameters could be optimized to match any specific muscle if the data were available; thus, as ongoing studies are com-

TABLE II. Passive material properties for the one-dimensional Titze (1996) fiber model.

Equation	Parameters description	Symbol	Ligament	TA (canine)
13–15	Passive stress at resting length	σ_0	13 kPa	2 kPa
13–15	Scaling of exponential stress	σ_2	45 kPa	55 kPa
13–15	Strain at zero stress	ϵ_1	−0.5	−0.5
13–15	Strain at exponential stress	ϵ_2	0.2	0.2
13–15	Exponential strain constant	B	6.8	4.5
15	Parallel time constant	t_p	0.08 s	0.050 s
15	Series time constant	t_2	0.07 s	0.044 s

TABLE III. Active (contractile) material properties for the one-dimensional Titze (1996) fiber model.

Equation	Active parameter description	Symbol	TA (canine)
18	Maximum active stress	σ_m	105 kPa
19	Maximum strain rate	$\dot{\epsilon}_m$	6.0 s^{-1}
20	Coeff. for active stress–strain	b	1.07
20	Strain at max. active stress	ϵ_m	0.4
21	Activation time	t_i	$0.01+0.05\epsilon_y$ s

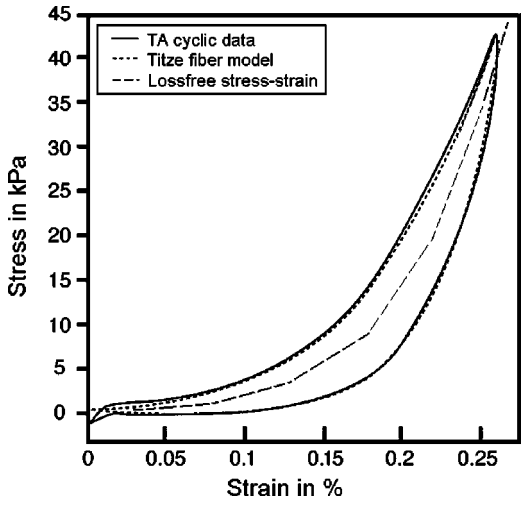


FIG. 5. The passive stress–strain response for the canine thyroarytenoid muscle. Shown are (1) measured cyclic (1 Hz) elongation data (solid line); (2) Titze fiber model response after being fit to data (dotted line); and (3) the piecewise stress–strain relation with no viscous losses (dashed line) representing the gel response.

pleted, parameter values can be obtained and used to refine the FEM posturing model.

D. Boundary conditions and constraints

Boundary conditions and constraints were used to simplify the FEM solution and to capture the key mechanics of abduction/adduction. The boundary conditions of the posturing model (shown in Fig. 6) consisted of four degree-of-freedom constraints: the inferior (bottom) border in the z direction ($d_z=0$) (surface 1); the posterior (back) border in the y direction ($d_y=0$) (surface 2); the shared medial border (left) near the anterior commissure and posterior glottis in the x direction ($d_x=0$) (surface 3); and the latero-anterior border in all directions (thyroid cartilage, $d_x=d_y=d_z=0$) (surface 4). Next, the CAJ was modeled by a constraint along a rocking–sliding axis (Fig. 6, dashed line). The vocal fold's superior and medial surfaces were allowed to move freely.

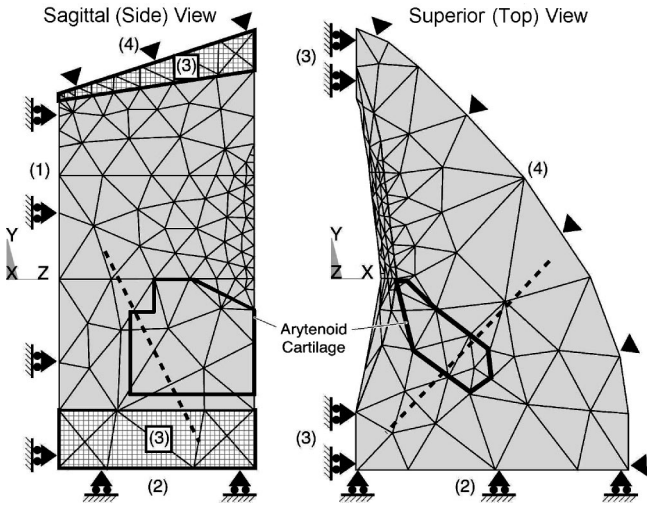


FIG. 6. The three-dimensional vocal fold model divided into 1721 elements with only surface element faces shown. Four surface constraints within the model: (1) inferior (bottom) border, (2) posterior (back) border, (3) medial border, and (4) latero-anterior (thyroid cartilage) border. The CAJ axis (dotted line) is shown passing though the arytenoid cartilage (represented by the dark outline).

A simple boundary constraint could not be easily formed from the complex combination of tissues posterior to the arytenoid cartilage (e.g., IA and PCA, ligaments, membranes, and portions of the cricoid cartilage). Therefore, the back border (Fig. 6, surface 2) was padded with two millimeters of isotropic soft tissue, which allowed the arytenoid to move with an effective soft tissue drag that simulated this complex combination of tissues (Fig. 2, point H). At this back boundary (Fig. 6, surface 2), the surface was constrained in the anterior–posterior direction ($d_y=0$).

The CAJ is essential to abduction/adduction posturing; therefore, the model was designed to preserve the key movement seen in the joint, a rocking–sliding motion (Selbie *et al.*, 1998). This motion was captured by extending the modeled arytenoid cartilage inferiorly beyond the anatomical facet location and then constraining the extension with a rocking–sliding axis. This axis was obtained from Selbie *et al.*, who calculated various joint axes about which the arytenoid rocked and slid by analyzing the shape of the facets. From the reported orientation of the joint axes, a rocking axis was calculated in terms of direction cosines for the right half of larynx: $l=0.609$, $m=0.619$, and $n=-0.495$. The axis appeared about 4 mm below the human CAJ facet and 2.5 mm below the lowest TA fiber insertion point in the posturing model.

E. Finite element implementation

A commercial FEM software package, ANSYS 5.7, was used to construct the model and solve the FEM equations. Key to the FEM implementation was simulating fibrous tissues as a fiber–gel composite, which was created by superpositioning fiber characteristics onto isotropic finite elements through special subroutines written in the ANSYS Parametric Design Language. The ANSYS *transient* option was used to include time dependence; nonlinearities were solved via the ANSYS *Newton–Raphson* option.

Modeling more than one type of these nonlinearities at a time in FEM increases computational costs significantly. Vocal fold posturing has not only large-strain geometric nonlinearities (i.e., large deformations), but also material behavior nonlinearities (e.g., stress–strain, as in Fig. 5). Therefore, computational costs greatly increase because, for a single time step solution, multiple iterations are required for both of these types of nonlinearities.

1. Fiber–gel implementation

Because of the complexity of the vocal fold posturing model, a submodel of an isolated TA muscle was used to show how fiber–gel was created. The same principles were then employed by the larger posturing model to simulate intrinsic vocal fold fibers.

The submodel was designed in ANSYS using 56 elements (eight-node brick) with *transient* and *Newton–Raphson* options enabled. The geometry was a simple column [Fig. 4(a)] that measured 18 mm, the approximate length of human male TA strands (Cox *et al.*, 1999); the cross-sectional area was 12.25 mm^2 (2.5 mm to a side),

slightly more than half the average reported TA area. Portions of the thyroid and arytenoid cartilages were represented by cartilage attachments (length of 1.8 mm) placed at the ends of the muscle. These end elements were assigned material property values, as specified in Table I. The cartilages were sufficiently stiff so that superimposed fiber forces applied at the cartilage–muscle interface would cause negligible deformations to the cartilage. The muscle elements were designed with values from Table I; however, a nonlinear Young's Modulus [Eq. (14)] (Table II) was implemented in a piecewise fashion using an ANSYS Material Table (similar to the dashed stress–strain curve, Fig. 5). Therefore, if the submodel (currently only a gel) were cyclically elongated at this point in its creation, it could not simulate the TA work-cycle response (as seen in Fig. 5, solid).

Fibrous characteristics were next added to the gel by superimposing viscous losses and contractile properties onto the lines of nodes (or solution points) parallel to the muscle's length. If any two adjacent nodes along one of these parallel lines were examined, these nodes would represent the end nodes of a single fiber model [Fig. 4(a)]. Fibrous characteristics superimposed onto the two nodes were in the form of forces calculated by the fiber model; these forces were always directed toward the adjacent node of the pair. Each of the four outer edges of the rectangular TA muscle was designed with fiber characteristics using ten fiber models in series. Thus, the TA submodel had 40 fiber models oriented in four lines, with the resulting forces superimposed on their respective nodes.

Two steps were used to solve the TA submodel for a given time increment. First, multiple Newton–Raphson iterations were used in the FEM solution for the deformation of the finite elements. Second, vector force magnitudes were calculated by the fiber model from the resulting deformation. These vector forces were added by (1) recording the new fiber–gel node locations; (2) calculating the fiber strain and strain history for each of the 40 small fiber models; and (3) applying the losses and contractile forces to the nodes. However, because the finite element gel was constructed with nonlinear (loss-free) passive properties (which were also accounted for in the fiber model), the forces applied to the nodes had to be reduced by the amount of passive force already in existence within the elements.

The differential equations of a single fiber model [Eqs. (21) and (22)] were solved by a fourth-order Runge–Kutta routine, requiring a time sampling of approximately 1000 Hz to be stable. In solving the TA submodel (consisting of FEM equations and 40 fiber models), the solution time sampling was allowed to be variable in order to keep the overall solution stable while minimizing computational time; the average time sampling was approximately 1100 Hz, with each time sampling costing about one second of CPU time on an HP B2600 workstation (i.e., one second of dynamic solution expended about 1100 s of CPU time).

The TA submodel was verified by testing the muscle for cyclic elongation (1 Hz) and contractile effects. The results were as desired: the elongation's resulting stress/strain curve mirrored the fiber model (similar to Fig. 5), and the contraction shortened and bulged the muscle appropriately [Fig. 4(b)].

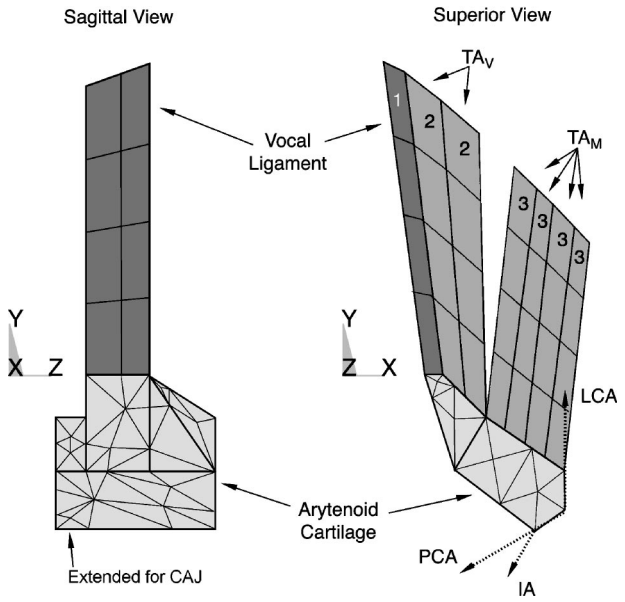


FIG. 7. The modeled arytenoid cartilage and fiber–gel composite volumes (only the outer element faces are shown). Three types of modeled fiber–gels: TA_M, TA_V, and vocal ligament. Intrinsic adductor and abductor muscles (PCA, LCA, and IA) attached near the muscular process. Dotted vectors in the superior view represent the applied forces ($x-y$ component shown only) of these three muscles.

2. Posturing model implementation

With the TA submodel created and the fiber–gel composite tested, the next step was to follow a similar pattern to implement the larger posturing model in ANSYS. The vocal fold's geometry (described in Sec. II B) was divided into hexahedral elements and hexahedral-reduced forms (prism and pyramid shapes) (Fig. 6, only surface element faces visible). The elements near the vocal fold margin were more finely structured to show greater deformation detail because they were of the most interest to future glottal shape studies. All elements were assigned their respective material properties from Tables I–III.

Boundary conditions and constraints (described in Sec. II D) were implemented and applied to the nodes as degree-of-freedom constraints. The CAJ's rocking–sliding axis was modeled using a constraint equation, which was assigned only to those nodes that were both on this axis and within the arytenoid's extended region (below the actual CAJ facet). The constrained motion of the individual nodes then combined to constrain the entire CAJ motion to a rocking–sliding cylindrical path.

The PCA, IA, and LCA were treated as point forces acting on the arytenoid (Fig. 7). These point forces were calculated from the fiber model using properties from Tables I–III. Although the muscles were only point forces, a virtual length was created by calculating the distance between the point at which force was applied on the arytenoid and a fixed point in space that represented the other end of the muscle (based on known lengths and orientations) (Table IV). These

TABLE IV. Anatomical properties of the canine abductor and adductor muscles (PCA, LCA, and IA) (Mineck *et al.*, 2000) and human TA muscle: the direction cosines, lengths, and cross-sectional areas of each muscle. The TA muscle fibers were grouped into five muscle bundles of nearly equal cross-sectional areas (TA total area $=63.8\ mm^2$), based on fiber measures as derived from multiple sources (as stated in the text).

Fiber direction	Direction cosines			Length (mm)	Cross-sectional area
	l	*m*	*n*		
PCA	−0.666	−0.228	−0.710	11.5	34.2 mm^2
LCA	−0.158	0.902	−0.403	14.4	21.2 mm^2
IA	−0.692	−0.627	−0.358	9.4	12.1 mm^2
TA_V	−0.187	0.932	−0.222	16.6	22% of total
	−0.135	0.945	−0.213	17.9	20% of total
	−0.129	0.963	−0.169	17.8	20% of total
TA_M	0.207	0.916	−0.247	19.8	19% of total
	0.288	0.9905	−0.223	20.8	19% of total

virtual lengths changed with the motion of the arytenoid, by which changes a strain and strain history could be calculated.

The vocal ligament and the TA muscle, both highly fibrous tissues intrinsic to the vocal folds, were designed as fiber–gel composites. The vocal ligament was modeled in a manner similar to the TA submodel discussed above, but with its own parameters (Table I and II) and without active properties. Its position was parallel to the vocal fold's medial edge (Fig. 7, Volume 1). The TA_V (Volume 2) and TA_M (Volume 3) were modeled as independent fiber–gel composites. The TA_M fibers were oriented in a lateral direction, while the TA_V fibers were oriented medially (Table IV).

As was done in the submodel, the total vector forces (from the fiber model) in the TA_M, TA_V, and vocal ligament had to be reduced by the amount of passive forces already within their finite elements. The cross-sectional areas of the TA fiber–gel in the posturing model (as shown in Fig. 7) were smaller than in the reported literature (Table I); therefore, the passive forces from the finite elements were smaller than the total passive force that should be in the TA. However, because the fiber model already included the total passive forces, the applied vector forces could compensate for these differences.

Like the TA submodel, the solution time sampling was allowed to be variable to maintain stability in the posturing model (consisting of 1721 finite elements with 168 superimposed fiber models). The average time sampling was approximately 1500 Hz, with each time sample costing an average of 36 s of CPU time on an HP B2600 workstation (i.e., one second of a dynamic solution expended approximately 54 000 s of CPU time).

F. Model solution with resulting postural movement measures

Glottal angle, as well as its rate of change, was measured from the model. The glottal angle was chosen as the output measure because it was the only absolute vocal fold measure that could be obtained from humans by the standard clinical endoscopic techniques used to understand vocal fold motion (Omori *et al.*, 1997). As only the right fold was modeled, the glottal angle was defined as twice the angle between the midline and the vocal process excursion with the anterior commissure as the vertex. Locating precise glottal landmark positions is difficult with endoscopic techniques; thus, variations were accounted for by using multiple landmarks from the model to represent a range of potential anterior commissure and vocal process positions. Two possible angle vertices were used, the superior and inferior edges of the anterior commissure [Fig. 2(b), both arrows in point D]; also, two possible vocal process positions were used, the superior edge and the most medial point of the vocal fold at the model's vocal process ($y=0$, at rest) (point E). Therefore, four different glottal half angles were calculated from five points: two possible anterior commissure positions, two possible vocal process positions, and the glottal midline.

III. RESULTS

A. Typical results

Each simulated muscle (i.e., IA, LCA, PCA, TA_V, and TA_M) was contracted individually with 100% activation. As previously discussed, the fiber model parameters were identical for all of these muscles, even though their anatomical properties were not. The LCA muscle contraction (Fig. 8) demonstrated a typical result of the four glottal angles and angle rates of the other muscles (see all muscle results in Fig. 9). When the LCA contracted, the top surface was deformed, the vocal fold was no longer planar, and the medial surface did not have the uniform curvature of the neutral position [Fig. 8(c)] [see Hunter *et al.* (in *preparation*) and Hunter and Titze (2004), for a multimedia representation of the dynamics]. After 100 ms of LCA contraction, the glottis had closed approximately 12.5° from cadaveric [Fig. 8(a)]. Depending on which landmark was used for the angle calculation, the peak angular rate varied between −120° and −150° per second and occurred at about 14.5 ms from LCA muscle activation [Fig. 8(b)]. The same three measures (the total change in glottal angle from the initial or cadaveric position, peak angular rate, and time to peak angular rate) were taken for the other four muscles (Fig. 9).

B. Validation

Three studies from the literature were used to validate the model's results. First, Hirano *et al.* (1988) found in a study of 20 excised human larynges (10 male) that the maximum glottal angle opening for males was an average of 35°

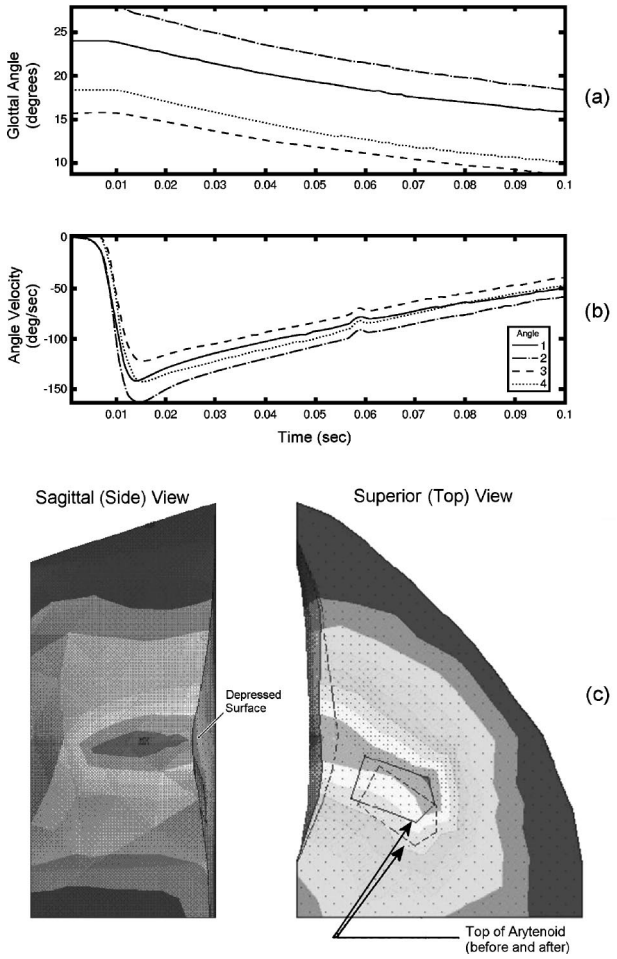


FIG. 8. Modeled lateral cricoarytenoid muscle contraction at full activity. Contours of four glottal angle measures taken from different landmark locations shown to 0.1 s (a) glottal angle and (b) the corresponding glottal angle rate (velocity). The magnitude of deformation (c) of the model (shading) showing the topology change from the initial position. Dotted lines represent edges at the initial position. The top of the vocal fold was depressed, as seen on the side view.

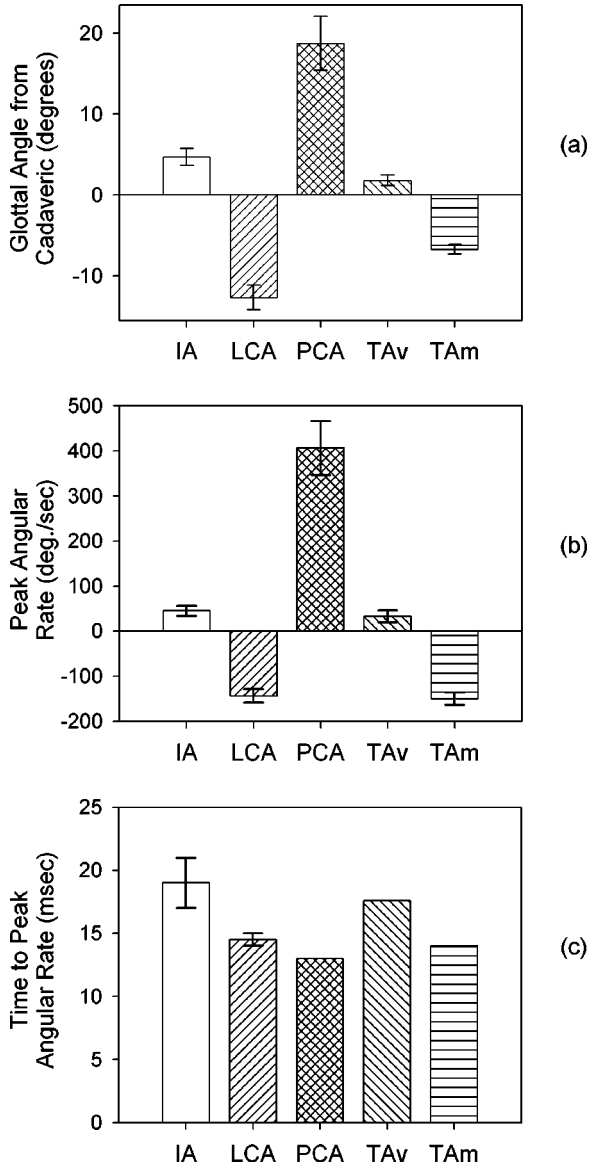


FIG. 9. The average of all four possible glottal angles variants depict the results of the individual 100% muscle contraction for 0.1 s: (a) the change in the glottal angle from the cadaveric (initial) position; (b) the peak glottal angle rate; and (c) the time from contraction start to the peak glottal angle rate.

(calculated from the maximum simulated abduction). The model had a comparable average maximum glottal angle opening of 31° (±5), as calculated by combining the results of a LCA and PCA contraction (both measured as a change from the cadaveric angle). Second, the dynamic posturing motions of each muscle in the model was nearly completed in all cases at 150 ms. This duration was within the range that Hillel (2001) measured for most abduction/adduction gesture durations (between 125 and 350 ms).

Finally, and most significantly, Cooke *et al.* (1997) used an endoscope to examine the vocal fold kinematics prior to various glottal onsets (e.g., breathy, normal, and hard). Using the published data provided in this study, the speed of glottal closure could be extracted. The average beginning glottal angle was 20° for the hard onset, which most likely had 100% LCA contraction. Cooke *et al.* found that the hard onset gesture duration was an average of approximately 0.1 s, where the abduction gesture duration was the time it took for the glottis to move from 80% to 20% maximum distance (or 60% total extent) and the gesture end was either glottal closure or phonatory onset. Assuming that the gesture end occurred with a glottal angle between 0° and 5°, the total gesture motion would be 15° to 20°. The result would be a gesture extent of 9° to 12°, or 60% total extent×(15° to 20° total glottal motion). The average glottal velocity over a gesture duration would then be 90° to 120°/s, or (9° to 12° glottal extent)/0.1 s gesture duration.

Comparing Cooke *et al.* (1997) to the posturing model, 80% to 20% of the maximum adductor motion (LCA contraction) occurred in about 0.08 s, with a gesture extent of about 7.5°, or 60% total extent ×12.5° total gesture motion. The average glottal angle velocity over the gesture duration was then 94° per second, or 7.5°/0.08 s. Though the average glottal velocity values were only approximate in both cases, the posturing model did closely predict the speed of adduction seen by Cooke *et al.*

C. Posturing observations and insights

The current model provided three insights into abduction/adduction. First, it is traditionally accepted that the speed of vocal onset/offset gestures depends on differences in the contractile times of laryngeal muscles. Pure physiological studies of the intrinsic laryngeal muscles via fiber typing experiments have suggested that the PCA should be one of the fastest muscles in the laryngeal system (Malmgren *et al.*, 1985) and that the TA muscle should be even faster (Teig *et al.*, 1978; Guida and Zorzetto, 2000). Although the model did not dispute these histological findings, it did suggest that the timing of abduction/adduction gestures and the resultant speed from muscle contractions in this complex multimuscle system is also affected by the passive mechanics of the vocal folds. As each muscle was contracted individually, the modeled abduction/adduction gestures were most sensitive to the PCA contraction, which resulted in the greatest maximum angular speed and the shortest rise time to that maximum speed when compared to the other muscles. The laryngeal muscles in the model were designed with identical material properties; therefore, the differences between the response speed from individual muscle contractions were the consequence of the anatomical and mechanical properties and not the contractile properties.

Further, the model provided insights into the muscle's participation in opening and closing the glottis. Full glottal opening in the model was accomplished by a PCA contraction, but full glottal closure was not accomplished by any one muscle in isolation. The LCA came closest to closing the glottis, and the TA_M also closed it significantly. These findings substantiated the study by Hillel (2001), in which he showed that the PCA could independently open the glottis and that the LCA, IA, and TA_M contracted simultaneously to close the glottis.

The third insight provided by the model was that the TA_V and the TA_M played different roles in abduction/adduction, with the TA_M significantly adducting the folds and the TA_V just slightly abducting the folds. These different capabilities corroborate the statement by Sanders *et al.* (1998) that the TA has more than one functionally different compartment. Further, since the TA is connected to both the thyroid cartilage and the arytenoid cartilage, it might be assumed that the movement of the thyroid by the cricothyroid joint (CTJ) would strongly affect arytenoid motion and, thus, abduction/adduction. However, the orientation of the TA_V and TA_M (Fig. 7) suggests that the CTJ, which elongates the vocal folds primarily in the anterio-posterior direction (with some vertical change), would thereby elongate the TA_V more than the TA_M. Thus, because the TA_V plays little role in abduction, and because the TA_V and TA_M had opposite effects on the glottal angle, the CTJ movement (acting primarily through the TA_V) should have little effect on abduction/adduction. This finding substantiates and provides mechanical insight into the observation of Woodson *et al.* (1998) that the cricothyroid muscle (CT) does not abduct or adduct the vocal folds; the model validates this observation because the only way for the CT to abduct/adduct the vocal folds would be through the CTJ. It should be noted that although the model shows that the TA_V does not play a significant role in abduction/adduction posturing, it does not negate its role in shortening (via the CTJ) and in vocal fold stiffening. Nevertheless, because these two types of posturing were beyond the scope of the current abduction/adduction study, they were not quantified and were left to future studies.

IV. DISCUSSION

The goal of this research was to design a three-dimensional, biomechanical model of laryngeal abduction/adduction. Because of the potential for significant computational costs in creating such a model, certain simplifying assumptions are necessary in order to make the solution practical; nevertheless, it is crucial to balance this need to simplify with the need to maintain those essential characteristics that would otherwise invalidate the model. The current model seemed to capture key characteristics of laryngeal abduction/adduction in promising ways. The simplifications used, however, did impose limitations that create opportunities for future research. Still, expanding the current model beyond straightforward refinements (e.g., improved material property values and differentiated muscle contraction times) would increase the already significant computational costs of the model.

A. Current limitations and future refinements

The posturing model was dependent on material property data, including active and passive muscle behavior. Much of this data was nonlinear (e.g., passive stress–strain curves); further, vocal folds often shorten 30% or elongate 50% from rest during some types of posturing. Therefore, it would not be unusual for vocal tissues to be in their nonlinear stress–strain region, where small changes in strain would produce large changes in stress. When measuring the stress–strain response of nonlinear tissue, an appropriate reference length is often difficult to measure. Because strain is a function of elongation with respect to reference length, errors in measuring the reference length might cause large misrepresentations of a muscle's passive stress.

These errors could be exacerbated by the prestrained nature of vocal tissues, which have seldom (if ever) been mentioned in literature when defining reference length. Perlman *et al.* (1984) observed that thyroarytenoid muscles became 22.5% to 41% shorter following dissection from viable tissue samples; other laryngeal muscles showed similar prestrained effects. Therefore, prestrained fibrous vocal tissues at equilibrium within the vocal system are already close to their nonlinear stress–strain region. This issue is particularly significant when modeling the vocal ligament, which has been observed to be more nonlinear than other vocal soft tissues (Min *et al.*, 1995). The posturing model used the ligament data from Min *et al.*, which did not mention prestraining or explain how the reference length, was defined. Given the highly nonlinear nature of the ligament stress–strain curve, misquantifying the prestraining could have had a large effect on the model's ability to predict the ligament's contribution to posturing. This limitation becomes particularly important if subsequent laboratory experiments confirm the preliminary observations of vocal ligament prestrain (between

40% and 50%). Until the sensitivity to such errors is known, the posturing model (as well as any other model depending on muscle data), is incomplete.

There are potentially many other ways to refine this model. First, the force representation of the interarytenoid muscle might not have been ideal and, thus, should be refined. All laryngeal muscle forces (except the thyroarytenoid) were simplified to single vector forces acting at a point on the arytenoid cartilage. The human interarytenoid muscle has two parts, the transverse and oblique portions, with the transverse part covering the entire posterior surface of the arytenoid; however, the modeled interarytenoid was based on a canine model that does not have these two portions. Because the two human interarytenoid portions probably have different roles in posturing, the average of the two may be quite different from the average canine interarytenoid orientation.

This limitation might be addressed if future refinements included multiple muscle fiber origin and insertion points for the interarytenoid and muscles other than the already included thyroarytenoid. The raw muscle bundle data studies by Cox *et al.* (1999) (human and canine thyroarytenoid and cricothyroid muscles) and Mineck *et al.* (2000) (canine abductory and adductory muscles) have recently become available to the authors for further analysis (only the overall averages were originally reported). The future analysis of these data will provide more details about the individual fibers, rather than just the overall averages; these more complete details would be beneficial to the community for use in models of the vocal folds. In addition, Selbie *et al.* (2002) used MRI scanning techniques to obtain three-dimensional human laryngeal cartilage geometries; however, only landmark distances were reported. The eventual availability of a full database of geometries and muscle strand measures from these studies would accelerate the refinement and realism of the FEM model.

Furthermore, the complex motion and properties of the cricoarytenoid joint were simplified to a rocking–sliding axis. Although the model captured most of the characteristics of this joint, a rocking–sliding axis is only a first-order approximation and might not have optimally demonstrated the interarytenoid muscle's contributions to posturing. It is possible that the interarytenoid requires degrees of freedom for the cricoarytenoid joint not represented by the rocking–sliding axis. However, such a revision could not be implemented without more experimental studies of movement restraints.

Future refinements of the model should also include motions other than abduction/adduction posturing, such as vocal fold elongation via the cricothyroid joint. This inclusion would nevertheless add a significant calculation time to an already computationally intensive model unless appropriate simplifications could be made. One possible simplification would be to reduce cricothyroid joint motion to a single degree of freedom like the rigid-body model described by Titze *et al.* (1988). However, this simplification (which could only elongate the folds) would not take into account the change in the thyroarytenoid's fiber direction (up and down) as the thyroid rotates with respect to the cricoid. Another option would be to model the entire joint (e.g., thyroid cartilage, cricoid cartilage, and surrounding structure) with finite elements. Nonetheless, adding just the thyroid cartilage to the current model would substantially increase the complexity of the model because of the need to also address the structures and tissues external to the thyroid, as well as their effects on the thyroid (e.g., tissue drag). Therefore, this development was left to a future study.

Another refinement might be to include the respective behavioral differences of the laryngeal muscles. Of the muscles used in this model, only the thryoarytenoid has been sufficiently measured and quantified to fill all of the modified-Kelvin fiber model parameters. The absence of these property differences, however, might have caused some overall misrepresentation of muscle effects. Nevertheless, a necessary first step was to test the physical plant independent of active muscle property differences. Currently, the muscle properties of the lateral cricoarytenoid, posterior cricoarytenoid, and interarytenoid are being studied (Alipour *et al.*, in preparation). Future versions of the model might include differences in muscle contraction and relaxation times.

B. Applications of the model

It is expected that future versions of the model could have applications to vocal pathology and vocal training. For example, one application would be to examine how mechanical stress is distributed throughout the tissue for a given posture (Gunter, 2003). This would be particularly important in studies of abnormalities in which stress is believed to be a contributing factor (e.g., hyperadduction, bowing, and excessive glottalization). Another application might be to model a scar or a tissue augmentation, for which the material properties or local volume could be selectively changed.

Finally, a future model could be used to improve phonosurgical procedures. For example, the model could simulate an arytenoid adduction or thyroplasty, which is used to medialize a paralyzed vocal fold. An optimization might be attempted to find out which technique (or combination of techniques) would deliver the best vocal quality and answer two important questions posed by Neuman *et al.* (1994): (1) Does variation in suture placement anterior to the thyroid cartilage affect forward tilting of the arytenoid cartilage, thereby allowing for a correction of vocal fold posturing at different vertical levels?; and (2) how do other medialization procedures (e.g., thyroplasties, where an implant pushes the vocal folds medially) affect arytenoid movement?

ACKNOWLEDGMENT

This work was supported by NIDCD grant No. DC04347-03 from the National Institutes of Health.

Alipour, F., Titze, I. R., Hunter, E. J., and Tayama, N., (in preparation), "Active and passive properties of canine abduction/adduction laryngeal muscles," Ann. Otol. Rhinol. Laryngol.

Alipour, F., and Titze, I. R. (**1999**). "Active and passive characteristics of the canine cricothyroid muscles," J. Voice **13**, 1–10.

Alipour-Haghighi, F., and Titze, I. R. (**1991**). "Elastic models of vocal fold tissues," J. Acoust. Soc. Am. **90**, 1326–1331.

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.
Bathe, K. J. (**1996**). *Finite Element Procedure* (Prentice–Hall, Englewood Cliffs, NJ).
Berry, D. A., Moon, J. B., and Kuehn, D. P. (**1999**). ''A finite element model of the soft palate,'' Cleft Palate Craniofac J. **36**, 217–223.
Broad, D. J. (**1968**). ''Some physiological parameters for prosodic description,'' SCRL Monograph Number 3, Speech Communications Research Laboratory, Santa Barbara, CA.
Chan, R. W., and Titze, I. R. (**1998**). ''Viscosities of implantable biomaterials in vocal fold augmentation surgery,'' Laryngoscope **108**, 725–731.
Cook, R. D., Malkus, D. S., and Plesha, M. E. (**1989**). *Concepts and Applications of Finite Element Analysis* (Wiley, New York).
Cooke, A., Ludlow, C. L., Hallett, N., and Selbie, W. S. (**1997**). ''Characteristics of vocal fold adduction related to voice onset,'' J. Voice **11**, 12–22.
Cooper, D. S., Shindo, M., Sinha, U., Hast, M. H., and Rice, D. H. (**1994**). ''Dynamic properties of the posterior cricoarytenoid muscle,'' Ann. Otol. Rhinol. Laryngol. **103**, 937–943.
Cox, K. A., Alipour, F., and Titze, I. R. (**1999**). ''Geometric structure of the human and canine cricothyroid and thyroarytenoid muscles for biomechanical applications,'' Ann. Otol. Rhinol. Laryngol. **108**, 1151–1158.
Dang, J., and Honda, K. (**1998**). ''A physiologic model of a dynamic vocal tract for speech production,'' Technical Report No. 43.70.Aj, 43.70.Bk, ATR Human Information Processing Research Laboratories.
Eckel, H. E., Sittel, C., Zorowka, P., and Jerke, A. (**1994**). ''Dimensions of the laryngeal framework in adults,'' Surg. Radiol. Anat. **16**, 31–36.
Eckel, H. E., and Sittel, C. (**1995**). ''Morphometry of the larynx in horizontal sections,'' Am. J. Otol. **16**, 40–48.
Farley, G. R. (**1996**). ''A biomechanical laryngeal model of voice F-0 and glottal width control,'' J. Acoust. Soc. Am. **100**, 3794–3812.
Guida, H. L., and Zorzetto, N. L. (**2000**). ''Morphometric and histochemical study of the human vocal muscle,'' Ann. Otol. Rhinol. Laryngol. **109**, 67–71.
Gunter, H. E. (**2003**). ''A mechanical model of vocal-fold collision with high spatial and temporal resolution,'' J. Acoust. Soc. Am. **113**, 994–1000.
Hillel, A. D. (**2001**). ''The study of laryngeal muscle activity in normal human subjects and in patients with laryngeal dystonia using multiple fine-wire electromypgraphy,'' Laryngoscope **111**, 1–47.
Hirano M., Kiyokawa, K., and Kurita, S. (**1988**). ''Laryngeal muscles and glottic shaping,'' in *Vocal Physiology: Voice Production, Mechanisms, and Functions*, edited by O. Fujimura (Raven Press, Ltd., New York).
Hirano, M., and Sato, K. (**1993**). *Histological Color Atlas of the Human Larynx* (Singular Pub. Group Inc., San Diego).
Hirano, M., Vennard, W., and Ohala, J. (**1970**). ''Regulation of register, pitch and intensity of voice. An electromyographic investigation of intrinsic laryngeal muscles,'' Folia Phoniatr. **22**, 1–20.
Honda, K. (**1983**). ''Variability analysis of laryngeal muscle activities,'' in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, Denver, CO: The Denver Center for the Performing Arts, edited by I. Titze and R. Scherer.
Hunter, E. J. (**2001**). ''Three-dimensional biomechanical model of vocal fold posturing,'' Ph.D. dissertation, University of Iowa.
Hunter, E. J., Titze, I. R., and Alipour, F. (**in preparation**). ''Review of range of arytenoid motion compared to models of vocal abduction/abduction,'' ARLO.
Hunter, E. J., Titze, I. R., (**2004**). ''Dynamics of a three-dimensional model of vocal fold abduction/adduction,'' NCVS Online Technical Reports, February 2004. Available from URL: http://www.nevs.org/nevs/library/tech.
Kotby, M. N., Basiouny, S. E., Amin, M., Garrett, D., Kirchner, J. A., and Kahane, J. C. (**1992**). ''Pattern of gross displacement of the vocal fold in adduction and abduction movements,'' Acta Oto-Laryngol. **112**, 349–352.
Li, S. T. (**1995**). ''Biologic biomaterials: Tissue-derived biomaterials (collagen),'' in *The Biomedical Engineering Handbook*, edited by J. D. Bronzino (CRC Press, Boca Raton), pp. 627–647.
Malmgren, L. T., Gacek, R., and Etzler, C. A. (**1985**). ''Muscle fiber types in the human posterior cricoarytenoid muscle: A correlated histochemical and ultrastructure morphometric study,'' in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, Denver, CO, The Denver Center for the Performing Arts, edited by I. Titze and R. Scherer, pp. 41–56.
Min, Y. B., Titze, I. R., and Alipour-Haghighi, F. (**1995**). ''Stress–strain response of the human vocal ligament,'' Ann. Otol. Rhinol. Laryngol. **104**(7), 563–569.
Mineck, C. W., Tayama, N., Chan, R., and Titze, I. R. (**2000**). ''Three-dimensional anatomic characterization of the canine laryngeal abductor and adductor musculature,'' Ann. Otol. Rhinol. Laryngol. **109**, 505–513.
Murry, T., Xu, J. J., and Woodson, G. E. (**1998**). ''Glottal configuration associated with fundamental frequency and vocal register,'' J. Voice **12**, 44–49.
Neuman, T. R., Hengesteg, A., Lepage, R. P., Kaufman, K. R., and Woodson, G. E. (**1994**). ''Three-dimensional motion of the arytenoid adduction procedure in cadaver larynges,'' Ann. Otol. Rhinol. Laryngol. **103**, 265–270.
Omori, K., Slavit, D. H., Matos, C., Kojima, H., Kacker, A., and Blaugrund, S. M. (**1997**). ''Vocal fold atrophy: quantitative glottic measurement and vocal function,'' Ann. Otol. Rhinol. Laryngol. **106**, 544–551.
Perlman, A. L. (**1985**). ''A technique for measuring the elastic properties of vocal fold tissue,'' Ph.D. dissertation. The University of Iowa.
Perlman, A. L., Titze, I. R., and Cooper, D. S. (**1984**). ''Elasticity of canine vocal fold tissue,'' J. Speech Hear. Res. **27**, 212–219.
Sanders, I., Rao, F., and Biller, H. F. (**1994**). ''Arytenoid motion evoked by regional electrical stimulations of the canine posterior cricoarytenoid muscle,'' Laryngoscope **104**, 456–462.
Sanders, I., Han, Y., Wang, J., and Biller, H. (**1998**). ''Muscle spindles are concentrated in the superior vocalis subcompartment of the human thyroarytenoid muscle. [Review] [45 refs],'' J. Voice **12**, 7–16.
Selbie, W. S., Zhang, L., Levine, W. S., and Ludlow, C. L. (**1998**). ''Using joint geometry to determine the motion of the cricoarytenoid joint,'' J. Acoust. Soc. Am. **103**, 1115–1127.
Selbie, W. S., Gewalt, S. L., and Ludlow, C. L. (**2002**). ''Developing an anatomical model of the human laryngeal cartilages from magnetic resonance imaging,'' J. Acoust. Soc. Am. **112**, 1077–1090.
Tayama, N., Chan, R. W., Kaga, K., and Titze, I. R. (**2001**). ''Geometric characterization of the laryngeal cartilage framework for the purpose of biomechanical modeling,'' Ann. Otol. Rhinol. Laryngol. **110**, 1154–1161.
Tayama, N., Chan, R. W., Kaga, K., and Titze, I. R. (**2002**). ''Functional definitions of vocal fold geometry for laryngeal biomechanical modeling,'' Ann. Otol. Rhinol. Laryngol. **111**, 83–92.
Teig, E., Dahl, H. A., and Thorkelsen, H. (**1978**). ''Actomyosin ATPase activity of human laryngeal muscles,'' Acta Oto-Laryngol. **85**, 272–281.
Titze, I. R. (**1988**). ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**1996**). ''Coupling of neural and mechanical oscillators in control of pitch, vibrato, and tremor,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular Pub. Group, San Diego), pp. 47–62.
Titze, I. R., Jiang, J., and Druker, D. G. (**1988**). ''Preliminaries to the body-cover model of pitch control,'' J. Voice **1**, 314–319.
Titze, I. R., and Sundberg, J. (**1992**). ''Vocal intensity in speakers and singers,'' J. Acoust. Soc. Am. **91**, 2936–2946.
Tomori, Z., Benacka, R., and Donic, V. (**1998**). ''Mechanisms and clinico-physiological implications of the sniff- and gasp-like aspiration reflex,'' Respir. Physiol. **114**, 83–98.
Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissue articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098.
Woodson, G., Murry, M. P., Schweizer, V., Hengesteg, A., Chen, N., and Yeung, D. (**1998**). ''Unilateral cricothyroid contraction and glottic configuration,'' J. Voice **12**, 335–339.
Yoshioka, H. (**1981**). ''Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese,'' Phonetica **38**, 236–251.

On the ability of a physiologically constrained area function model of the vocal tract to produce normal formant patterns under perturbed conditions[a)]

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

(Received 12 June 2003; revised 22 January 2004; accepted 1 February 2004)

An area function model of the vocal tract is tested for its ability to produce typical vowel formant frequencies with a perturbation at the lips. The model, which consists of a neutral shape and two weighted orthogonal shaping patterns (modes), has previously been shown to produce a nearly one-to-one mapping between formant frequencies and the weighting coefficients of the modes [Story and Titze, J. Phonetics, **26**, 223–260 (1998)]. In this study, a perturbation experiment was simulated by imposing a constant area ''lip tube'' on the model. The mapping between the mode coefficients and formant frequencies was then recomputed with the lip tube in place and showed that formant frequencies ($F1$ and $F2$) representative of the vowels [ʊ,o,u] could no longer be produced with the model. However, when the mode coefficients were allowed to exceed their typical bounding values, the mapping between them and the formant frequencies was expanded such that the vowels [ʊ,o,u] were compensated. The area functions generated by these exaggerated coefficients were shown to be similar to vocal-tract shapes reported for real speakers under similar perturbed conditions [Savariaux, Perrier, and Orliaguet, J. Acoust. Soc. Am., **98**, 2428–2442 (1995)]. This suggests that the structure of this particular model captures some of the human ability to configure the vocal-tract shape under both ordinary and extraordinary conditions. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1689347]

PACS numbers: 43.70.Bk [Al] Pages: 1760–1770

I. INTRODUCTION

The ability of the speech motor system to respond to a perturbation or constraint imposed on some part of the vocal tract is well known. In ''everyday'' experience it is not uncommon to observe talkers balancing the needs of speech production simultaneously with some other demand on the articulators. For example, speaking while clenching an object between the teeth (e.g., pen, toothpick, etc.) requires articulatory patterns that differ from those of typical speech. Yet, assuming the perturbation is not too extreme, speakers are generally able to produce intelligible speech (although the quality of the speech sound may carry information concerning the nature of the perturbation).

Articulatory compensation of this sort has often been systematically studied with a perturbation paradigm. In such experiments, some part of a talker's articulatory system is deliberately perturbed or constrained relative to the typical manner of speaking. As an example, Lindblom, Lubker, and Gay (1979) showed that speakers quickly compensated (within the first glottal cycle) for the imposition of a bite block constraining the distance between the upper and lower teeth, such that measured formant frequencies were nearly identical to those in an unperturbed condition. Whether experimental or otherwise, imposed perturbations limit the mechanical range of motion of an articulatory structure and apparently cause the speech motor system to alter the movement of other articulatory structures so that appropriate (or nearly so) acoustic characteristics are maintained.

Other investigations have considered constraints that limited the possible cross-sectional areas within specific portions of the vocal tract, rather than perturbation of a single articulator. Sorokin, Olshansky, and Kozhanov (1998) showed that laryngectomized patients, whose lower vocal tracts were radically altered by surgery, could learn to produce vowels with formant frequencies nearly the same as those recorded prior to the surgery. Similar results were reported by Hamlet and Stone (1976, 1978), who fitted speakers with an oral prosthesis that altered the morphology of the oral cavity. Speakers modified their speech production to compensate for the change in vocal-tract shape. In these latter two cases, compensation occured on a much longer time scale (about 1–2 weeks) than for a single articulator perturbation, but speakers nonetheless found new articulation patterns that produced the desired acoustic signals.

An experiment designed specifically to observe the response of the entire vocal-tract shape to a perturbation was reported by Savariaux, Perrier, and Orliaguet (1995), henceforth referred to as ''SPO95.'' They made use of a small Plexiglas tube to perturb the vocal-tract shape during production of the vowel [u] of 11 native French speakers. The tube imposed a constraint on the cross-sectional area of the vocal tract at, and just posterior to, the lips. The task for each speaker in the study was to produce the perturbed [u] as close to their natural production as possible. The observations, based on midsagittal x-ray images and formant fre-

[a)]A preliminary version of this paper was presented at the 140th Meeting of the Acoustical Society of America.
[b)]Electronic mail: bstory@u.arizona.edu

 0001-4966/2004/115(4)/1760/11/$20.00

quency analysis of audio recordings, showed that the speakers fell into two distinct groups. The first group showed a small amount of tongue displacement in response to the lip tube. Consequently, the formant frequencies in this perturbed condition were not effectively maintained near those of the unperturbed [u] vowel. The second group adapted to the lip tube by displacing the tongue posteriorly, causing the primary constriction location to be moved backward. This had the effect of increasing the length and cross-sectional areas of the front vocal-tract cavity. Acoustically, these modifications of the vocal-tract shape did reduce the difference between the formants in the unperturbed and perturbed conditions, but did not achieve full compensation. Only one speaker displaced the tongue body backward far enough to fully compensate for the perturbation in terms of $F1$ and $F2$; however, $F3$ was significantly higher than in the unperturbed condition.

While relatively simple in design, this lip-tube experiment provides an ideal test case for the development of speech production models that are intended to account for and possibly replicate human speech. Guenther, Hampson, and Johnson (1998) simulated the lip-tube perturbation with their model of speech acquisition and production in order to test its compensation abilities. Their model responded to the lip tube by retracting the tongue to create a pharyngeal constriction, similar to the one subject in the SPO95 study who achieved full compensation in terms of $F1$ and $F2$. In general, a model should be able replicate the extreme articulations that humans can produce, even though most of the time a speaker may opt for less extreme productions.

The focus of this paper concerns the control of formant frequencies based on an area function model of the vocal tract. Specifically, the interest is in modeling the systematic deformation of the vocal-tract shape to appropriately position the formant frequencies for vowel sounds, even under perturbed conditions. The model used in this study is based on area functions obtained from magnetic resonance imaging (Story, Titze, and Hoffman, 1996) and consists of a ''neutral'' shape and two orthogonal shaping patterns (called *modes*) that are parametrically weighted to generate a wide range of vowel shapes (Story and Titze, 1998). Similar models have been developed from factor analysis of midsagittal tongue profiles obtained from midsagittal x-ray pictures (Harshman, Ladefoged, and Goldstein, 1977; Maeda, 1990) or with principal components analysis of area functions derived from sagittal x-ray images (Meyer, Wilhelms, and Strube, 1989; Yehia, Takeda, and Itakura, 1996). In these models, 2–5 factors or components have been found to account for the majority of the variance, facilitating development of low-dimensional models.

The present model has been used for mapping formant frequencies extracted from natural speech to plausible time-varying area functions (Story and Titze, 1998), and for suggesting some possible mechanisms of voice quality differences among speakers (Story and Titze, 2002). But, it has not been tested with regard to perturbations of the vocal tract. That is, if a certain section of the vocal tract is somehow constrained, can the parameters of the area function model be altered so that a desired acoustic output is maintained? The answer to this question has implications for assessing whether this model is merely descriptive of the area function set on which it is based, or if it is also potentially predictive of humans' abilities to configure the vocal-tract shape under both ordinary and extraordinary conditions.

Toward this goal, a simulation was performed that partially replicated the lip-tube experiment reported by SPO95. The specific aim was to determine whether a reorganization and extension of the control parameters of the area function model would generate a response to the lip-tube perturbation similar to those shown in SPO95. The paper will first briefly review the formulation of the area function model and demonstrate the relation between the model's parameters and formant frequencies. The simulation of the lip-tube perturbation and subsequent compensation strategy will then be described and compared to the results of SPO95.

II. AREA FUNCTION MODEL

The model used in this study specifies a vocal-tract area function $V(x)$ as

$$V(x)=\frac{\pi}{4}[\Omega(x)+q_1\phi_1(x)+q_2\phi_2(x)]^2, \tag{1}$$

where x represents the distance from the glottis (Story and Titze, 1998). The x-dependent terms on the right side of the equation were empirically derived from a principal components analysis (PCA) of a ten vowel set of area functions acquired with magnetic resonance imaging (Story *et al.*, 1996). The PCA was performed on the equivalent diameters of the original ten area functions (see Story and Titze, 2002 for an explanation); thus, the sum of the terms in brackets represents a collection of diameters extending from the glottis to the lips. The squaring operation and scaling factor of $\pi/4$ converts the diameters to areas. The $\Omega(x)$ is the mean diameter function across the ten vowels and $\phi_1(x)$ and $\phi_2(x)$ are the two most significant (in terms of variance accounted for) orthogonal basis functions resulting from the PCA. For purposes of this model, the basis functions are referred to as *modes*. These three functions are shown graphically in Figs. 1(a)–(c); the $\phi_1(x)$ and $\phi_2(x)$ have been smoothed with eighth-order polynomials fitted to the original basis functions. The coefficients of the modes, q_1 and q_2, determine the vocal-tract shape, and their values for reconstructing the ten vowel area functions are given in Fig. 1(d). When $q_1=q_2=0$, the area function specified as $(\pi/4)\,\Omega^2(x)$ produces nearly evenly spaced formant frequencies, which suggests it to be effectively a ''neutral'' vocal-tract configuration. Consequently $\Omega(x)$ is referred to as the ''neutral'' diameter function.

Vocal-tract length can also be varied. Under normal conditions it is set to a constant value of 17.5 cm, but will be extended during simulation of the perturbation in later sections.

The independent effect of each mode [Fig. 1(a) or Fig. 1(b)] on both the area function and the frequency locations of the first two formants, $F1$ and $F2$, can be observed by linearly increasing the numerical value of either mode coefficient (q_1 or q_2) in incremental steps from its minimum to

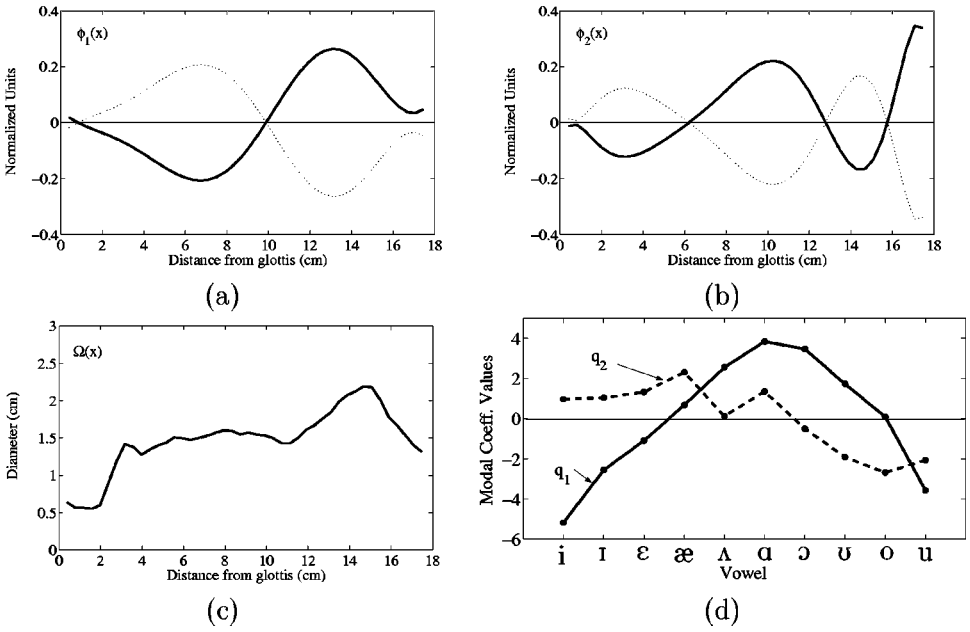


FIG. 1. Representation of an adult male vocal tract with "modes" derived from principal components analysis: (a) first mode shape ϕ_1 (dotted line is the reflection of the solid line); (b) second mode shape ϕ_2; (c) mean diameter function $\Omega(x)$; and (d) mode coefficients for reconstructing the original ten vowels with Eq. (1).

maximum value [see Fig. 1(d) for numerical values], while holding the other coefficient constant at 0. Area functions for each step were generated with Eq. (1), while corresponding formant frequencies were determined by finding the prominent peaks in the computed frequency response function (Sondhi and Schroeter, 1987). Figures 2(a) and (c) show area functions that represent the extremes of the coefficient ranges for each mode along with the neutral shape. Formant frequencies are shown in Figs. 2(b) and (d) as functions of the respective coefficient value. The two modes similarly affect $F1$ and oppositely affect $F2$. That is, as q_1 increases from -5.17 to $+3.85$, $F1$ increases from 300 to 744 Hz while $F2$ decreases from 2184 to 1139 Hz. When q_2 is increased from -2.68 to 2.31, $F1$ rises from 385 to 681 Hz, as does $F2$, increasing from 902 to 1832 Hz.

This relation between the mode coefficients and formant frequencies is further exemplified by plotting, in a 2D plane, the linear change of both mode coefficients, along with those coefficient pairs characterizing the original vowels [Fig. 3(a)]. The corresponding formant frequencies are similarly plotted in Fig. 3(b). For illustrative purposes, the orientation of the plots in both cases has been manipulated so that the traditional articulatory vowel space (i.e., *front* at the left side, *high* at the top) can be shown. The first mode is observed to approximate a continuum extending from a high/front vowel to one that is low/back, while the second mode, which is orthogonal to the first, approximates a low/front to high/back continuum. The physiologic relevance is that the modes seem to capture the basic phonetic dimensions of the tongue's ability to deform the vocal tract into vowel shapes.

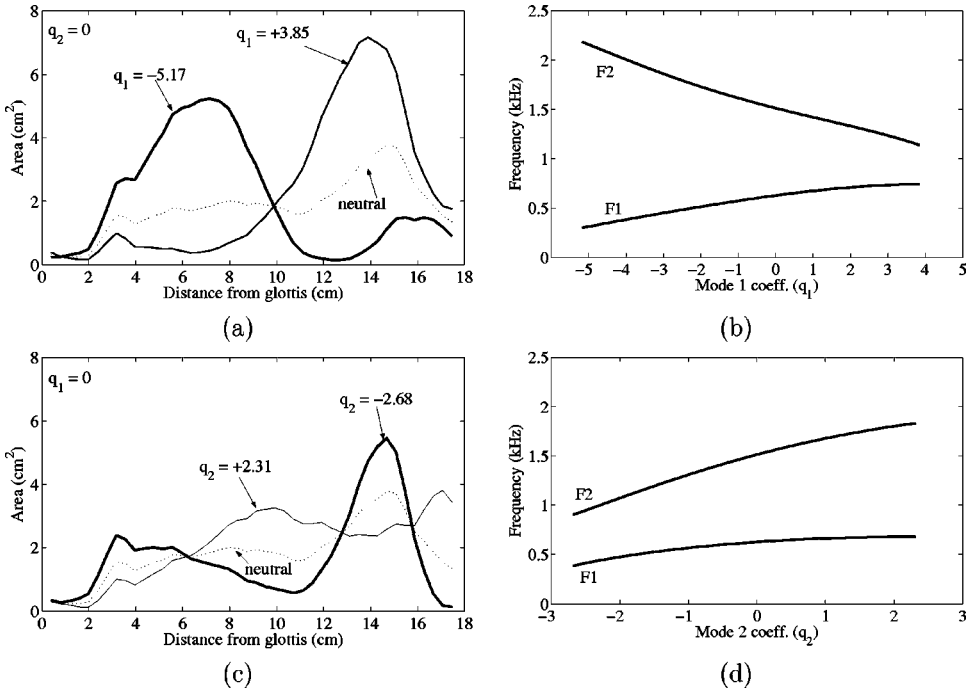


FIG. 2. Spatial and formant frequency effects resulting from superposition of the modes (ϕ_1 and ϕ_2) on the neutral vocal-tract area function. (a) Case where $(q_1,q_2)=(-5.17,0.0)$ (thick) and $(q_1,q_2)=(+3.85,0.0)$ (thin); (b) $F1$ and $F2$ formant frequencies that result from a continuum of q_1 values ranging from -5.17 (left side) to $+3.85$ (right side) while $q_2=0$; (c) case where $(q_1,q_2)=(0.0,-2.68)$ (thick) and $(q_1,q_2)=(0.0,+2.31)$ (thin); (e) $F1$ and $F2$ formant frequencies resulting from a continuum of q_2 values ranging from -2.68 (left side) to $+2.31$ (right side) while $q_1=0$.

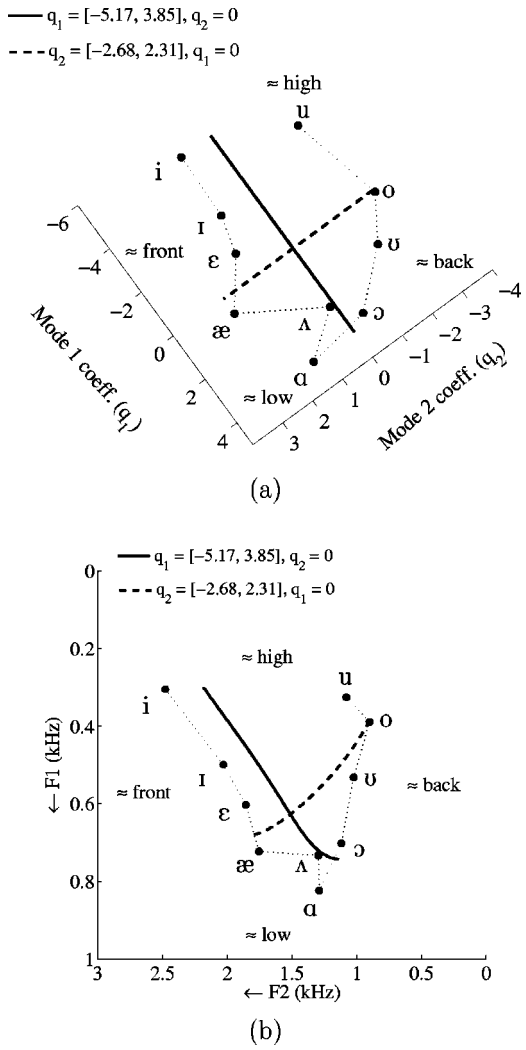


FIG. 3. Relation between the q_1-q_2 coefficient plane and the $F1-F2$ formant plane. (a) The solid and dashed lines represent the ranges of q_1 and q_2, respectively. The coefficient pairs characterizing the original vowels are shown with solid dots connected by the dotted line. The axes have been rotated so that the traditional articulatory dimensions of vowels (high/low and front/back) are represented. (b) Formant pairs represented along the solid line correspond to those coefficient pairs along the solid line in (a). Similarily, the formants along the dashed line correspond to the dashed line in (a) and the formant pairs of the original ten vowels are indicated with solid dots connected with the dotted line.

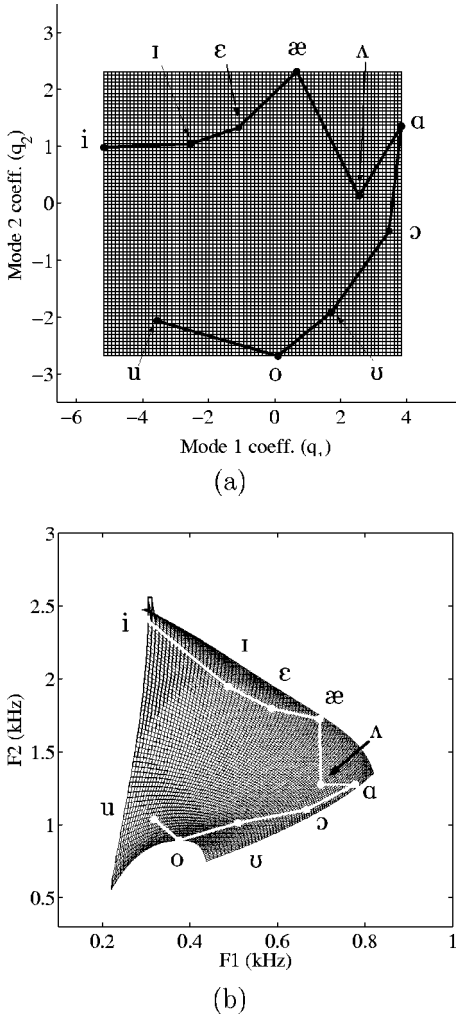


FIG. 4. Mapping of 6400 (q_1,q_2) coefficient pairs to the same number of $F1-F2$ formant pairs, where the solid circles represent the coefficients and formant frequencies for the original ten vowels: (a) coefficient mesh; (b) corresponding $F1-F2$ mesh.

However, because they were determined from area functions spanning the entire length of the vocal tract, the modes must necessarily contain contributions from the other articulators as well as the tongue. Hence, each mode may be thought to represent a coordination of the articulators that simplifies control of the tract shape under normal speaking conditions.

The orthogonality of the modes allows each vowel to be represented by a vector combination of the two mode coefficient values (i.e., a coefficient pair in the q_2 vs q_1 plane) as depicted in the figure. This suggests that a wide range of possible vocal-tract shapes might be generated with other vector combinations of the coefficients that did not exist in the original ten-vowel set. Combining incremental values of the mode coefficients throughout their respective ranges generates a matrix of coefficient pairs. Figure 4(a) shows an 80×80 grid of 6400 coefficient pairs (q_1,q_2) based on the ranges of coefficient values given in Fig. 1(d). The solid dots and dark connecting line indicate the coefficient pairs that reconstruct the original ten vowels. For each coefficient pair in the grid, an area function was created with Eq. (1). Formant frequencies corresponding to each area function were then determined by the same method mentioned previously. The first two formants ($F1$ and $F2$) are plotted as pairs in the formant ''mesh'' given in Fig. 4(b). Note that for both the coefficient and formant plots, the orientation of axes has now been set so that values increase from left to right and from bottom to top. At almost every point in this mesh, there is a formant pair that uniquely corresponds to a coefficient pair in Fig. 4(a). The possible exception is along the upper border where the density of the formant pairs becomes high enough in some places that multiple sets of coefficients may produce area functions with essentially the same formant pattern. Nonetheless, the relation between the coefficient and formant meshes effectively provides a nearly one-to-one mapping between formant frequencies and area functions [via Eq. (1)].

III. SIMULATION OF A VOCAL-TRACT PERTURBATION

In this section, the experiment reported by SPO95 was simulated, by imposing (numerically) the same lip-tube perturbation on the vocal-tract area function as they imposed on real speakers. A coefficient-to-formant relation, like the one shown in Fig. 4, was computed twice under the perturbed condition. In the first, *no* strategy was used to alter the vocal-tract shape in response to the perturbation, while for the second, the minimum and maximum values for each coefficient were allowed to exceed their original boundaries.

For practical reasons, the SPO95 study focused only on perturbation of the vowel [u]. But, for the simulated experiment described in this section, the computed relation between mode coefficients and formant frequencies allows for a view of the *entire* vowel space [produced by Eq.(1)] under both normal and perturbed conditions. Thus, vocal-tract shape changes for any vowel can potentially be observed.

A. Lip-tube constraint without response

The area function model [Eq. (1)] was constrained at the lip end by setting the cross-sectional areas of the final 0.8 cm of vocal-tract length to be 4.9 cm^2. In addition, an extra section of the same cross-sectional area and 1.2 cm in length was added to the end of the vocal tract. Thus, the total lip-tube length was 2.0 cm long and contributed an increase of 1.2 cm to the original vocal-tract length. These tube dimensions are the same as those used by SPO95. The imposed lip tube is demonstrated with the neutral area function $(\pi/4)\,\Omega^2(x)$ in Fig. 5(a), where the area is constant between 16.7–18.7 cm from the glottis. The same tract shape is shown in Fig. 5(b), but as a ''pseudo-midsagittal'' projection. This view is analogous to the anatomical perspective of a real speaker in the midsagittal plane (Story, Titze, and Hoffman, 2001) and is generated by plotting each element of an area function as a line of equivalent diameter perpendicular to a profile of the vocal tract. This view will be used later to demonstrate the compensatory responses to the lip tube.

Using exactly the same coefficient matrix as in the previous section [Fig. 4(a)], but with the lip-tube perturbation in place, a new $F1-F2$ mesh was generated. Figure 6 shows this new ''perturbed'' mesh overlaid on the original (also shown are the formants of the original ten vowels). An apparent discontinuity is observed in the upper-left corner where [i]-like vowels are produced. In these cases, the radiation losses at the lips caused the bandwidths of the upper formants to increase, obscuring the $F2$ peaks in the frequency response. Thus, the peak-picking algorithm used for determining the formant frequencies misinterpreted $F3$ as $F2$. For the the main portion of the mesh, the one-to-one mapping between coefficients and formant pairs is retained, but the presence of the lip tube has compressed the range of $F2$. In particular, the upper edge is lowered relative to the original, due in large part to the increased vocal-tract length imposed by the tube. This causes the original vowel formant pairs for [æ] and [ɑ] to now fall outside the mesh, although just barely. More significant, however, is the lower edge of the mesh, along which the minimum values of $F2$ are significantly higher than those for any of the three vowels [ʊ, o, u] (denoted with black dots). Thus, these vowels can no longer be adequately produced by Eq. (1) when the lip perturbation is imposed. Hypothetically, this is equivalent to a speaker void of volitional effort to compensate for the lip tube.

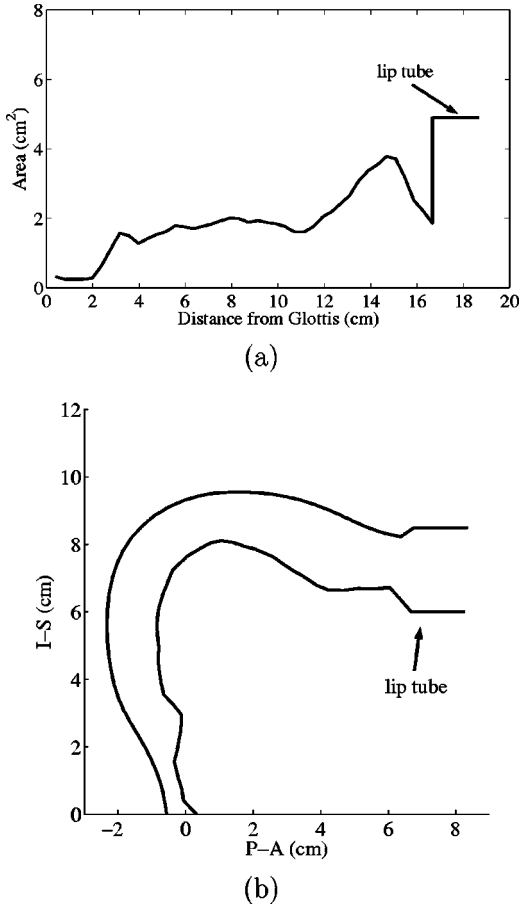


FIG. 5. The neutral vocal-tract shape shown with the lip-perturbation tube (2 cm in length with cross-sectional area of 4.9 cm^2). (a) area function with the lip tube; (b) pseudo-midsagittal plot of the neutral shape with the lip tube in place.

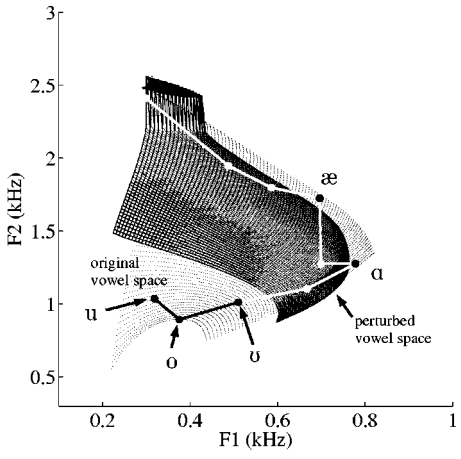


FIG. 6. $F1-F2$ grid that results when the lip tube is in place *and* the same set of modal coefficient pairs are used as in Fig. 4(a). For comparison, the original $F1-F2$ grid is shown underneath with light dots. Note that the lower border of the mesh has been moved upward such that [ʊ, o, u] are no longer within the vowel space. Also, the upper border has been moved downward, leaving [æ] and [ɑ] just outside the new mesh.

B. Lip-tube constraint with compensatory response

In this section, a strategy is presented that allows the perturbed model [Eq. (1) with lip tube] to generate new, compensatory vocal-tract shapes capable of supporting formant frequencies appropriate for the [ʊ, o, u] vowels. To respond to a perturbation such as the lip tube, it is assumed that a reorganization of motor control commands to the speech articulators is required. If successful, a modified vocal-tract shape would result, producing formant frequencies that are the same or nearly the same as those in the unperturbed condition. In the area function model used here, such a reorganization of control commands can be realized by modifying the allowed magnitudes of the mode coefficients. In this sense, the coefficients serve as abstract ''motor commands'' that activate the modes to shape or reshape the vocal tract.

A simulated compensatory response was carried out by allowing the mode coefficients q_1 and q_2 to become more positive and more negative than their original maximum and minimum values, respectively. Accordingly, a new coefficient matrix (grid) was generated that is expanded relative to the original, the boundaries of which can be seen as dashed lines in Fig 7(a). To highlight the increased ranges of both coefficients, the boundaries of the original grid from Fig. 4(a) are also shown in the plot (as solid lines). In this new coefficient space, the rightmost (positive) q_1 boundary is increased by 50% while the left (negative) boundary was unchanged. For q_2, the magnitude of the lower (negative) boundary was increased by 50% (making it a larger negative value), but the upper boundary was left the same as in the original (these amounts of increase or decrease were determined by preliminary simulations).

The $F1-F2$ mesh corresponding to the expanded coefficient space is shown in Fig. 7(b). The upper border of the mesh, including the discontinuity in the upper-left corner, is identical to that shown previously in Fig. 6(b). This is expected since the q_2 coefficient was not increased in the positive direction. However, even in preliminary simulations (not shown) where the positive boundary of q_2 was increased, the upper border of the mesh did not increase significantly in the $F2$ direction. Apparently, the added length of the vocal tract, due to the lip tube, imposes an upper bound on $F2$.

Of greater relevance to the vowels [ʊ, o, u], is the change in the lower border of the mesh, where $F2$ has been decreased enough that the original formant pairs corresponding to these vowels now also exist within the new vowel space. In addition, the new mesh contains a ''tail'' in which $F1$ and $F2$ could take on values as low as 0.3 and 0.6 kHz, respectively. Thus, expanding the coefficient space has allowed the vocal-tract shape to be compensated for the presence of the lip tube.

Coefficient pairs (q_1, q_2) specifically for the vowels [ʊ, o, u] are indicated in Fig. 7(a) for both the original and perturbed/compensated cases. The arrows are suggestive of ''routes'' to compensation of the vocal-tract shape for each vowel when the lip tube is introduced. In other words, under normal speaking conditions the coefficient pairs given by the open circles are appropriate for production of these vowels, but when the lip tube is present the coefficient pairs must be modified to those values represented by the solid dots.

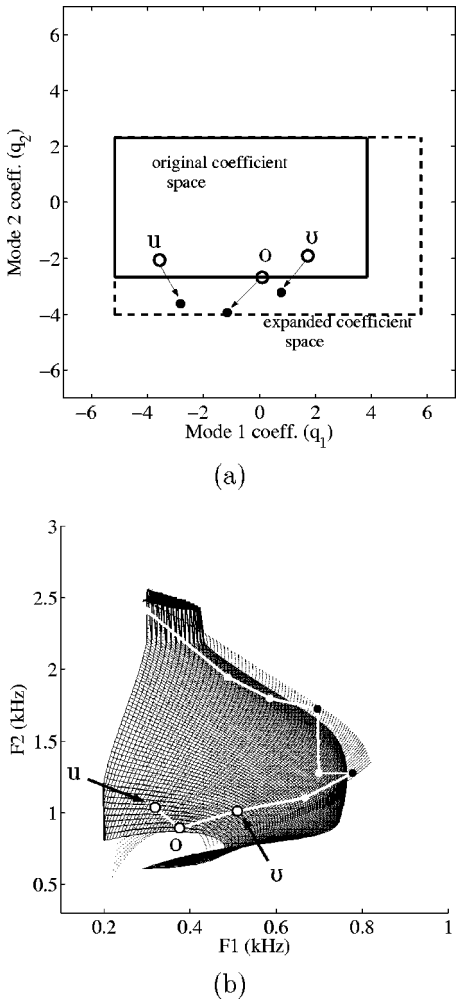


FIG. 7. Mapping of coefficient pairs to $F1-F2$ formant pairs when q_1 and q_2 were extended beyond their original limits. (a) The coefficient space in which q_1 ranged from -5.17 to $+5.77$ and q_2 from -4.03 to $+2.31$ is indicated by the outer rectangle (dashed lines), while the bounds of the original coefficient grid from Fig. 4(a) are shown with solid lines. (b) The $F1-F2$ pairs corresponding to the expanded coefficient space. The original vowel formants are shown with the thick line (and dots), while the original $F1-F2$ grid is shown underneath. Note that the [ʊ, o, u] are now fully compensated.

New area functions for [ʊ] and [o] were generated with Eq. (1), but using the coefficient values specified by their respective points in the expanded coefficient space. They are presented in Fig. 8 as pseudo-midsagittal plots superimposed with similar plots of the original vowel shapes. The compensation of the vocal tract for both vowels consists of a decrease in the cross-sectional area within the midtract constriction (located at L_c in the figure), and a slight expansion of the pharynx. Note that in each case, the perturbed/compensated vocal-tract shape generates the same values of $F1$ and $F2$ as the original shape.

A new area function for the [u] vowel could be similarly generated. However, to facilitate a more detailed comparison to the results of SPO95, three possible compensations for the [u] were explored, as indicated in Fig. 9. The first compensation is the same one shown previously for [u] in Fig. 7, but is now denoted as [u_a]. The other two compensations shown

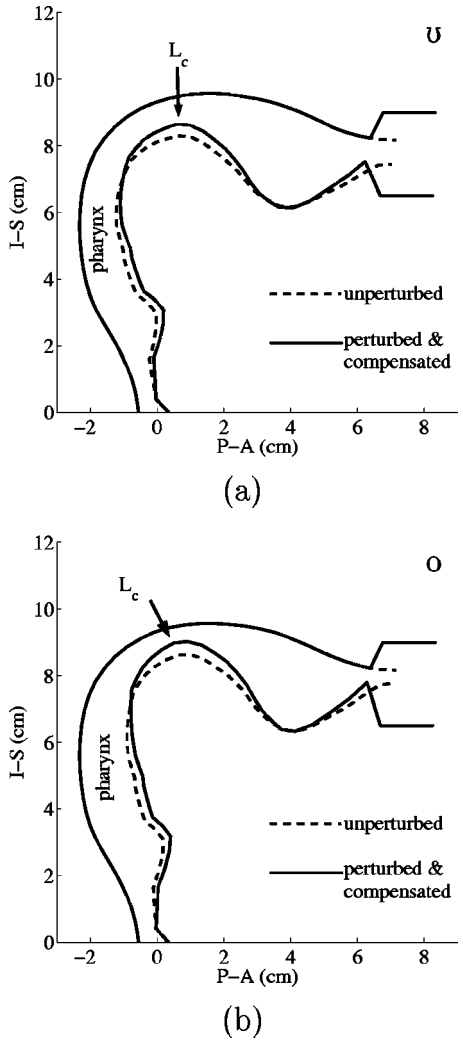


FIG. 8. Comparison of original (dashed line) and perturbed+compensated (thick line) vocal tract shapes for ʊ and o shown as pseudo-midsagittal plots. The perturbed+compensated shapes were generated from the points indicated for these vowels in Fig. 7(a). L_c denotes the locations of minimum cross-sectional area.

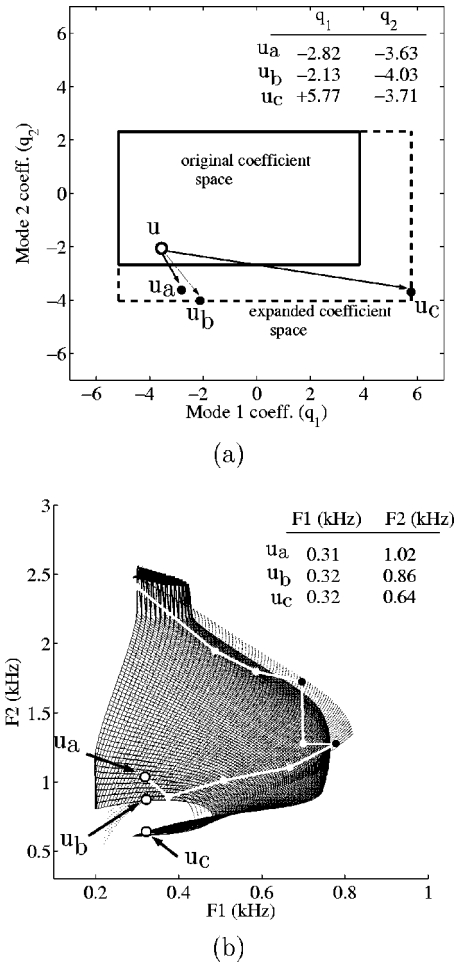


FIG. 9. Replot of the mapping of coefficient pairs to $F1-F2$ formant pairs shown in Fig. 7. However, three possible compensations for the [u] vowel are shown with points marked with u_a, u_b, u_c. Note that point u_c in coefficient space requires $q_1 = +5.77$ and corresponds to an $F1-F2$ pair on the "tail" of the mesh.

in the figure, [u_b] and [u_c], were chosen to maintain nearly the same value of $F1$ but have progressively lower values of $F2$. Typical frequencies for $F1$ of an adult male [u] vowel tend to be between about 0.25–0.40 kHz, while $F2$ has been variously reported to range from 0.87 kHz to nearly 1.2 kHz (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995; Lee, Potamianos, and Narayanan, 1999). In SPO95, the measured values of $F1$ and $F2$ in a normal speaking condition ranged from 0.224–0.327 kHz and 0.601–0.768 kHz, respectively. Thus, the $F1-F2$ pairs chosen for the three cases [see the upper-right corner of Fig. 9(b) for numerical values] roughly span a typical range of formant frequencies for a male [u].

From Fig. 9(a), it can be seen that there is only a slight difference in coefficient values for [u_a] and [u_b]. However, increasing q_1 from -2.82 to -2.13 and decreasing q_2 from -3.63 to -4.03 creates enough of a change in the tract shape that $F2$ drops from 1.02 to 0.86 kHz. For the [u_c] compensation, q_1 is increased to a value of $+5.77$, while q_2 is similar to the other two cases. Such a large increase in q_1 is the means by which a vocal-tract shape can be attained that is capable of producing formant frequencies located on the "tail" of Fig. 9(b).

Area functions for each of the three [u] cases were generated with Eq. (1) using the coefficient values specified in Fig. 9(a). They are shown in Fig. 10 in pseudo-midsagittal form and superimposed with the original [u] shape. In addition, a frequency response plot is presented for each case. Relative to the original [u], both [u_a] and [u_b] [Figs. 10(a) and (c)] result in vocal-tract shapes with a small posterior shift in the location of the midtract constriction (i.e., from L_c to L_c') as well as a decrease in its cross-sectional area (at L_c'). Both also have an expanded front cavity and slightly constricted upper pharynx, although the changes are a bit more extensive in the case of [u_b]. The frequency response plots in Figs. 10(b) and (d) reiterate the information already presented in Fig. 9(b) for $F1$ and $F2$, but also show that the upper formants, $F3$ and $F4$, are slightly higher in the compensated conditions than the original vowel.

The vocal-tract shape produced in the [u_c] condition [Fig. 10(c)] is radically different than either the original [u] or the [u_a] and [u_b] conditions. Its prominent features are a widely expanded oral cavity and severely constricted phar-

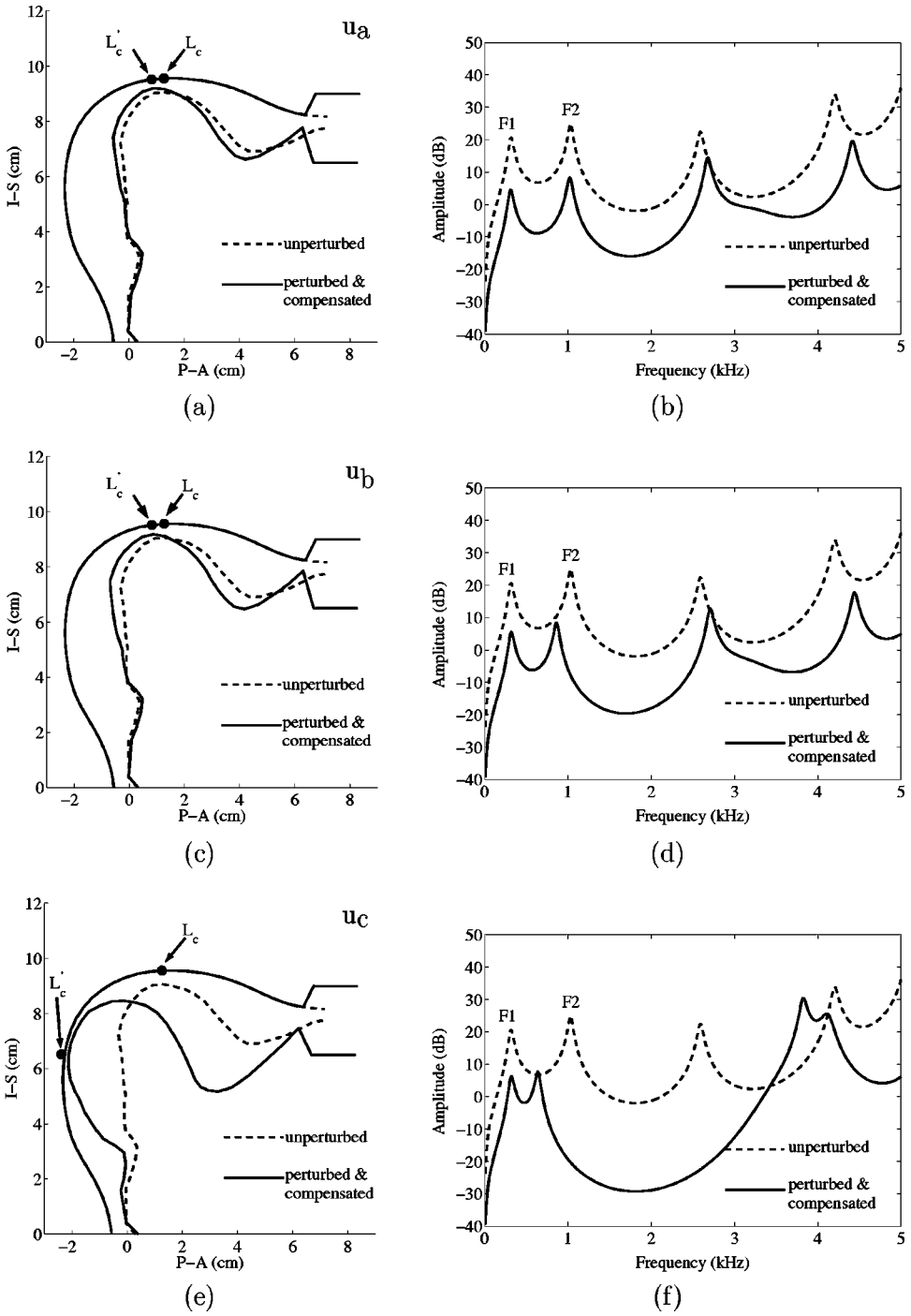


FIG. 10. Comparison of original and perturbed+compensated vocal-tract shapes for [u_a], [u_b], and [u_c] shown as pseudo-midsagittal plots. The perturbed+compensated shapes were generated from the points indicated in Fig. 9(a). Also shown are the corresponding frequency response functions. (a) Original [u] (dashed line) with [u_a] (solid line); (b) frequency responses corresponding to the two shapes in (a). (c) Original [u] (dashed line) with [u_b]; (d) frequency responses corresponding to the two shapes in (c). (e) Original [u] (dashed line) with [u_c]; (f) frequency responses corresponding to the two shapes in (e).

ynx. In addition, the primary constriction location has been moved from the posterior portion of the oral cavity (L_c), to a point well into the upper part of the pharynx (L_c'). It is also noted that the cross-sectional area at L_c' reaches a minimum value of about 0.02 cm^2, which is unrealistically small for a typical vowel (although see Story *et al.*, 1998, p. 475 for a discussion of small measured cross-sectional areas) and would likely give rise to turbulent sound generation (Stevens, 1998). This may represent a point at which the model has been driven beyond reasonable limits and would need to be *adapted* (i.e., modify the model itself) to accept any further expansions in the coefficient space. However, as will be seen in the next section, the [u_c] case still seems to emulate an extreme compensatory response produced by one of the speakers in SPO95.

The frequency responses of [u_c] and the original [u] are shown in Fig. 10(f). The $F1$ and $F2$ for [u_c] are located at 0.32 and 0.64 kHz, respectively, as has already been indicated in Fig. 9(b). While this compensation significantly lowers $F2$ relative to the other cases, $F3$ is observed to increase by more that 1 kHz. This drastic increase in the third formant is an apparent ''cost'' of choosing a coefficient pair that corresponds to the ''tail'' in the $F1-F2$ space.

In Fig. 11, $F3$ is added as a third dimension to the formant mesh [i.e., Fig. 9(b)]. The first plot [Fig. 11(a)] is a perspective view showing all three $F1$, $F2$, and $F3$ dimen-

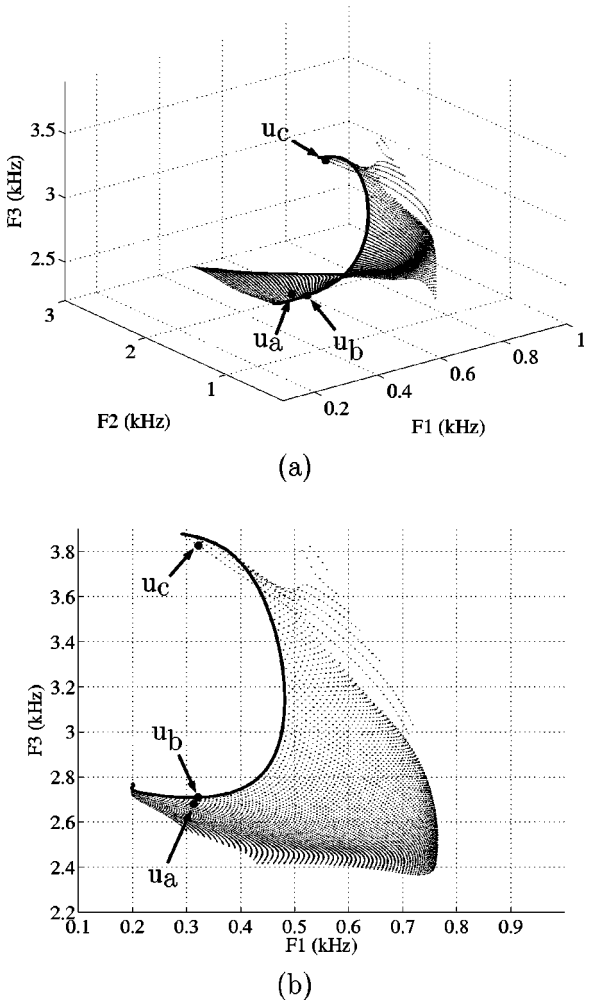


FIG. 11. The same formant vowel space as previously shown in Fig. 9(b), but including the third formant ($F3$) as an additional dimension. In this view, $F3$ is observed to significantly increase whenever an $F1-F2$ pair is chosen on the tail of the mesh [see Fig. 10(b)]. (a) Perspective view with $F1$, $F2$, and $F3$; (b) $F3$ vs $F1$.

sions, whereas the second [Fig. 11(b)] is a rotated view indicating only $F3$ versus $F1$. In either case, a large increase in $F3$ is assured for any point on the tail of the mesh. Hence, within the constraints of the model defined by Eq. (1) and the lip-tube perturbation, an $F2$ value less than about 0.85 kHz can only be attained by allowing $F3$ to move upward in frequency.

C. Comparison to the results of Savariaux *et al.* (1995)

Midsagittal tracings of the vocal tract were shown for several speakers in SPO95, where each one was representative of a different degree of compensation for the lip tube. Two sets of these tracings are reproduced in Fig. 12, where the normal production of the [u] vowel is indicated with a dashed line and the perturbed versions are shown with solid lines [note that these are from the last repetition in their experiment (condition PL)]. For both cases, the midtract constriction in the perturbed condition is located at a point L_c' that is posterior and inferior to the constriction location L_c in the normal condition. Additionally, the oral cavity is widened

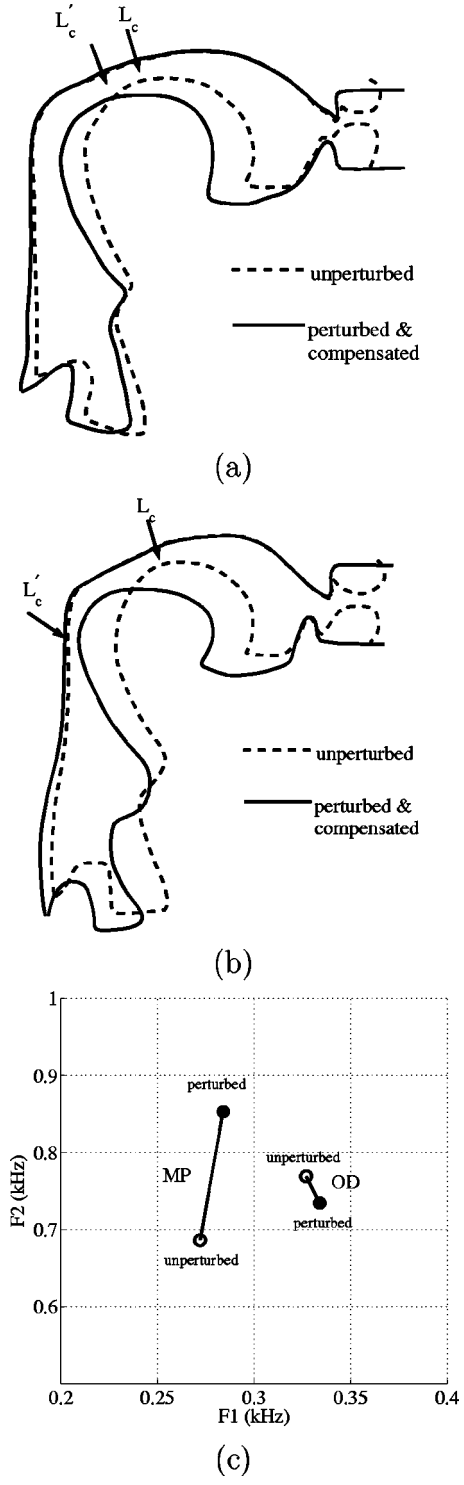


FIG. 12. Adaptation of midsagittal tracings reported in SPO95, where the vocal-tract shape under normal production of the [u] vowel is indicated by dashed lines and the perturbed condition (denoted PL in their study) by solid lines. (a) Subject MP; (b) Subject OD; (c) Formant pairs plotted in the $F2$ vs $F1$ plane for each subject's production of the normal (open circle) and perturbed/adapted (solid circle) [u].

in the perturbed condition. However, the speaker in Fig. 12(a) (SPO95's subject MP) did not move the constriction as far back or reduce its cross-sectional area to the same degree as did the speaker in Fig. 10(b) (their subject OD). As can be seen in Fig. 10(c), MP's adaptation was only partially successful in keeping $F1$ and $F2$ at frequency values similar to

the normal [u] vowel, since $F2$ was still nearly 170 Hz too high. OD's adaptation was extensive enough to essentially produce a full compensation for the lip tube in terms of $F1$ and $F2$; however, $F3$ (not shown) was about 350 Hz higher than in the normal condition (SPO95, p. 2432).

These results for these real speakers mirror those presented in the previous section that were produced by the area function model. Specifically, the compensatory tract shape changes in Figs. 10(a) and (c) may be considered somewhat analogous to the speaker MP. The places of constriction are similar in both the real and modeled cases. However, the degree of constriction produced by the speakers appears to be less severe than for the modeled versions, perhaps explaining why MP's $F1$ and $F2$ were only partially compensated.

The more extensive vocal-tract modification in Fig. 10(e) suggests similarities to SPO95's speaker OD because both have tight constrictions, almost excessively widened oral cavities, and fully compensated $F1$ and $F2$ frequencies. Furthermore, $F3$ for both the modeled compensation and speaker OD are higher than in a normal [u] production. Also noteworthy is that during the collection of audio data in the SPO95 experiment, the speakers were allowed 21 trials of [u] production following the placement of the lip tube. During this process speaker OD maintained a nearly constant $F1$ while $F3$ was allowed to be consistently higher than in the normal condition. But, unlike any of the other speakers, $F2$ oscillated above and *below* the normal value throughout the trials (SPO95, p. 2440); the final value of $F2$ was slightly below the normal $F2$. This seemingly inconsistent behavior would perhaps be expected if a speaker's $F1-F2$ vowel space under perturbed conditions was similar to Figs. 9(b) and 11. That is, the variation in $F2$ over the course of many trials may be due to the speaker choosing between a location on the tail portion (e.g., $[u_c]$) or some point above it in the main body of the mesh (e.g., $[u_b]$). Moving back and forth between these two points would create drastically different vocal-tract shapes, forcing $F2$ to change in a stepwise manner, while $F1$ could potentially be left nearly unchanged (as OD seemed to do). However, according to Fig. 11, $F3$ should also oscillate as the tract shape shifts from $[u_b]$ to $[u_c]$. This behavior is not apparent in the SPO95 data.

IV. DISCUSSION

The results of this study have shown that an area function model of the vocal tract based on two orthogonal shaping components (modes) can compensate for a perturbation to maintain a desired pattern of formant frequencies, at least in terms of $F1$ and $F2$. Furthermore, the shapes of the vocal tract produced by the model in the perturbed/compensated conditions are representative of previously reported tract shapes of real speakers under similar conditions. What seems to be significant is that the model itself did not require any modification, only the polarity and magnitude of the weighting coefficients for the modes needed to be changed in order to carry out the compensation. However, because of the small cross-sectional areas (≈ 0.02 cm^2) state produced in the constriction for the compensated condition, it could be argued that the model was driven beyond reasonable limits. Perhaps so, but to keep this study relatively simple, it is noted that the model was constrained to maintain the same tract length for all of the perturbed conditions. Thus, the compensation of $F1$ and $F2$ had to originate entirely from modifications of cross-sectional areas. In contrast, the midsagittal tracings in SPO95 suggest that localized length changes may have occurred near the laryngeal end of the vocal tract as the speakers attempted to compensate for the lip tube. Hence, some of the compensation may have been provided by lengthening the vocal tract, easing the necessary severity of the constriction area. While a more accurate simulation would include such effects, the similarities of the results in the present study and those from SPO95 nonetheless suggest that the modes capture at least some aspects of vocal-tract shaping that carry beyond normal speech.

To the degree that these orthogonal modes capture basic movement patterns of the *tongue*, the results are similar to several studies of lingual articulation for vowels. An analysis of electromyographic signals recorded during vowel production (Maeda and Honda, 1994; Honda, 1996) showed that the action of two antagonistic muscle pairs, hyoglossus (HG)–genioglossus posterior (GGp) and styloglossus (SG)–genioglossus anterior (GGa) coincided with tongue-shaping characteristics observed in midsagittal x-ray images that appear to be analogous to the modes in the present study. The HG–GGp pair was shown to be responsible for a tongue position that is either front/high or back/low, similar to the effect of the first mode on the vocal-tract shape and corresponding formant frequencies (see Fig. 3). The back/high or front/low tongue positions were generated by the SG–GGa pair, analogous to the second mode.

However, the modes on which the current model is based were derived from an analysis performed on area functions measured directly from 3D images of the *entire* vocal tract rather than midsagittal tongue profiles. Thus, the spatial characteristics accounted for by these two modes suggest that they may represent some type of coordinative structure or synergy for which constraints between muscles are established, allowing for motor control in terms of functional units rather than individual muscles (e.g., Kelso, 1995; Löfqvist, 1997). Hence, if each mode is considered to be a functional unit of learned muscle synergies that efficiently and uniquely affects the acoustic properties of the vocal tract, then perhaps it follows that they could be more or less ''activated'' to alter the tract shape to meet some acoustic need. Under normal speaking conditions, the levels of activation would be tightly linked to their acoustic consequences via a mapping like the one shown in Fig. 4. Under conditions of vocal-tract perturbation, a new mapping is established that links exaggerated activation patterns to acoustic consequences that are as similar as possible to those in the normal condition. But, as suggested above, the structure of the functional units (modes) need not change, just the level of activation. Of course, if the perturbation is excessive, the functional units (i.e., the model for the present study) may need to be modified as well.

V. CONCLUSION

The main question posed in this study was to determine whether the area function model specified by Eq. (1) could be made to compensate for a lip-tube perturbation by modifying only the control parameters of the model and not its underlying structure. The results suggest that, indeed, the model can compensate by changing only the weighting coefficients of each mode, albeit to values that are well beyond those in the normal conditions. The compensated vocal-tract shapes were also shown to be similar to those produced by real speakers (SPO95), suggesting that the model is potentially predictive of humans' abilities to configure the vocal-tract shape under both ordinary and extraordinary conditions.

ACKNOWLEDGMENTS

The author would like to thank Pascal Perrier and an anonymous reviewer for their helpful comments on an earlier version of this manuscript. A preliminary version of this paper was presented at the 140th Meeting of the Acoustical Society of America. This work was supported by NIH R01-DC04789.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). ''A theoretical investigation of reference frames for the planning of speech movements,'' Psychol. Rev. **105**(4), 611–633.

Hamlet, S., and Stone, M. (**1976**). ''Compensatory vowel characteristics resulting from the presence of different types of experimental prostheses,'' J. Phonetics **4**, 199–218.

Hamlet, S., and Stone, M. (**1978**). ''Compensatory alveolar consonant production induced by wearing a dental prosthesis,'' J. Phonetics **6**, 227–248.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). ''Factor analysis of tongue shapes,'' J. Acoust. Soc. Am. **62**(3), 693–707.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**(5), 3099–3111.

Honda, K. (**1996**). ''Organization of tongue articulation for vowels,'' J. Phonetics **24**, 39–52.

Kelso, J. A. S. (**1995**). *Dynamic Patterns: The Self-Organization of Brain and Behavior* (MIT Press, Cambridge, MA).

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). ''Acoustics of children's speech: Developmental changes of spectral and temporal parameters,'' J. Acoust. Soc. Am. **105**(3), 1455–1468.

Lindblom, B., Lubker, J., and Gay, T. (**1979**). ''Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation,'' J. Phonetics **7**, 147–161.

Löfqvist, A. (**1997**). ''Theories and models of speech production,'' in *The Handbook of Phonetic Sciences*, edited by W. Hardcastle and J. Laver (Blackwell, Oxford, UK).

Maeda, S. (**1990**). ''Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model,'' in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht, pp. 131–149).

Maeda, S., and Honda, K. (**1994**). ''From EMG to formant patterns of vowels: The implication of vowel spaces,'' Phonetica **51**, 17–29.

Meyer, P., Wilhelms, R., and Strube, H. W. (**1989**). ''A quasiarticulatory speech synthesizer for German language running in real timem,'' J. Acoust. Soc. Am. **86**(2), 523–539.

Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**(2), 175–184.

Savariaux, C., Perrier, P., and Orliaguet, J. P. (**1995**). ''Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production,'' J. Acoust. Soc. Am. **98**, 2428–2242.

Sondhi, M. M., and Schroeter, J. (**1987**). ''A hybrid time-frequency domain articulatory speech synthesizer,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**(7), 955–967.

Sorokin, V., Olshansky, V., and Kozhanov, L. (**1998**). ''Internal model in articulatory control: Evidence from speaking without larynx,'' Speech Commun. **25**, 249–268.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). ''Vocal tract area functions from magnetic resonance imaging,'' J. Acoust. Soc. Am. **100**(1), 537–554.

Story, B. H., and Titze, I. R. (**1998**). ''Parametrization of vocal tract area functions by empirical orthogonal modes,'' J. Phonetics **26**, 223–260.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). ''Vocal tract area functions for an adult female speaker based on volumetric imaging,'' J. Acoust. Soc. Am., **104**(1), 471–487

Story, B. H., Titze, I. R., and Hoffman, E. A. (**2001**). ''The relationship of vocal tract shape to three voice qualities,'' J. Acoust. Soc. Am. **109**, 1651–1667.

Story, B. H., and Titze, I. R. (**2002**). ''A preliminary study of voice quality transformation based on modifications to the neutral vocal tract area function,'' J. Phonetics **30**, 485–509.

Yehia, H. C., Takeda, K., and Itakura, F. (**1996**). ''An acoustically oriented vocal-tract model,'' IEICE Trans. Inf. Syst. **E79-D**(8), 1198–1208.

An approach to real-time magnetic resonance imaging for speech production[a)]

Shrikanth Narayanan
Departments of Electrical Engineering, Computer Science, and Linguistics, University of Southern California, Los Angeles, California 90089

Krishna Nayak[b)]
Department of Electrical Engineering, Stanford University, Stanford, California 94305

Sungbok Lee
Departments of Electrical Engineering and Linguistics, University of Southern California, Los Angeles, California 90089

Abhinav Sethy
Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089

Dani Byrd[c)]
Department of Linguistics, University of Southern California, Los Angeles, California 90089

(Received 29 July 2003; revised 2 January 2004; accepted 13 January 2004)

Magnetic resonance imaging (MRI) has served as a valuable tool for studying static postures in speech production. Now, recent improvements in temporal resolution are making it possible to examine the dynamics of vocal-tract shaping during fluent speech using MRI. The present study uses spiral k-space acquisitions with a low flip-angle gradient echo pulse sequence on a conventional GE Signa 1.5-T CV/i scanner. This strategy allows for acquisition rates of 8–9 images per second and reconstruction rates of 20–24 images per second, making veridical movies of speech production now possible. Segmental durations, positions, and interarticulator timing can all be quantitatively evaluated. Data show clear real-time movements of the lips, tongue, and velum. Sample movies and data analysis strategies are presented. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1652588]

PACS numbers: 43.70.Jt [AL] Pages: 1771–1776

I. INTRODUCTION

A perennial challenge in speech production research is the ability to examine real-time changes in the shaping of the vocal tract. These deformations in the vocal tract serve two critical functions in speech production: (1) Creation of supraglottal sources through the coordination of appropriate articulator configurations and aerodynamic conditions, and (2) acoustic filtering of the laryngeal and supraglottal sources by actively modifying the shape of the vocal tract to produce distinctive spectral patterns for the different speech sounds. Hence, spatiotemporal information about speech movements is critical not only to understanding and modeling the speech production process but also to a thorough understanding of speech acoustics. Many approaches are available to the researcher for obtaining information about the rapid and complex movements of the mouth and face that participate in creating speech sounds—e.g., electropalatography (EPG) to examine linguapalatal contact, point-movement tracking (e.g., x-ray microbeam, magnetometry) for dynamic oral information, or ultrasound for examining tongue-surface contours in the mouth and pharynx. However, none of these techniques yields real-time moving images of articulators along the entire length and diameter of the vocal tract.[1] We report for the first time a novel high speed MRI technique for imaging the moving vocal tract in real time.

A. Prior work on dynamic imaging

Cine x-ray techniques had been popular in speech research to obtain lateral (midsagittal) images of the vocal tract (Fant, 1960; Delattre and Freeman, 1968; Perkell, 1969; Subtelny *et al.*, 1972; Giles and Moll, 1975). Cross-sectional vocal-tract areas (area functions) were estimated from these midsagittal images for acoustic modeling. The use of such radiographic techniques has declined significantly, mainly due to the radiation risks involved. Computer-aided tomography (Kiritani *et al.*, 1977) is capable of yielding cross-sectional information but still suffers from radiation risks and relatively low speed of imaging. Ultrasound and MRI have proved to be viable alternatives for such investigations.

Ultrasound provides an acceptable means of studying tongue shapes and movements during speech production (Stone, 1990). The technology is safe and noninvasive, and the imaging speed is suitable for studying the dynamics of speech production. However, the entire vocal tract cannot be studied at once by this method. Moreover, due to the presence of the airway above the tongue, the palate cannot be imaged, while due to the presence of air space below, the tongue tip/blade cannot be successfully captured. Neverthe-

[a)]Portions of this work were presented at the 2003 Nashville meeting of the Acoustical Society of America [Narayanan *et al.*, J. Acoust. Soc. Am. **113**(4,2), 2258 (2003)].
[b)]Now affiliated with the USC Department of Electrical Engineering.
[c)]Electronic mail: dbyrd@usc.edu.

less, modified ultrasound techniques have helped to further the understanding of the 3D model of tongue, either by using multiple scanning procedures (Watkin and Rubin, 1989) or by using other parallel instrumentation measurements such as x-ray microbeam (Stone, 1990) or palatography (Stone *et al.*, 1992).

Magnetic resonance imaging (MRI) is a powerful tool for obtaining vocal-tract geometry data and does not involve any known radiation risks. The images have good signal-to-noise ratio, are amenable to computerized 3D modeling, and provide excellent structural differentiation. In addition, the tract (airway) area and volume can be directly calculated. The low image-sampling rate, however, has largely limited MRI to the study of sustained speech sounds, corresponding to ''static'' tract shape (Baer *et al.*, 1991; Greenwood *et al.*, 1992; Moore, 1992; Sulter *et al.*, 1992; Dang *et al.*, 1993; Narayanan *et al.*, 1995; Story, 1995; Narayanan *et al.*, 1997; Alwan *et al.*, 1997; Ong and Stone, 1998, Narayanan *et al.*, 1999). Some of these studies used multiple repetitions of an utterance to reconstruct the 3D volume of the vocal tract images from static postures. These data have been valuable in providing hitherto unknown details of the 3D vocal-tract morphology and interspeaker variation during speech production. MRI data have also been used to improve models of the acoustics of speech sounds including vowels (Story *et al.*, 1996), liquids (Bangayan *et al.*, 1996; Espy-Wilson *et al.*, 2000), and fricatives (Narayanan and Alwan, 2000).

In the past few years, improvements in temporal resolution have allowed MRI to move from being limited to imaging static postures (Mády *et al.*, 2001, 2002; Demolin *et al.*, 2000). Progress toward increased temporal resolution in MR imaging is challenged by issues related to poor signal-to-noise ratio (SNR) and susceptibility artifacts when fast imaging protocols, such as fast gradient echo techniques and echo planar imaging, are involved. At this point, it is worthwhile to clarify the use of the terms *dynamic* MRI and *real-time* MRI in the present article. We use the term *dynamic* to refer to the creation of images from an actively articulating subject, rather than a static postural source. We reserve the term *real-time* MRI to refer specifically to directly capturing or acquiring moving image data in real time. That is, *dynamic* refers to the source, *real time* to its acquisition.

The dynamic MRI technique (Foldvik *et al.*, 1993) provides a way to capture valuable kinematic information. It relies on gated scanning on numerous repetitions of the same speech sequence to reconstruct the impression of articulatory movement in time. Note that the reconstructed sequence is drawn from across several repetitions. Using a multiplanar dynamic MRI technique, Shadle *et al.* (1999) formed pseudo-time-varying images of the vocal tract using a simultaneously recorded audio signal and the scanned images acquired while an utterance is repeated. The technique showed positive results in terms of getting fairly accurate volumetric measurements of area functions and tongue volumes.

In the challenging area of real-time MRI in which motion is imaged directly, Demolin *et al.* (2000) have shown significant improvement towards real-time MRI for speech production by the use of an ultrafast implementation of turbo spin echo sequence (TSE) with an acquisition speed of about 4 images/second for a single plane. This initial work on real-time MRI and work by Mády *et al.* (2002) was promising and worth further investigation.

While these MRI advances represent a significant improvement in the quality of information that could be attainted about changes in tongue, lip, and velum positioning over time, they are still not close to the temporal resolution necessary for capturing the dynamic characteristics of tongue movement, which demands a *minimal* sampling rate on the order of 20 Hz. Such high MR imaging rates present significant technical challenges. We present one approach to dynamic real-time MR imaging that successfully addresses these challenges.

II. TECHNICAL DESCRIPTION

Real-time MRI has been developed for several applications including cardiac imaging, abdominal imaging, and interventional imaging, which also demands a high temporal resolution for tracking a moving structure (Kerr *et al.*, 1995; 1997; Santos *et al.*, 2002; Nayak *et al.*, 2001). We pursue real-time imaging of the upper airway, which for speech also requires high spatial and temporal resolution and resilience to image artifacts. We utilized gradient echo imaging with a fast interleaved spiral acquisition strategy. On a conventional 1.5-T imager with high-speed gradients, we were able to acquire images with 110-ms temporal resolution. Critically, this allows effective reconstruction rates of 24 frames per second using a sliding window technique.

A. Experimental methods

Experiments were performed on a GE Signa 1.5-T system with gradients supporting 40-mT/m amplitude and 150-T/m/s slew rate and receiver capable of 4-μs sampling ($\pm$125-kHz receiver bandwidth). A generic head coil, not specially adapted for vocal-tract imaging, was used in all studies. The institutional review board of Standford University approved the imaging protocols, and each subject was screened for magnetic resonance imaging risk factors and provided informed consent in accordance with institutional policy.

Sequences were implemented within a custom real-time imaging (RTI) framework previously described (Kerr *et al.*, 1995; Santos *et al.*, 2002). Interactive control, continuous reconstruction, and display of images were performed on a workstation adjacent to the scanner.

B. Pulse sequence design

Previous approaches to imaging the upper airway in the vocal tract have been limited to standard 2DFT (two-dimensional Fourier transform) and 3DFT sequences (Shellock *et al.*, 1992). These techniques produce good quality static images but are not efficient enough to capture dynamic events. For rapid imaging of the upper airway (UA), we focused on using the spiral readout scheme for accelerating acquisition. Spirals are an alternative scheme for sampling the spatial frequency domain (k-space), in which data are acquired in a spiraling pattern. Twenty interleaved spirals together form a single image. Spirals are highly time effi-

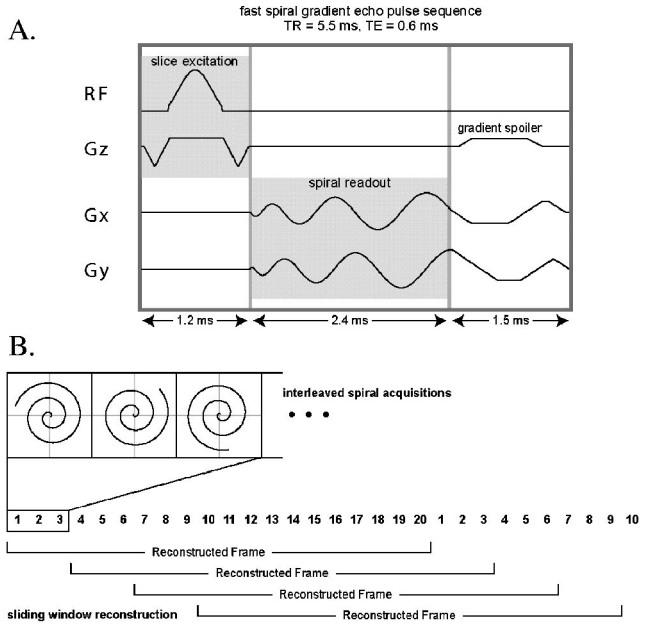


FIG. 1. (A) Pulse sequence used in real-time upper airway imaging. A 1.4-ms slice selective excitation (640-μs rf) is followed by a 2.4-ms spiral readout, and gradient refocusing in X and Y, and crushing in Z. The TR is 5.5 ms. Twenty interleaves are used to achieve 112 pixels over a nominal 20-cm FOV (30-cm FOV was used in most studies). Complete images are acquired every 110 ms, and were reconstructed at 24 frames/s using a sliding window. (B) Acquisition timing. In spiral imaging, the frequency domain is sampled using spiral-shaped trajectories (top). Twenty interleaved spirals are required to form a single image. In dynamic real-time imaging, we continuously reacquire these 20 interleaves (middle). Each image is based on 110 ms of acquired data; however, using sliding window reconstruction, images can be reconstructed at a higher rate as new interleaves become available (bottom). Note that motion within the 110-ms window will result in motion artifacts and/or blurring.

cient (Meyer *et al.*, 1992; Nayak *et al.*, 2001) and have excellent motion properties (Nishimura *et al.*, 1995), as compared to the widely used 2DFT Cartesian approach. Their main limitation is blurring due to off-resonance. This is a particular problem in the UA because the difference in magnetic susceptibility between air and tissue (Schenk, 1996) causes large amounts of off-resonance near to air–tissue interfaces. The amount of blurring experienced in spiral imaging is proportional to the amount of phase that accrues during each readout. To mitigate this effect, we used extremely short 2.4-ms readouts.

The pulse sequence is shown in Fig. 1(A). A 640-μs excitation is followed by 2.4-ms readouts and a gradient spoiler. The total TR is 5.5 ms. Twenty interleaves are required to achieve 1.8-mm resolution over a nominal 20-cm FOV (field of view). A slice thickness of 5 mm was used. (It should be noted that a 3-mm slice thickness would be possible with the same pulse sequence but with a sacrifice in SNR.) In our studies, a 30-cm FOV was used to prevent aliasing artifacts from the large volume captured by the head coil. This resulted in an effective in-plane resolution of 2.7 mm. Complete images were acquired every 110 ms, and were reconstructed at 24 frames per second using a sliding window (Holsinger *et al.*, 1990). A schematic representation of this procedure is given in Fig. 1(B). Basically, this means that while each image is acquired over 110 ms, since data for each image are acquired in pieces, images can be reconstructed more often as portions of the frequency-domain data are updated.

FIG. 3. The [l] constriction in ''Say peal leap again;'' the real-time MRI movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

III. EXAMPLE STUDIES

Two subjects (SN and KN) were recorded in a preliminary study producing normal speech using this protocol. The pulse sequence used is that given in Fig. 1. Study 1, in which both subjects participated, included English sentences varying the syllable position of /n, r, l/. Study 2, in which only SN participated, included sentences in Tamil varying among five liquids. This dataset matches that of a point-movement tracking study found in Narayanan *et al.* (1999).

Results show clear imaging of the entire vocal tract and real-time movements of the lips, tongue, and velum. Seg-

FIG. 2. The [l] constriction in ''Say pea leap again;'' the real-time MRI movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

FIG. 4. The [n] constriction in ''Say bean knee again;'' the real-time MRI movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

FIG. 5. The Tamil retroflex [ɭ] constriction from ''...paɭam;'' the real-time MRI movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

mental durations, positions, and interarticulator timing can all be quantitatively evaluated. Sample movies and data analysis strategies are presented below.

A. English sonorants

The sentences used in Study 1 include
''Say——again.''
pea leap, peal heap, peal leap, peal-y, pea reap, pier heap, pier reap, peer-y, be knee, bean he, bean knee, bean-y

Captured frames from movies of the moving vocal tract for three sentences are presented (see Figs. 2–4) and files of the entire real-time movie for each can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

B. Tamil liquids

The sentences used in Study 2 to examine liquid consonants in Tamil embed stimuli of the sort ''pali'' [sacrifice], ''paɭi'' [nonce], ''paɻi'' [blame], ''pari'' [horse], and ''paṟi'' [pluck] in a frame utterance; the full set of sentences and a phonetic description can be found in Narayanan *et al.* (1999). Captured frames from movies of the moving vocal tract for the first three consonants given above are presented in Figs. 5–7, and files of the entire real-time movie for each can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

It is noteworthy that Movie 5 indicates the temporal as well as spatial extent of the retroflex constriction and release. In Fig. 8, a selection from this movie is presented to allow for the inspection of the vocal-tract shape changes over multiple frames simultaneously. It indicates that the tongue tip curling for the retroflex lateral in [paɭam] begins well before the *preceding* labial closure is achieved.

FIG. 6. The Tamil [l] constriction from ''... palam;'' the real-time MRI movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

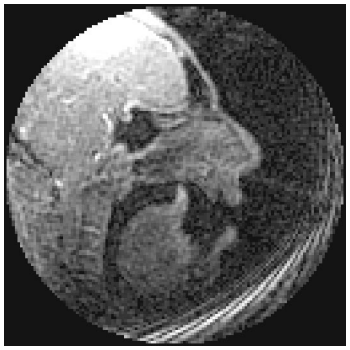


FIG. 7. The Tamil [ɻ] constriction from ''...paɻam;'' the real-time MRI movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

IV. SKETCH OF ANALYSIS CHALLENGES

The method presented above promises to generate vast amounts of data that need to be processed efficiently and effectively. This opens up a number of data analysis challenges.

A. Data segmentation and validation

We hope to replace the current laborious process of hand-segmenting the image data to identify areas of linguistic interest by an automated, *data-driven* process. A first goal is to enable automatic methods for extracting relevant regions of interest from the image sequences. Our ongoing work focuses on methods to automatically segment and track the real-time MRI data using Kalman snakes and optical flow (Kass *et al.*, 1987; Cootes *et al.*, 1994; Gautama and van Hulle, 2002). At this time, this has allowed us to quantitatively track events during articulation; for example, the open-

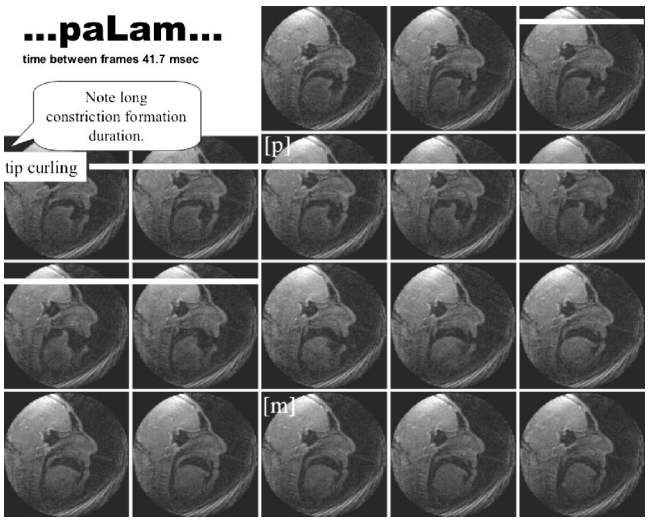


FIG. 8. An example sequence for ''...paɭam...'' with 41.7 ms between frames. The middle panel in the top row shows the beginning of tongue tip curling toward forming the retroflex lateral; the maximal tongue retroflexion can be seen in the first panel of row 3. Note the formation of the word-initial and word-final labials are also marked in the middle frames of rows 2 and 4, respectively.

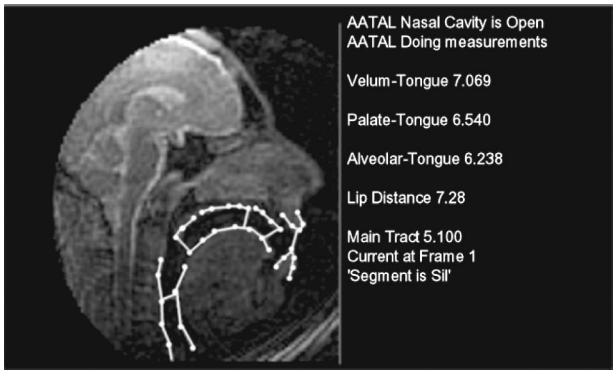


FIG. 9. A snapshot of the automatic tracking method of MR images developed in the study. In the left panel, each group of points connected by line represents a Kalman snake (i.e., contour), which is individually updated through an optical flow method applied to the sequence of images. In this demonstration the midsagittal dimensions (represented by lines connecting the upper and lower walls of the vocal tract) at some positions along the vocal tract are computed and corresponding values are displayed in the right panel. (A) The real-time MRI movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004.

ing and closing of the velic port was detected based on a edge detection and linking method. We present a preliminary movie (frame captured in Fig. 9) that shows automatic airway tracking and measurements (the movie can be downloaded at sail.usc.edu/production/rtmri/jasa2004). Future work will focus on issues related to calibration and validation of such automatic processing schemes. Once validated, these data can be valuable for modeling, such as in performing degree-of-freedom analysis and motion parametization of articulators using PCA applied to the point-tracking data generated by image analysis.

B. Combining data types

Parallel point-tracking (EMMA magnetometer; see Perkell *et al.*, 1992) data collected previously for the same Tamil stimuli and speaker (Narayanan *et al.*, 1999) provide a future opportunity to investigate data alignment strategies for MRI and point-tracking technologies. For example, key region tracking from MRI can be supplemented and/or validated by using magnetometer data. In the future, imagable beads can help coregistration for validation purposes (Byrd *et al.*, 1999). In addition to improving the overall time resolution of the combined data, point-tracking data also provide us with dynamical information in terms of velocity and acceleration for different points on active articulators.

C. New avenues

The ability to acquire real-time images of the speaking vocal tract can provide vital data for computational modeling of the production process that are not currently otherwise available. Intergestural coordination can be evaluated—for example, an understanding of the production of English [r] can be enhanced by the fact that the pharyngeal, oral, and labial components of the consonant can be simultaneously quantified (Alwan *et al.*, 1997), not only for the quality of the constrictions but also for the coordination among the constrictions (note that point-tracking technology, in contrast, cannot be employed in the pharynx). Rhotic mechanisms can be examined cross-linguistically, including rare or poorly described rhotics such as the uvular.

We also feel that this real-time MRI tool can provide heretofore unavailable details that will inform our understanding of interspeaker variability, since individual differences in the morphology of the vocal tract can be directly associated with differences in speech production behavior.

D. Technical improvements on the horizon

We have identified and are pursuing a number of avenues to further improve both the speed and quality of the image acquisition. First, our team is in the process of designing a new neck coil for better SNR for airway region. Second, we are exploring options for the acquisition of (even compromised) audio signals during imaging (this has been done fairly successfully in prior static vocal-tract imaging studies). Finally, a significant effort is being devoted to new pulse sequence designs. In spiral imaging (such as the one reported here), off-resonance causes a blurring of signal (Noll *et al.*, 1991) predominantly at the air–tissue interface. When more than $\pi/4$ of phase is accrued during a readout this blurring is significant (Noll *et al.*, 1992), and this small pervasive blurring can be observed throughout our image sequences, for example, at the tongue surface. Another fast alternative is echo-planar imaging, where off-resonance results in a shifting of signal along the phase encode direction. The amount of shift in pixels is roughly equal to the amount of phase accrued over a readout divided by 2π. In areas of high off-resonance, echo-planar can result in a warping artifact but will retain image sharpness. An additional benefit of echo-planar is that the FOV can be limited in the readout direction using the analog filter, enabling scanning with a smaller FOV. Both of these effects may enable sharper and higher resolution real-time imaging of the upper airway. While preliminary experiments have shown improved spatial resolution (1.56 mm) and improved sharpness, further investigation is needed to establish the performance of echo-planar in the presence of rapid motion with speech.

V. CONCLUSION

Using a new approach, real-time images of the moving vocal tract with MRI are, for the first time, possible at rates that permit meaningful investigations of speech production dynamics ($\geqslant$20 images per second). Future challenges for this technology include improvements in image quality and in quantitative evaluation of continuously varying vocal-tract shaping.

ACKNOWLEDGMENTS

This work was supported by NIH Grant DC 03172. The authors extend their thanks to Phil Hoole and an anonymous reviewer for their helpful comments.

[1]Cinefluorography does permit this but is not useful due to the need to protect human subjects.

Alwan, A., Narayanan, S., and Haker, K. (**1997**). "Toward articulatory-

acoustic models for liquid consonants based on MRI and EPG data. II. The rhotics,'' J. Acoust. Soc. Am. **101**, 1078–1089.
Baer, T., Gore, J., Gracco, C., and Nye, R. (**1991**). ''Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels,'' J. Acoust. Soc. Am. **90**, 799–828.
Bangayan, P., Alwan, A., and Narayanan, S. (**1996**). ''From MRI and acoustic data to articulatory synthesis: A case study of the lateral approximants in American English,'' in Proceedings of the International Conference. Spoken Language Processing (Philadelphia, PA), pp. 793–796.
Byrd, D., Browman, C. P., Goldstein, L., and Honorof, D. (**1999**). ''Magnetometer and x-ray microbeam comparison,'' in *Proceedings of the 14th International Congress of Phonetic Sciences*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (American Institute of Physics, New York), pp. 627–630.
Cootes, T. F., Hill, A., Taylor, C. J., and Haslam, J. (**1994**). ''The use of active shape models for locating structures in medical images,'' Image Vis. Comput. **12**, No. 6, 355–366.
Dang, J., Honda, K., and Suzuki, H. (**1993**). ''MRI measurements and acoustic investigation of the nasal and paranasal cavities.'' J. Acoust. Soc. Am. **94**, No. 3, 1765.
Delattre, P., and Freeman, D. C. (**1968**). ''A dialect study of American r's by x-ray motion picture,'' Linguistics, An International Review, Vol. 44, pp. 29–68.
Demolin, D., Metens, T., and Soquet, A. (**2000**). ''Real time MRI and articulatory coordinations in vowels speech production,'' Proceedings of the 5th Speech Production Seminar, pp. 86–93.
Espy-Wilson, C. Y., Boyce, S. E., Jackson, M. T. T., Narayanan, S., and Alwan, A. (**2000**). ''Acoustic modeling of the American English /r/,'' J. Acoust. Soc. Am. **108**, No. 1, 343–356.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton and Co., The Hague).
Foldvik, A., Kristainsen, U., and Kvaerness, J. (**1993**). ''A time-evolving three-dimensional vocal tract model by means of MRI,'' Proceedings of Eurospeech, pp. 557–560.
Gautama, T., and van Hulle, M. M. (**2002**). ''A phase-based approach to the estimation of the optical flow field using spatial filtering,'' IEEE Trans. Neural Netw. **13**(5), 1127–1136.
Giles, S. B., and Moll, K. L. (**1975**). ''Cineflurographic study of selected allophones of English /1/,'' Phonetica **31**, 206–227.
Greenwood, A. R., Goodyear, C. C., and Martin, P. A. (**1992**). ''Measurement of vocal tract shapes using magnetic resonance imaging,'' IEE Proc. I (Commun., Speech Vision) (UK), Vol. 139, No. 6, pp. 553–560.
Holsinger, A., Wright, R. C., Riederer, S. J., Farzaneh, F., Grimm, R. C., and Maier, J. K. (**1990**). ''Real-time interactive magnetic resonance imaging,'' Magn. Reson. Med. **14**, No. 3, 547–553.
Kass, M., Witkin, A., and Terzopoulos, A. D. (**1987**). ''Snake: Active contour model,'' Int. J. Comput. Vis. **1**, 321–331.
Kerr, A. B., Pauly, J. M., Meyer, C. H., and Nishimura, D. G. (**1995**). ''New strategies in spiral MR fluoroscopy,'' Proceedings of SMR, 3rd Annual Meeting (Nice, FR), p. 99.
Kerr, A. B., Pauly, J. M., Hu, B. S., Li, K. C. P., Hardy, C. J., Meyer, C. H., Macovski, A., and Nishimura, D. G. (**1997**). ''Real-time interactive MRI on a conventional scanner,'' Magn. Reson. Med. **38**, 355–367.
Kiritani, S., Tateno, Y., and Iinuma, T. (**1977**). ''Computer tomography of the vocal tract,'' in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and F. S. Cooper (University of Tokyo press, Tokyo), pp. 203–206.
Mády, K., Sader, R., Zimmermann, A., Hoole, P., Beer, A., Zeilhofer, H., and Hannig, C. (**2001**). ''Use of real-time MRI in assessment of consonant articulation before and after tongue surgery and tongue reconstruction.'' in *Speech Motor Control in Normal and Disorderd Speech*, edited by B. Maassen, W. Hulstijn, R. Kent, H. Peters, and P. van Lieshout, 4th International Speech Motor Conference, Nijmegen, pp. 142–145.
Mády, K., Sader, R., Zimmermann, A., Hoole, P. Beer, A., Zeilhofer, H., and Hannig, C. (**2002**). ''Assessment of consonant articulation in glossectomee speech by dynamic MRI,'' Proceedings of ICSLP, Denver.
Meyer, C. H., Hu, B. S., Nishimura, D. G., and Macovski, A. (**1992**). ''Fast spiral coronary artery imaging,'' Magn. Reson. Med. **28**, 202–213.
Moore, C. A. (**1992**). ''The correspondence of vocal tract images with volumes obtained from magnetic resonance images,'' J. Speech Hear. Res. **35**, 1009–1023.
Narayanan, S., and Alwan, A. (**2000**). ''Noise source models for fricative consonants,'' IEEE Trans. Speech Audio Process. **8**, No. 3, 328–344.
Narayanan, S., Alwan, A., and Haker, K. (**1995**). ''An articulatory study of fricative consonants using magnetic resonance imaging,'' J. Acoust. Soc. Am. **98**, 1325–1347.
Narayanan, S., Alwan, A., and Haker, K. (**1997**). ''Toward articulatory-acoustic models for liquid consonants based on MRI and EPG data. I. The laterals,'' J. Acoust. Soc. Am. **101**, 1064–1077.
Narayanan, S., Byrd, D., and Kaun, A. (**1999**). ''Geometry, kinematics, and acoustics of Tamil liquid consonants,'' J. Acoust. Soc. Am. **106**, 1993–2007.
Nayak, K. S., Pauly, J. M., Yang, P. C., Hu, B. S., Meyer, C. H., and Nishimura, D. G. (**2001**). ''Real-time interactive coronary MRA,'' Magn. Reson. Med. **46**, No. 3, 430–435.
Nishimura, D. G., Irarrazabal, P., and Meyer, C. H. (**1995**). ''A velocity k-space analysis of flow effects in echo-planar and spiral imaging,'' Magn. Reson. Med. **33**, No. 4, 549–556.
Noll, D. C., Meyer, C. H., Pauly, J. M., Nishimura, D. G., and Macovski, A. (**1991**). ''A homogeneity correction method for magnetic resonance imaging with time-varying gradients,'' IEEE Trans. Med. Imaging **10**, No. 4, 629–637.
Noll, D. C., Pauly, J. M., Meyer, C. H., Nishimura, D. G., and Macovski, A. (**1992**). ''Deblurring for non-2D Fourier transform magnetic resonance imaging,'' Magn. Reson. Med. **25**, 319–333.
Ong, D., and Stone, M. (**1998**). ''Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics,'' Phonoscope **1**, 1–13.
Perkell, J. S. (**1969**), ''Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study,'' (Cambridge, The MIT Press).
Perkell, J. S., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). ''Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements,'' J. Acoust. Soc. Am. **92**, No. 6, 3078–3096.
Santos, J. M., Wright, G., Yang, P. C., and Pauly, J. M. (**2002**). ''Adaptive architecture for real-time imaging systems,'' Proceedings of ISMRM, 10th Annual Meeting (Honolulu, HI), p. 468.
Schenk, J. F. (**1996**). ''The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds,'' Med. Phys. **23**, No. 6, 815–850.
Shadle, C. H., Mohammad, M., Carter, J., and Jackson, P. J. B. (**1999**). ''Dynamic magnetic resonance imaging: New tools for speech research,'' Proceedings of the 14th Int. Cong. Phon. Sci., pp. 623–626.
Shellock, H. G., Schatz, C. J., Julien, P. M., Silverman, J. M., Steinberg, F., Foo, T. K., Hopp, K. L., and Westbrook, P. R. (**1992**). ''Dynamic study of the upper airway with ultrafast spoiled GRASS MR imaging,'' J. Magn. Reson Imaging **2**, 103–107.
Stone, M. (**1990**). ''A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data,'' J. Acoust. Soc. Am. **87**, 2207–2217.
Stone, M., Faber, A., Raphael, L. J., and Shawker, T. H. (**1992**). ''Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [ʃ], and [l],'' J. Phonetics **20**, No. 2, 253–270.
Story, B. (**1995**). ''Physiologically based speech simulation using an enhanced wave-reflection model of the vocal tract,'' Ph.D. dissertation, University of Iowa.
Story, B., Titze, I., and Hoffman, E. (**1996**). ''Vocal tract area functions from magnetic resonance imaging,'' J. Acoust. Soc. Am. **100**, 537–554.
Subtelny, J. D., Oya, N., and Subtelny, J. D. (**1972**). ''Cineradiographic study of sibilants,'' Folia Phoniatr. **24**, 30–50.
Sulter, A., Miller, D., Wolf, R., Schutte, H., Wit, H., and Mooyaart, E. (**1992**). ''On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI,'' Magn. Reson. Imaging **10**, 365–373.
Watkin, K. L., and Rubin, J. M. (**1989**). ''Pseudo-three-dimensional reconstruction of ultrasonic images of the tongue,'' J. Acoust. Soc. Am. **85**, 496–499.

The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults[a)]

Susan Nittrouer[b)]
Utah State University UMC 6840, Logan, Utah 84322-6840

(Received 11 September 2003; revised 6 January 2004; accepted 9 January 2004)

Adults whose native languages permit syllable-final obstruents, and show a vocalic length distinction based on the voicing of those obstruents, consistently weight vocalic duration strongly in their perceptual decisions about the voicing of final stops, at least in laboratory studies using synthetic speech. Children, on the other hand, generally disregard such signal properties in their speech perception, favoring formant transitions instead. These age-related differences led to the prediction that children learning English as a native language would weight vocalic duration less than adults, but weight syllable-final transitions more in decisions of final-consonant voicing. This study tested that prediction. In the first experiment, adults and children (eight and six years olds) labeled synthetic and natural CVC words with voiced or voiceless stops in final C position. Predictions were strictly supported for synthetic stimuli only. With natural stimuli it appeared that adults and children alike weighted syllable-offset transitions strongly in their voicing decisions. The predicted age-related difference in the weighting of vocalic duration was seen for these natural stimuli almost exclusively when syllable-final transitions signaled a voiced final stop. A second experiment with adults and children (seven and five years old) replicated these results for natural stimuli with four new sets of natural stimuli. It was concluded that acoustic properties other than vocalic duration might play more important roles in voicing decisions for final stops than commonly asserted, sometimes even taking precedence over vocalic duration. [DOI: 10.1121/1.1651192]

PACS numbers: 43.71.Ft, 43.71.An [RLD] Pages: 1777–1790

I. INTRODUCTION

In 1955, Denes reported that vowel duration for the noun *use* (as in *the use*) was shorter than vowel duration for the verb *use* (as in *to use*). In a complementary perceptual test he found that the proportion of voiced judgments increased as vowel duration increased. The finding was revolutionary for its time because, contrary to the consensus opinion at that time, it showed that listeners can make phonetic judgments with information other than temporally discrete pieces of the spectral structure. Since that report, the relation between vowel duration and syllable-final consonant voicing has been well-studied in both speech production and perception.

Regarding speech production, there is no question that syllables with voiced final stops generally have longer vowels than syllables with similar phonetic structures in similar contexts, but with voiceless final stops. This effect is so pervasive that Chen (1970) went so far as to call it a *language-universal* phenomenon. Of course, the phenomenon cannot really be universal across languages, even if for the simple reason that some languages do not permit syllable-final obstruents. However, even among those that do, a few languages have been identified that lack a vowel-length distinction based on syllable-final consonant voicing. For example, Flege and Port (1981) reported that native Arabic speakers do not differentiate vowel length for Arabic words ending in voiced and voiceless final stops. This absence of a vowel-length effect for even a few languages demonstrates that the effect is not an inevitable consequence of syllable production. Nonetheless, most languages that permit syllable-final obstruents demonstrate a vowel-length distinction based on the voicing feature of the final consonant. In particular, studies of English have consistently demonstrated this effect (e.g., Chen, 1970; Crowther and Mann, 1992, 1994; Flege and Port, 1981; House and Fairbanks, 1953; Peterson and Lehiste, 1960). To be completely accurate, the entire voiced portion of a (stressed) syllable is shorter preceding a voiceless obstruent than preceding a voiced obstruent. In addition to the vowel nucleus, the voiced portion may consist of transitions into and out of the vowel nucleus, as well as sonorant consonants (Raphael, Dorman, Freeman, and Tobin, 1975; Raphael, Dorman, and Liberman, 1980). For this reason, the terms *vocalic length* and *vocalic duration* will be used in this manuscript instead of *vowel length* and *vowel duration*.

It is not clear why vocalic duration should be longer when the final obstruent is voiced than when it is voiceless. One reason that has been suggested is that speakers begin the closure gesture sooner for voiceless consonants because these closures require greater force than that required for voiced consonants, and people tend to begin relatively difficult tasks sooner than easier tasks (Malécot, 1970). However, there are numerous arguments against this suggestion, including the simple fact that the vocalic-length distinction for

[a)]Portions of this work were presented at the 141st meeting of the Acoustical Society of America, Chicago, June 2001.
[b)]Electronic mail: nittrouer@cpd2.usu.edu

final-consonant voicing is not universal. Be that as it may, given that the phenomenon is at least prevalent across languages with syllable-final obstruents we would expect vocalic length to be used by listeners of those languages in perception. And, indeed, numerous reports have shown that vocalic duration influences voicing judgments for syllable-final consonants made by adult speakers of languages with a vocalic-length distinction associated with final consonant voicing (Crowther and Mann, 1992, 1994; Denes, 1955; O'Kane, 1978; Raphael, 1972; Raphael *et al.*, 1975; Raphael *et al.*, 1980). Other acoustic properties of syllables ending with voiced or voiceless final stops have also been found to influence adults' voicing judgments (Hogan and Rozsypal, 1980; Summers, 1988; Wardrip-Fruin, 1982), with spectral characteristics associated with the vocal-tract closing gesture apparently being weighted particularly heavily (Hillenbrand, Ingrisano, Smith, and Flege, 1984). In fact, Hillenbrand *et al.* concluded that the release burst and voicing during closure contribute little to voicing decisions for final stops. This conclusion makes sense *prima facie* because in natural speech final stops may not be released, and speakers do not always voice during closure for obstruents (e.g., Klatt, 1976). However, the precise acoustic correlate of the vocal-tract closing gesture that contributes most to voicing decisions for final stops remains unclear. Fischer and Ohde (1990) tried to separate the contributions of first formant (F1) transition rate and F1 frequency at voicing offset to these decisions, and concluded that of the two, F1-offset frequency was weighted more strongly by adult listeners. Unfortunately, they did not manipulate higher formants, which also vary in frequency at voicing offset as a function of consonant voicing (Nittrouer, Estee, Lowenstein, and Smith, submitted). Furthermore, the changing (i.e., dynamic) nature of formant transitions is generally considered to be critical to speech recognition for both adults and children (e.g., Browman and Goldstein, 1990; Miranda and Strange, 1989; Nittrouer, Manning, and Meyer, 1993; Strange, 1989; Sussman, MacNeilage, and Hanson, 1973), owing in part to demonstrations that listeners can understand signals in which sinusoids are substituted for center formant frequencies (e.g., Remez, Rubin, Pisoni, and Carrell, 1981). In these ''sinewave'' signals, many acoustic properties traditionally associated with phonetic perception are missing. The current study did not separate the influences of formant transitions from discrete frequencies at voicing offset. Nonetheless, we considered entire transitions to have contributed to voicing decisions for final stops, when they contributed at all, because this is the signal property generally considered to influence voicing decisions in similar studies (e.g., Hillenbrand *et al.*, 1984; Wardrip-Fruin, 1982).

The current study focused on children's weighting of vocalic duration and syllable-offset transitions in decisions of voicing for syllable-final stops. This topic was selected for study because earlier work has suggested that young children in the range of three to eight years of age prefer dynamic signal properties to other sorts of properties for making phonetic decisions (e.g., Morrongiello, Robson, Best, and Clifton, 1984; Nittrouer, 1992; Parnell and Amerman, 1978). In particular, three studies have examined voicing decisions for syllable-final stops by three-year-olds, six-year-olds, and adults (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). Evidence across the three studies supports the suggestion that children fail to weight vocalic duration as strongly as adults, instead relying on formant transitions near voicing offset to make voicing decisions about final stops. While this result matches more general suggestions that young children weight dynamic signal properties (i.e., formant transitions) most strongly, and then gradually learn how other acoustic properties signal phonetic identity in their native language, the question of children's perception of syllable-final stop voicing was worth another look because of some irregularities in stimulus construction in the existing three studies. For example, step size on the vocalic-duration continuum in Greenlee was as large as 70 ms for some steps, instead of the 20–30 ms more commonly used in studies with only adult listeners.

One other study has also looked at children's abilities to use vocalic duration in voicing decisions about syllable-final stops. Lehman and Sharf (1989) created synthetic versions of *beet* and *bead* that differed in vocalic duration only (i.e., formant transitions were consistent across stimuli). Adults and children of the ages of five, eight, and ten years labeled these stimuli. The authors reported that category boundaries were similar across groups, but that the functions became steeper with increasing age. This age-related change in slopes indicates that vocalic duration was being weighted more heavily by older listeners, but unfortunately this study could make no comparison of the relative weighting of vocalic duration and syllable-offset transitions because formants were not manipulated. When only one acoustic property varies across stimuli in a labeling task, listeners must turn their perceptual attention to that property, if they are to do the task at all. Because Lehman and Sharf provide no data on how many children attempted to do the task, but failed, we are unable to estimate how many children were unable to make the required shift in their perceptual attention.

The question of how children weight vocalic duration and formant transitions in their voicing decisions for syllable-final stops is particularly intriguing because of the contradictory predictions that would be made based on children's speech perception capacities and on the nature of input to children. Again, studies of children's speech perception lead to the prediction that children would weight formant transitions at voicing offset as much as or more than adults do, but that children would weight vocalic duration less than adults. In addition to developmental studies, this prediction is supported by cross-linguistic data showing that adult native speakers of languages that either fail to have syllable-final obstruents or fail to make a vocalic-length distinction based on the voicing of the final consonant do not weight vocalic duration as much as adult native speakers of languages that make a vocalic-length distinction based on final-consonant voicing (Crowther and Mann, 1992, 1994; Flege and Wang, 1989). Such language-specific results for perceptual weighting strategies indicate that learning must be involved in the acquisition of the strategies under investigation.

On the other hand, Ratner and Luberoff (1984) showed that language directed to infants and toddlers (9 months to 2

years, 3 months) actually exaggerates the vocalic-length distinction, while often deleting the syllable-final consonant itself.[1] The acoustic characteristics that signal specific phonetic distinctions in a language are commonly reported to shape the listening strategies of infants by the time they are roughly one year of age (e.g., Jusczyk, 1997). Therefore, the Ratner and Luberoff finding leads to the prediction that young children would use vocalic length in making their voicing decisions concerning syllable-final stops as much as adult native speakers of the language they are learning. The current study tested these contradictory predictions.

II. EXPERIMENT 1: REPLICATING AND EXTENDING CROWTHER AND MANN

Crowther and Mann (1992, 1994) presented synthetic tokens of *pot* and *pod* in which the voicing of the final stop was signaled by the duration of the vocalic syllable portion and the offset frequency of F1. They found that listeners whose native language either did not permit syllable-final stops (Mandarin and Japanese) or did not show a difference in vocalic length as a function of the voicing of the final stop (Arabic) weighted vocalic duration less than native speakers of American English. No group differences were observed in the weighting of F1-offset frequency. Crowther and Mann concluded from their results that native-language experience generally shapes one's strategies for speech perception, but that some acoustic properties may be more malleable by experience than others. For this particular phonetic distinction, vocalic duration seems to have been more malleable than F1-offset frequency. Crowther and Mann offered no suggestions for what might make the weighting of vocalic duration more malleable than the weighting of F1-offset frequency, but the finding is in agreement with results from developmental studies. As described in the Introduction, developmental studies of speech perception have found that children generally weight dynamic components of the signal (i.e., formant transitions) more than other signal properties, including temporal properties (e.g., Morrongiello *et al.*, 1984; Nittrouer, 1992; Parnell and Amerman, 1978). In particular, several studies have reported that children between 3 and 6 years of age weight syllable-offset transitions more than adults, but vocalic duration less, in decisions of final-stop voicing (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). In sum, using formant transitions for the purpose of making phonetic decisions seems to be the ''default'' strategy. Apparently people learn to use other acoustic properties for making phonetic decisions through their experiences with a native language. Thus, the adult, non-native speakers of English who served as listeners in Crowther and Mann (1992, 1994) may have been exhibiting these default strategies in their labeling responses. It follows that children, who have less English experience than adults, might be expected to perform similarly to the non-native listeners in those two studies. This experiment tested that prediction.

A. Method

1. Listeners

Twenty nine adults between the ages of 19 and 39 years participated, as well as 30 children between 7 years, 11 months and 8 years, 5 months, and 29 children between 5 years, 11 months and 6 years, 5 months. In addition, 12 children between 3 years, 11 months and 4 years, 5 months participated, but seven of these children were unable to label stimuli reliably for any of the three sets of stimuli tested. As a result, testing with 4-year-olds was discontinued, and no data from this age group were included.

All participants were native speakers of American English, and had to meet additional criteria to participate. They had to pass a hearing screening of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB HL. Children needed to score at or above the 30th percentile on the Goldman-Fristoe Test of Articulation, Sounds-In-Words subtest (Goldman and Fristoe, 1986). Children could have had no more than six episodes of otitis media before their second birthday. Adults were administered the reading subtest of the Wide Range Achievement Test—Revised (Jastak and Wilkinson, 1984), and needed to demonstrate at least an 11th grade reading level.

2. Equipment and materials

Testing took place in a soundproof booth, with the computer that controlled the experiment in an adjacent room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. Stimuli were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones. The experimenter recorded responses with a keyboard connected to the computer. Two hand-drawn pictures (8 in.$\times$8 in.) were used to represent each response label in each experiment, such as a *buck* (a male deer) and a *bug*. Gameboards with ten steps were also used with children: they moved a marker to the next number on the board after each block of test stimuli. Cartoon pictures were used as reinforcement and were presented on a color monitor after completion of each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

3. Stimuli

Three sets of stimuli were created for this experiment: synthetic *pot/pod*, synthetic *buck/bug*, and natural *buck/bug*. Both sets of synthetic stimuli were created using the Sensyn Laboratory Speech Synthesizer. The synthetic *pot/pod* stimuli were identical to the stimuli used by Crowther and Mann (1992, 1994), except that Crowther and Mann used three settings for F1-offset frequency and we used only two: the highest and lowest values used by them. The vocalic duration of these stimuli varied from 100 to 260 ms, in 20 ms steps. All vocalic portions were preceded by a 50 ms interval of aspiration noise. Fundamental frequency (f0) started at 138 Hz and fell linearly throughout the stimulus to an offset frequency of 95 Hz. F3 was constant throughout at 2460 Hz. F2 was constant at 1160 Hz until 50 ms before offset, at which time it began rising to an ending frequency of 1425 Hz. In all stimuli, F1 was constant at 675 Hz until 50 ms before offset, at which time it fell to either 555 Hz (most

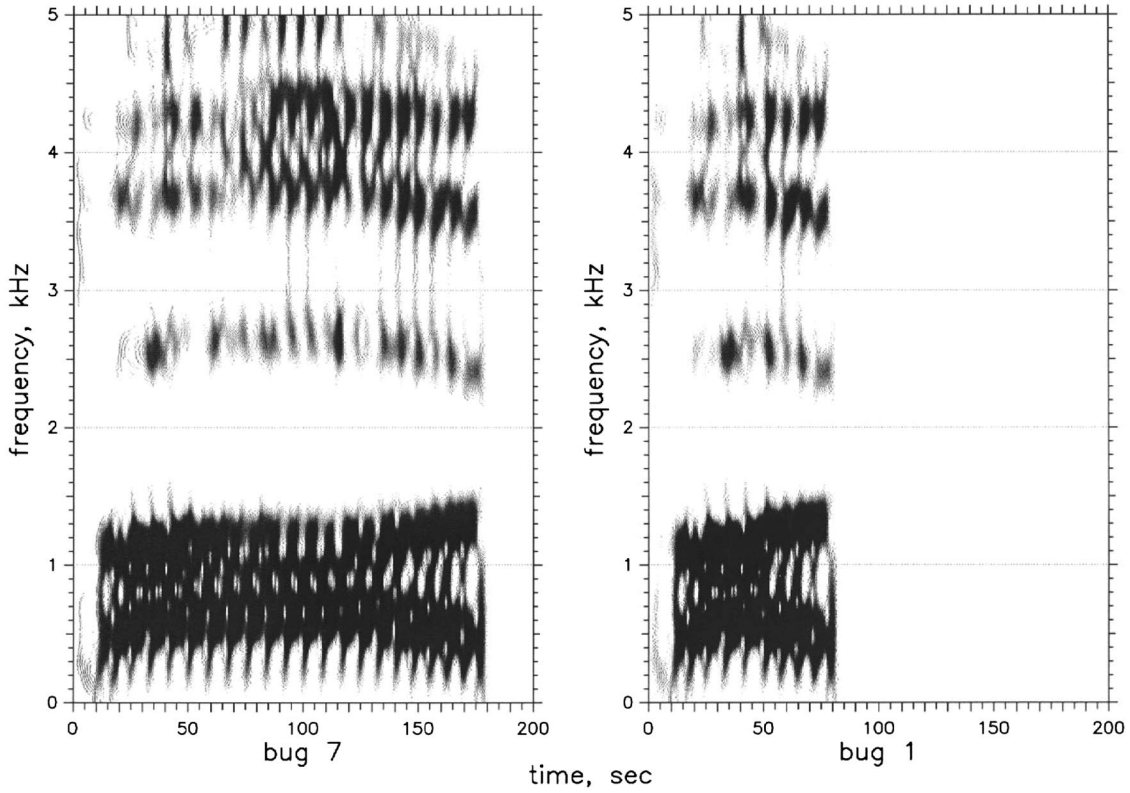


FIG. 1. Spectrograms of the longest (bug 7) and shortest (bug 1) *bug* stimuli created from the same natural token.

pot-like) or 355 Hz (most *pod*-like). Thus there were 18 stimuli: two F1-offset frequencies × nine vocalic durations.

The synthetic *buck/bug* stimuli were created to be as similar as possible to the *pot/pod* stimuli, while still evoking *buck* and *bug* responses. The vocalic duration of these stimuli also varied from 100 to 260 ms in 20 ms steps. The f0 began at 138 Hz and fell throughout to 95 Hz. However, both F2 and F3 had offset transitions: F2 was constant at 1000 Hz until 50 ms before offset, at which time it rose linearly to 1800 Hz. F3 was constant at 2700 Hz until 50 ms before offset, at which time it fell to its ending frequency of 2000 Hz. F1 started at 400 Hz, and rose linearly to 600 Hz over the first 50 ms. For half of the stimuli F1 remained at 600 Hz until stimulus offset. For the other half of the stimuli, F1 fell over the final 50 ms to 450 Hz. Thus there were 18 of these stimuli: two F1-offset frequencies × nine vocalic durations.

The natural *buck/bug* stimuli were created from natural tokens of an adult, male speaker producing these words in isolation. The speaker was a native speaker of American English. He produced ten tokens of each word, in randomized order. The three tokens of each word that matched each other most closely in vocalic duration and f0 contour were selected for modification. With each token, the release burst and any voicing during closure was deleted. Vocalic length was then manipulated either by reiterating a single pitch period from the most stable spectral region of the syllable (to lengthen syllables) or by deleting pitch periods from the most stable spectral region of the syllable (to shorten syllables). For both kinds of manipulation, care was taken to align signal portions at zero crossings so no audible clicks resulted. Also, initial and final formant transitions were not disrupted. Seven stimuli varying in vocalic duration from roughly 85 to 176 ms were created from each token this way. These endpoint values were selected because they match the mean lengths of natural *buck* and *bug*, but clearly the continuum was briefer than that of the synthetic stimuli. The step size was 15 ms (2 pitch periods) on average, but obviously these steps varied slightly according to small differences in f0 across stimuli. The mean durations of the three stimuli made from *buck* and of the three stimuli made from *bug* at each step were within 5 ms of each other. This procedure for manipulating natural tokens of words ending in voiced and voiceless final stops differs from procedures used in most earlier studies, in that we used both voiced and voiceless tokens, and we did not disrupt offset transitions. Figure 1 shows the longest and shortest stimuli created from the same *bug* token, and shows that this method of modifying stimuli was successful in only affecting the duration of the stable syllable portion. In all, 42 natural stimuli were created: seven vocalic durations × two kinds of offset transitions × three tokens of each.

4. Procedures

All participants attended two testing sessions. At the first session, screening tasks were completed first. Next, one set of the *buck/bug* stimuli (either synthetic or natural) was presented, with the choice of which set to present randomized across listeners. At the second session, the *pot/pod* and the other set of *buck/bug* stimuli were presented.

The same procedures were followed for each set of stimuli. Practice items were presented before the testing began. Practice items consisted of the best exemplars of each

category, which were the stimuli that should most strongly evoke voiced and voiceless percepts. For example, the best exemplar of *pot* was the stimulus with a 100 ms vocalic portion and the 555 Hz F1 offset. The best exemplar of *pod* was the stimulus with a 260 ms vocalic portion and the 355 Hz F1 offset. For the two synthetic sets of stimuli, there were just two best exemplars (one voiced and one voiceless). Each of these was played five times in random order, and the listener had to respond to at least nine of them correctly to proceed to testing. For the natural *buck/bug* stimuli, there were the six best exemplars (three *buck* and three *bug*). Each of these was played twice, with the 12 stimuli presented in random order. The listener had to respond correctly to at least 11 to proceed to testing.

During testing, ten blocks of stimuli were presented. There were 18 stimuli per block during testing with the two sets of synthetic stimuli. There were 14 stimuli per block during testing with the natural *buck/bug* stimuli. Because there were actually three tokens of each natural stimulus (i.e., each vocalic duration at each level of syllable-offset transitions), the program randomly selected one of the tokens to present during the first block, and then repeated this random selection during the next block without replacement. After the first three blocks, this process was repeated until testing was completed.

Listeners responded by saying the label and pointing to the picture that represented their selection. Having both kinds of responses served as a check that participants were paying attention to the task. In the rare instance that a listener pointed to one picture, but said the other label, the experimenter gave a gentle reminder to pay attention and the stimulus was replayed. To have their data included in the final analysis, participants needed to give at least 80% correct responses to the best exemplars during testing. This requirement was an additional assurance that data were analyzed only from participants who maintained attention to the task.

For children, cartoon pictures were displayed on the monitor and a bell sounded at the end of each block. They moved a marker to the next space on a gameboard after each block as a way of keeping track of how much more time they had left in the test.

Several methods have been used to characterize the weights assigned to various acoustic properties in labeling tasks. Traditionally, each listener's data are plotted as cumulative distributions of the proportion of one response (e.g., *pod* responses) across levels of the acoustic property manipulated in a continuous fashion (vocalic duration in this experiment) for each level of the acoustic property manipulated in a noncontinuous fashion (formant-offset frequencies in this experiment). Best-fit lines are then obtained, often using probit analysis (Finney, 1964), and slopes and distribution means (i.e., phoneme boundaries) computed. This method can extrapolate so that phoneme boundaries outside of the range tested can be obtained. However, we typically impose limits on the values that extrapolated phoneme boundaries can take, restricting them to 3.5 steps beyond the lowest and highest value tested. Because the null hypothesis is that listeners of all ages will show similar differences between phoneme boundaries, this restriction serves only to constrain the probability of rejecting that null hypothesis. The mean slope of the functions is taken as an indication of the weight assigned to the continuously varied property: the steeper the functions, the more weight that was assigned to that property. The separation between functions at the phoneme boundaries (for each level of the noncontinuous property) is taken as an indication of the weight assigned to that noncontinuous property: the greater the separation, the greater the weight that was assigned. More recently, some investigators (e.g., Turner *et al.*, 1998) have started computing partial correlation coefficients (partial *r*s) for each of the acoustic properties manipulated in the experiment and the proportion of one response option given (i.e., looking at how well each acoustic property predicts responses). Both kinds of metrics were computed in this experiment.

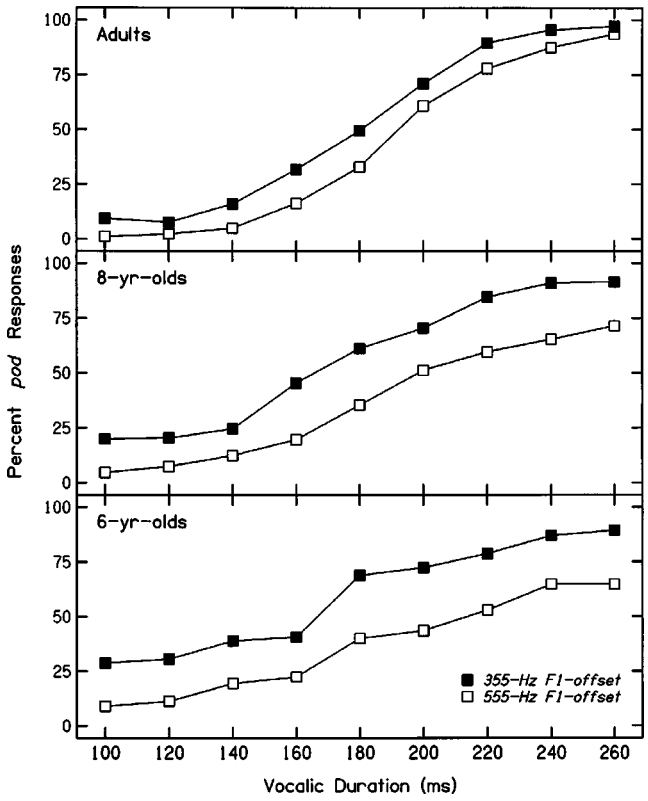


FIG. 2. Mean labeling functions for each age group for synthetic *pot/pod* stimuli, Experiment 1.

B. Results

1. Pot/pod

Data were excluded for twelve 6-year-olds and four 8-year-olds because they either failed to label nine out of ten practice stimuli correctly or did not maintain 80% correct responses during testing itself. Data were included for 29 adults, 26 8-year-olds, and 17 6-year-olds.

Figure 2 shows mean labeling functions from each age group for the synthetic *pot/pod* stimuli. Table I provides two estimates of the weighting of vocalic duration (mean slopes and partial *r*s for vocalic duration) and two estimates of the weighting of F1-offset frequency (separation in phoneme boundaries and partial *r*s for F1-offset frequency). In many

TABLE I. Mean slopes, separations in functions (at phoneme boundaries), and partial *r*s (for vocalic duration and F1 offset) for the *pot/pod* labeling task, Experiment 1. *Note*: Standard deviations (SDs) are given in parentheses.

	Adults	8-year-olds	6-year-olds
Weight assigned to vocalic duration			
Mean slope, across functions	0.64 (0.21)	0.43 (0.21)	0.31 (0.18)
Mean partial *r* for vocalic duration	0.91 (0.04)	0.78 (0.20)	0.72 (0.17)
Weight assigned to F1-offset frequency			
Mean separation in functions	0.93 (0.93)	2.91 (2.99)	3.64 (3.12)
Mean partial *r* for F1 offset	0.14 (0.14)	0.33 (0.23)	0.41 (0.23)

labeling experiments, slope is given with a physical reference (e.g., change in probit units per ms of change in vocalic duration). In this experiment, however, step size on the vocalic-duration continuum differed for the two synthetic stimulus sets and for the one natural stimulus set (20 ms per step for the synthetic stimuli; 15 ms per step for the natural). For that reason, the slope is given here as the change in probit units per step on the vocalic-duration continuum. Phoneme boundaries represent the step on the vocalic-duration continuum at which the function reaches the 50th percentile. The separation in phoneme boundaries is given here using steps again. Both Fig. 2 and Table I indicate that the weight that was assigned to vocalic duration increased with increasing age: the functions became steeper, and partial *r*s for vocalic duration increased. At the same time, the weight that was assigned to the F1-offset frequency diminished with increasing age: the separation in functions decreased, as did partial *r*s for F1-offset frequency.

One-way analyses of variance (ANOVAs), with age as the factor, were performed on mean slopes (across the functions with the 555 Hz and 355 Hz F1 offsets), mean separations in phoneme boundaries, and partial *r*s for vocalic duration and F1-offset transition. Because the main effect of age was significant for all these measures, post hoc *t* tests were also performed. Results of these ANOVAs and *t* tests are presented in Table II. Precise results are given for any analysis with *p* less than 0.10. When *p* is greater than 0.10, results are simply described as nonsignificant (NS). For post hoc *t* tests, both actual, computed *p* values as well as the Bonferroni significance levels, adjusted for multiple *t* tests, are provided. In all cases, the results of the statistical tests generally support impressions of age-related differences gleaned from Fig. 2 and Table I: Children's labeling functions were shallower than those of adults, but were more widely separated. Children's partial *r*s were smaller than those of adults for vocalic duration, but were greater for F1-offset. Although it appears from Table I that 8-year-olds performed more like adults than did 6-year-olds, no statistically significant differences were observed between 6- and 8-year-olds on any of the measures.

2. Synthetic buck/bug

Fourteen 6-year-olds, 13 8-year-olds, and six adults were unable to reach the criteria for having their data included on either training or testing. Consequently, data are included for 15 6-year-olds, 17 8-year-olds, and 23 adults.

TABLE II. Results of ANOVAs for the *pot/pod* labeling task, Experiment 1. *Note*: Degrees of freedom for the main effect of age are 2, 69. Degrees of freedom for the post hoc *t* tests are 69.

	F or *t*	*p*	Bonferroni significance level
Slopes			
Age effect	15.83	<0.001	
6- vs 8-year-olds	−1.80	NS	NS
6-year-olds vs adults	−5.30	<0.001	<0.001
8-year-olds vs adults	−3.92	<0.001	<0.001
Partial *r*s for vocalic duration			
Age effect	9.69	<0.001	
6- vs 8-year-olds	−1.19	NS	NS
6-year-olds vs adults	−4.07	<0.001	<0.001
8-year-olds vs adults	−3.23	0.002	<0.01
Separation in phoneme boundaries			
Age effect	7.96	<0.001	
6- vs 8-year-olds	0.96	NS	NS
6-year-olds vs adults	3.64	<0.001	<0.01
8-year-olds vs adults	3.01	0.004	<0.05
Partial *r*s for F1-offset frequency			
Age effect	11.16	<0.001	
6- vs 8-year-olds	1.22	NS	NS
6-year-olds vs adults	4.34	<0.001	<0.001
8-year-olds vs adults	3.51	<0.001	<0.01

Figure 3 shows mean labeling functions for each group, and Table III provides the same estimates of the weighting of vocalic duration and F1-offset frequency, as shown in Table I. As found for the *pot/pod* stimuli, children appear to have weighted vocalic duration less than adults, but to have

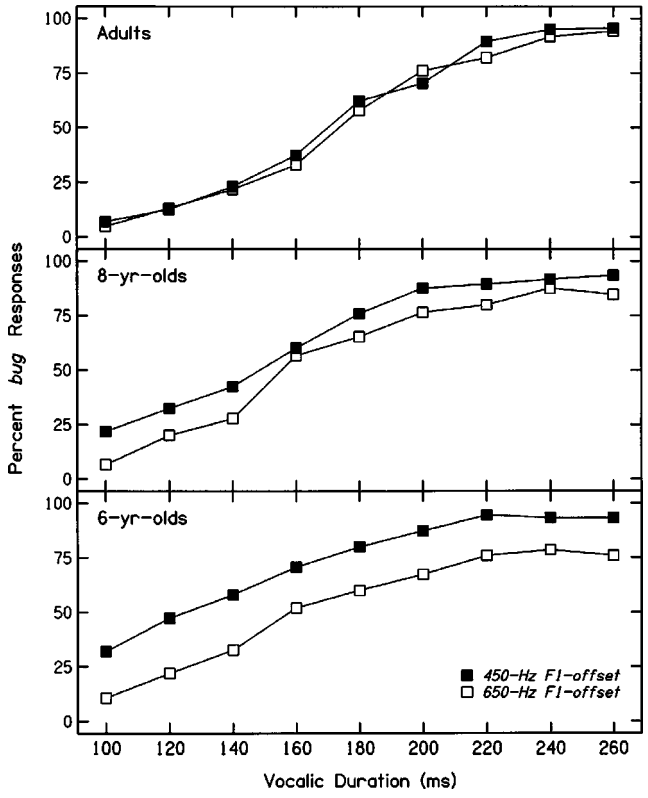


FIG. 3. Mean labeling functions for each age group for synthetic *buck/bug* stimuli, Experiment 1.

TABLE III. Mean slopes, separations in functions (at phoneme boundaries) and partial *r*s (for vocalic duration and F1-offset) for the synthetic *buck/bug* labeling task, Experiment 1. *Note*: Standard deviations (SDs) are given in parentheses.

	Adults	8-year-olds	6-year-olds
Weight assigned to vocalic duration			
Mean slope, across functions	0.53	0.41	0.37
	(0.18)	(0.19)	(0.25)
Mean partial *r* for vocalic duration	0.91	0.83	0.74
	(0.04)	(0.13)	(0.13)
Weight assigned to F1-offset frequency			
Mean separation in functions	0.24	1.51	2.80
	(0.82)	(1.64)	(1.75)
Mean partial *r* for F1 offset	0.06	0.18	0.37
	(0.09)	(0.16)	(0.16)

weighted F1-offset frequency more. Results for the one-way ANOVAs and post hoc *t* tests, shown in Table IV, confirm these impressions. Unlike the *pot/pod* stimuli, however, there were some statistically significant differences in performance between the 6- and 8-year-olds, with the 8-year-olds performing a little more similarly to adults.

3. Natural buck/bug

One 6-year-old and two 8-year-olds failed to reach criteria on either the training or testing for having their data included. Therefore data were analyzed for 28 6-year-olds, 28 8-year-olds, and 29 adults.

Figure 4 shows mean labeling functions for each age group, and Table V provides the same estimates of the weighting of vocalic duration and formant offsets as provided for the synthetic *pot/pod* and *buck/bug* stimuli. A very different pattern of responding is apparent for these natural stimuli than what was found for the synthetic stimuli. Labeling functions are widely separated, depending on whether stimuli were created from tokens with voiced or voiceless final stops. This pattern is apparent for listeners in all three age groups.

TABLE IV. Results of ANOVAs for the synthetic *buck/bug* labeling task, Experiment 1. *Note*: Degrees of freedom for the main effect of age are 2, 52. Degrees of freedom for the post hoc *t* tests are 52.

	F or *t*	*p*	Bonferroni significance level
Slopes			
Age effect	3.46	=0.039	
6- vs 8-year-olds	−0.53	NS	NS
6-year-olds vs adults	−2.42	=0.019	<0.010
8-year-olds vs adults	−1.93	=0.059	NS
Partial *r*s for vocalic duration			
Age effect	13.95	<0.001	
6- vs 8-year-olds	−2.61	=0.012	<0.05
6-year-olds vs adults	−5.26	<0.001	<0.001
8-year-olds vs adults	−2.57	=0.013	<0.05
Separation in phoneme boundaries			
Age effect	15.48	<0.001	
6- vs 8-year-olds	2.61	=0.012	<0.05
6-year-olds vs adults	5.53	<0.001	<0.001
8-year-olds vs adults	2.85	0.006	<0.05
Partial *r*s for F1-offset frequency			
Age effect	22.33	<0.001	
6- vs 8-year-olds	3.80	<0.001	<0.01
6-year-olds vs adults	6.68	<0.001	<0.001
8-year-olds vs adults	2.72	0.001	<0.05

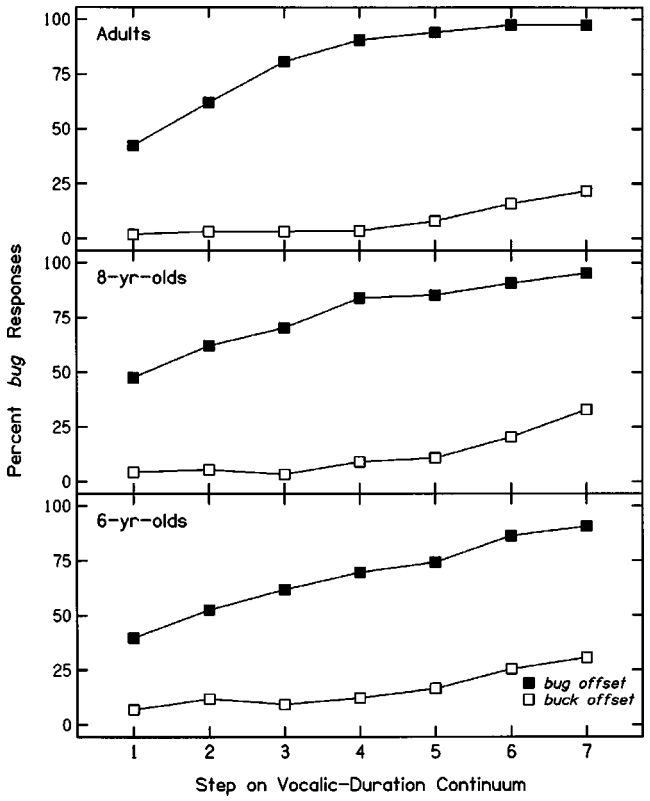


FIG. 4. Mean labeling functions for each age group for natural *buck/bug* stimuli, Experiment 1.

Table VI shows the results of the one-way ANOVAs conducted on all four measures. Unlike the statistical results for both sets of synthetic stimuli, these results differ for the two measures of each acoustic property. For the measures of weight assigned to vocalic duration (i.e., slopes and partial *r*s for vocalic duration), the analysis of slope showed a significant age effect, but the analysis of partial *r*s did not. It appears from Fig. 4 that this result is due to adults' function for stimuli with *bug* offsets being steeper than those of either children's group. Functions for stimuli with *buck* offsets appear similar in shape across age groups. One-way ANOVAs conducted on slope for each function separately confirm this

TABLE V. Mean slopes, separations in functions (at phoneme boundaries), and partial *r*s (for vocalic duration and F1-offset) for the natural *buck/bug* labeling task, Experiment 1. *Note*: Standard deviations (SDs) are given in parentheses.

	Adults	8-year-olds	6-year-olds
Weight assigned to vocalic duration			
Mean slope, across functions	0.48	0.33	0.28
	(0.27)	(0.17)	(0.21)
Mean partial *r* for vocalic duration	0.31	0.35	0.40
	(0.18)	(0.20)	(0.21)
Weight assigned to formant offsets			
Mean separation in functions	7.89	7.92	7.27
	(2.83)	(2.61)	(2.77)
Mean partial *r* for formant offsets	0.86	0.84	0.74
	(0.15)	(0.13)	(0.20)

TABLE VI. Results of ANOVAs for the natural *buck/bug* labeling task, Experiment 1. *Note*: Degrees of freedom for the main effect of age are 2, 82. Degrees of freedom for the post hoc *t* tests are 82.

	F or *t*	*p*	Bonferroni significance level
Slopes			
Age effect	5.73	=0.005	
6- vs 8-year-olds	−0.78	NS	NS
6-year-olds vs adults	−3.24	=0.002	<0.01
8-year-olds vs adults	−2.45	=0.017	<0.05
Partial *r*s for vocalic duration			
Age effect	1.46	NS	
Separation in phoneme boundaries			
Age effect	0.50	NS	
Partial *r*s for formant offsets			
Age effect	4.75	=0.01	
6- vs 8-year-olds	−2.40	=0.019	<0.10
6-year-olds vs adults	−2.88	=0.005	<0.05
8-year-olds vs adults	−0.46	NS	NS

impression. The main effect of age is significant only for functions for stimuli with *bug* offsets, $F(2,82)=10.59$, $p<0.001$. Post hoc *t* tests done on slopes of functions for stimuli with *bug* offsets reveal the same trends as seen in Table VI: There is no statistically significant difference in slopes for 6- and 8-year-olds' functions, but there are significant differences for 6-year-olds versus adults, $t(82)=-4.41$, $p<0.001$ (Bonferroni significance level=<0.001), and for 8-year-olds versus adults, $t(82)=-3.30$, $p=0.001$ (Bonferroni significance level=<0.01). In brief, it appears from Fig. 4 that listeners of all ages gave very few *bug* responses when formant offsets failed to signal that the vocal tract was closing (i.e., when stimuli had *buck* formant offsets). However, when formant offsets signaled a closing vocal tract, adults provided *bug* responses more consistently, at shorter vocalic durations than children did.

For the measures of weight assigned to formant offsets (i.e., separations in functions and partial *r*s for formant offsets), the analysis of separations in functions showed no age effect, but the analysis of partial *r*s did. This last result can actually be traced to 6-year-olds weighting formant offsets less than adults or 8-year-olds did.

C. Discussion

This experiment was undertaken to examine whether children would be found to weight formant transitions more and vocalic duration less than adults in decisions of syllable-final stop voicing, as has been reported by three earlier studies (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). For this purpose, the synthetic stimuli of Crowther and Mann (1992, 1994) were replicated, and another set of synthetic stimuli was created that closely matched those stimuli. Stimuli were also created from natural tokens of words ending in voiced and voiceless final stops. Adults' results for both sets of synthetic stimuli closely matched those of Crowther and Mann: vocalic duration explained variation in their responses to a large extent, with F1 transitions explaining little. This result agrees with the traditional view of the role of vocalic duration in decisions of syllable-final consonant voicing. It was found that 6- and 8-year-old children weighted vocalic duration less than adults, and weighted F1-offset transitions more. This result strongly supports the assertion that children attend largely to dynamic signal components in speech perception, gradually modifying their perceptual strategies to weight other components more as they gain experience with their native language (e.g., Nittrouer, Manning, and Meyer, 1993).

The third set of stimuli included in this experiment was generated from natural tokens of a speaker saying *buck* and *bug*. All information that might normally be available from the closure and release was removed, but differences between stimuli that exist in the syllable nuclei and margins remained. For example, kinematic studies have shown that jaw opening is faster and more extensive for words with voiceless, rather than voiced, final stops (Gracco, 1994; Summers, 1987). It follows that the rate of formant transition at the beginnings of syllables is faster for words with voiceless stops at the end, and maximum F1 frequency is greater for words with voiceless, rather than voiced, final stops. For example, Nittrouer *et al.* (submitted) found that F1 at the syllable center was 58 Hz higher in *buck* than in *bug*. Finally, the rate and final frequency of all formant transitions at the syllable's end can differ, depending on stop voicing. In this experiment, when all these acoustic properties varied naturally, partial *r*s indicated that listeners of all ages weighted vocalic duration very little. Instead, listeners largely weighted the other acoustic properties that correlate with the voicing of the final stop. Based on the studies of others (e.g., Hillenbrand *et al.*, 1984), it seems likely that the property that accounted for most of the variation in listeners' responses was the collective pattern across formants of syllable-final transitions. But there is no way to know that for sure without further experiments. In any event, these results with stimuli created from natural tokens require a tempering of the conclusions reached with synthetic stimuli.

There is some reason to suggest, however, that partial *r*s did not reveal the whole story in this case. When asked to label stimuli created from natural tokens, adults had steeper functions for those stimuli with syllable-offset transitions appropriate for voiced final stops, rather than those with transitions appropriate for voiceless final stops. That is, their *bug* responses approached 100% at briefer vocalic durations, and remained at that level of responding across much of the vocalic-duration continuum. Steep labeling functions are generally taken as an indication that listeners weighted the property represented on the *x* axis strongly, and so it would be reasonable to conclude that adults weighted vocalic duration in their voicing decisions when the final stop indicated complete vocal-tract closure. Children, on the other hand, show no indication of weighting vocalic duration strongly for any stimuli.

Of course, any time results for one set of stimuli are so strikingly different from results for other, similar stimuli, the possibility arises that perhaps those results are spurious. This concern was exacerbated here because the range of vocalic durations used was briefer in the natural stimuli than in either set of synthetic stimuli. Consequently, before the conclusion was firmly reached that vocalic duration actually has little influence on listeners' judgments of syllable-final con-

sonant voicing in natural listening conditions, it seemed important to examine perceptual responses for other stimuli created from natural tokens. For that reason, another experiment was conducted.

III. EXPERIMENT 2: NATURAL STIMULI

The purpose of this experiment was to test whether adults' and children's labeling responses for a wider range of natural stimuli would reveal similar weighting strategies to those found for the stimuli created from natural tokens in Experiment 1. In this second experiment, four sets of stimuli were created from natural tokens of word pairs ending in voiced or voiceless final stops, using the same procedure as was used to create the natural stimuli in Experiment 1 (i.e., manipulating vocalic duration by reiterating or deleting pitch periods at the most stable spectral region of the syllable).

A. Method

1. Listeners

Eleven adults between the ages of 20 and 31 participated. In addition, 13 7-year-olds (between 6 years, 11 months and 7 years, 5 months), 25 5-year-olds (between 4 years, 11 months and 5 years, 5 months) and 17 $3\frac{1}{2}$-year-olds (between 3 years, 5 months and 3 years, 11 months) participated. All participants needed to meet the same criteria as those in Experiment 1.

2. Equipment and materials

The same equipment and materials were used in this experiment as in Experiment 1, except that new pictures were created because different stimuli were used.

3. Stimuli

Four sets of stimuli were created, with care taken to vary the place of constriction across vowels and consonants: *cop/cob*, *boot/booed*, *feet/feed*, and *pick/pig*. The same male speaker who provided tokens of the natural *buck* and *bug* used in Experiment 1 provided tokens for use in the creation of these stimuli. The same procedures as used in Experiment 1 were used to modify these stimuli to create seven-step continua, varying in vocalic duration. In particular, care was again taken not to disrupt offset transitions. Of course, the four vowels used in this experiment differ in intrinsic duration, and so the continua differed in range: *cop/cob* varied from 82 to 265 ms; *boot/booed* varied from 97 to 258 ms; *feet/feed* varied from 93 to 255 ms; and *pick/pig* varied from 62 to 178 ms. For each set there were 42 stimuli: seven steps on the vocalic duration continuum $\times$ two formant-offset conditions $\times$ three tokens of each.

4. Procedures

The procedures for this experiment were essentially the same as for Experiment 1. All participants attended two sessions. However, all four sets of stimuli could be presented to individual listeners only in the adult and 7-year-old groups. This was because only adults and 7-year-olds could complete two sets of stimuli at the first session, after completing the screening procedures. These listeners were then presented with the final two sets of stimuli at the second session. The order of presentation of sets was randomized across these older participants. The $3\frac{1}{2}$- and 5-year-olds were presented with only three sets of stimuli: one set at the first session (after the hearing and speech screening tasks), and two sets at the second session. The stimuli that would be presented to a child were selected in a serial fashion. That is, the first child tested in each age group was presented with three sets of stimuli, randomly selected. The next child was presented with the fourth set, and the first two sets presented to the first child. The third child tested was then presented with the third and fourth sets, as well as the first, and so on.

One additional training procedure was incorporated into the protocol in this second experiment. Before administering the training with the best exemplars, unaltered stimuli were presented for practice. These unaltered stimuli had whatever voicing during closure was present in the original stimuli, as well as the release bursts. As in practice with the best exemplars, there were six of these stimuli (i.e., three of the word with a voiceless final stop and three of the word with a voiced final stop). Each stimulus was presented two times, and the listener had to respond to 11 of these items correctly in order to receive the training with the best exemplars.

B. Results

Data from the $3\frac{1}{2}$-year-olds were not included because not enough children in this age group were able to complete the tasks. Of the 17 children in this age group participating, two failed the hearing screening, and two refused to cooperate at all. Of the remaining 13 children, some became uncooperative on the second day after the presentation of one set of stimuli, and so could not be presented with the other set. In all, these $3\frac{1}{2}$-year-olds were presented with 33 sets of stimuli. As a group they reached the training criterion with the unaltered stimuli for 24 of these sets. Subsequently, the criterion for training with the best exemplars was reached for 17 sets of stimuli. However, $3\frac{1}{2}$-year-olds reached the 80% correct criterion for endpoint stimuli during testing for just five sets of stimuli. That was not enough to provide useful information about the weighting strategies of these young listeners.

One 5-year-old failed the hearing screening, and one failed to reach criterion on the Goldman-Fristoe Test of Articulation. Regarding the training with unaltered stimuli, one 5-year-old failed to reach criterion for *feet/feed* and one failed to reach criterion for both *cop/cob* and *boot/booed*. Several 5-year-olds failed to reach either the best-exemplar training criterion or the testing criterion for having their data included in the analyses. Specifically, four 5-year-olds (out of 18 tested) failed to reach criterion for *pick/pig*; seven (out of 17 tested) failed to reach criterion for *cop/cob*; five (out of 15 tested) failed to reach criterion for *feet/feed*; and eight (out of 18 tested) failed to reach criterion for *boot/booed*.

Only one 7-year-old failed to reach the testing criterion for having data included in the analyses for *feet/feed*, and one failed to reach the testing criterion for *pick/pig*. Data for one adult participant was lost for *cop/cob* because the experimenter neglected to save the data after testing.

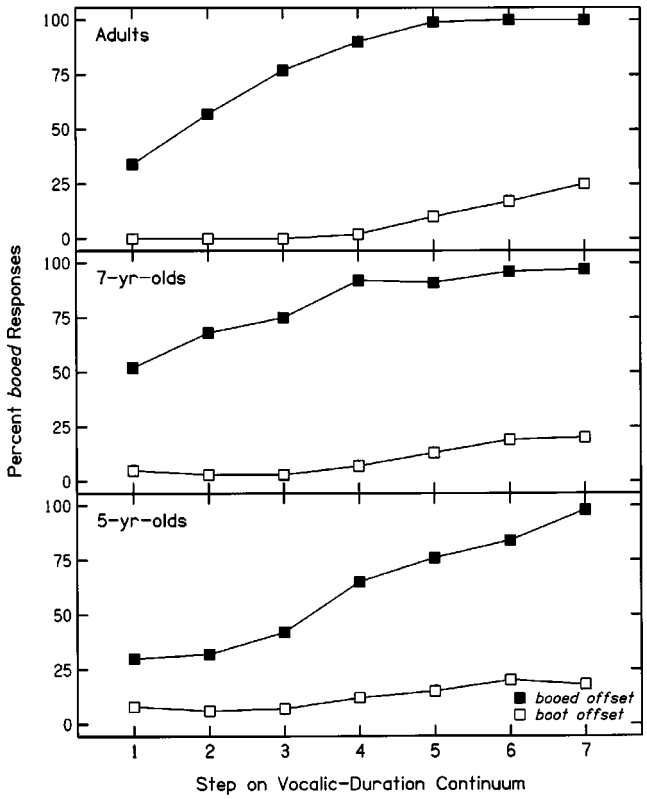


FIG. 5. Mean labeling functions for each age group for natural *boot/booed* stimuli, Experiment 2.

Results for all four sets of stimuli were similar, and are illustrated in Fig. 5, showing mean labeling functions for *boot/booed*. Clearly, listeners in all three age groups showed fairly flat functions with large separations between functions. This pattern of results traditionally is interpreted as indicating that listeners weighted only slightly the property varied continuously (in this case, vocalic duration) and weighted heavily the property varied dichotomously (in this case, syllable-offset transitions).

Two estimates of the weighting of vocalic duration and syllable-offset transitions were used in Experiment 1. Both estimates gave similar results, except where the weighting of vocalic duration in decisions for natural stimuli were concerned. In that case, partial *r*s indicated that adults and children weighted vocalic duration similarly, but slopes showed that adults weighted vocalic duration more than children. This age-related difference was particularly evident when stimuli were created from words with voiced final stops. Because of this result, it was concluded that slopes and separations between labeling functions provided more accurately described labeling results. At the same time, however, it is difficult to compare the weighting of the two acoustic properties using slopes and separations between functions because they provide different metrics. Partial *r*s allow us to do this kind of comparison. Consequently, both kinds of estimates are presented for this second experiment, but with an emphasis on slopes and separations between functions.

TABLE VIII. Mean separations in functions (at phoneme boundaries), Experiment 2. *Note*: Standard deviations (SDs) are given in parentheses.

	Adults	7-year-olds	5-year-olds
cop/cob	3.75	1.62	2.95
	(2.57)	(3.84)	(3.27)
feet/feed	5.01	6.18	6.59
	(2.42)	(1.35)	(3.34)
boot/booed	7.02	7.39	6.75
	(2.34)	(2.51)	(1.40)
pick/pig	9.42	9.03	9.87
	(2.97)	(2.42)	(2.08)

Table VII shows partial *r*s for each stimulus set. From these estimates it appears that listeners in all age groups performed similarly by weighting syllable-offset transitions strongly while weighting vocalic duration not as much. The one-way ANOVAs (with age as the main effect) done on each set of the partial *r*s provided eight analyses: one for vocalic duration and one for formant offsets for each of the four stimulus sets. Unlike results for natural *buck/bug* in Experiment 1, where a significant age effect was found for partial *r*s of formant offsets, none of these ANOVAs revealed a significant age effect. At least for *cop/cob* and *boot/booed*, both of which had ranges of group means similar to that found for natural *buck/bug* in Experiment 1, this failure to find a significant effect might be due to smaller sample sizes in this second experiment, compared to the first.

Mean separations in functions for each stimulus set are presented in Table VIII. From these estimates it is clear that listeners in all age groups weighted syllable-offset transitions similarly and strongly. The one-way ANOVAs done on separations in phoneme boundaries revealed no significant age effects.

Table IX shows mean slopes across functions, and for stimuli created from words with voiceless and voiced final stops separately. As was found for natural *buck/bug* in Experiment 1, it appears as if there is an age-related increase in

TABLE VII. Partial *r*s for each stimulus set, Experiment 2. *Note*: Standard deviations (SDs) are given in parentheses.

	Vocalic duration			Formant offsets		
	Adults	7-year-olds	5-year-olds	Adults	7-year-olds	5-year-olds
cop/cob	0.48	0.54	0.55	0.77	0.66	0.66
	(0.20)	(0.15)	(0.18)	(0.17)	(0.12)	(0.19)
feet/feed	0.56	0.49	0.52	0.69	0.72	0.69
	(0.20)	(0.11)	(0.22)	(0.18)	(0.07)	(0.26)
boot/booed	0.37	0.43	0.46	0.84	0.76	0.71
	(0.22)	(0.19)	(0.13)	(0.15)	(0.13)	(0.18)
pick/pig	0.22	0.28	0.22	0.93	0.88	0.90
	(0.18)	(0.18)	(0.17)	(0.08)	(0.10)	(0.08)

TABLE IX. Mean slopes, across functions, and for voiceless and voiced stimuli separately, Experiment 2. *Note*: Standard deviations (SDs) are given in parentheses.

	Adults	7-year-olds	5-year-olds
Mean slope, across functions			
cop/cob	0.56	0.37	0.34
	(0.22)	(0.19)	(0.16)
feet/feed	0.67	0.44	0.31
	(0.16)	(0.18)	(0.24)
boot/booed	0.60	0.36	0.34
	(0.22)	(0.17)	(0.12)
pick/pig	0.37	0.26	0.25
	(0.23)	(0.12)	(0.14)
Mean slope, voiceless stimuli			
cop	0.49	0.27	0.24
	(0.28)	(0.31)	(0.17)
feet	0.45	0.25	0.20
	(0.24)	(0.22)	(0.18)
boot	0.35	0.17	0.20
	(0.31)	(0.21)	(0.17)
pick	0.22	0.17	0.23
	(0.18)	(0.17)	(0.32)
Mean slope, voiced stimuli			
cob	0.62	0.47	0.41
	(0.24)	(0.16)	(0.18)
feed	0.88	0.62	0.41
	(0.22)	(0.23)	(0.31)
booed	0.85	0.55	0.49
	(0.36)	(0.37)	(0.17)
pig	0.52	0.35	0.27
	(0.36)	(0.21)	(0.18)

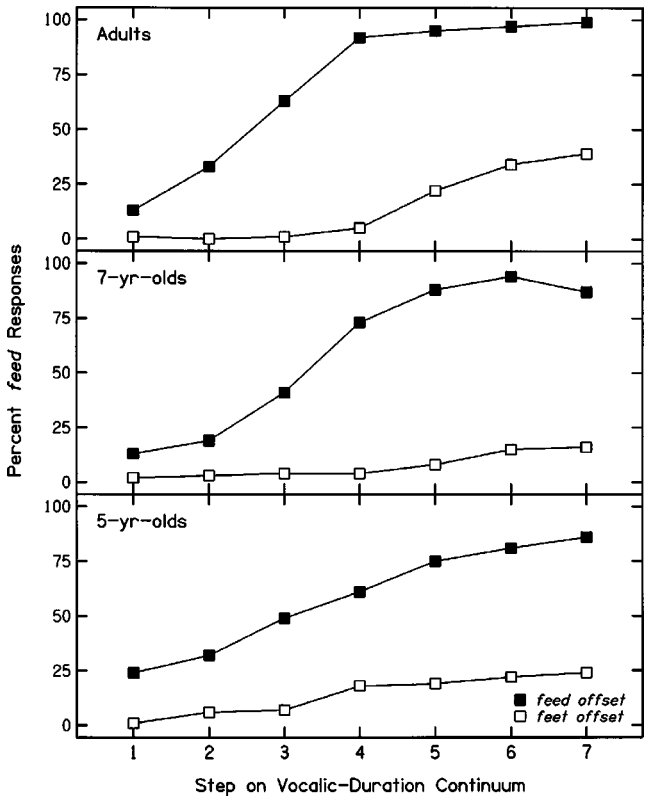


FIG. 6. Mean labeling functions for each age group for natural *feet/feed* stimuli, Experiment 2.

mean slopes, especially for slopes computed across functions and for slopes from stimuli with offset transitions appropriate for voiced final stops. There appears to be less of an age-related increase in slopes for stimuli with offset transitions appropriate for voiceless final stops. Because the analyses of slopes for the natural *buck/bug* in Experiment 1 resulted in a significant age effect for *bug*, but not for *buck*, one-way ANOVAs were computed for stimuli with syllable-offset transitions appropriate for voiced and voiceless final stops separately. As with *buck/bug*, in every case stimuli with syllable-offset transitions appropriate for voiced final stops showed significant, or close to significant, age effects, indicating that adults' functions were steeper than those of children: *booed*, $F(2,30)=3.79$, $p=0.034$; *cob*, $F(2,29)=3.06$ $p=0.062$; *feed*, $F(2,30)=8.89$, $p<0.001$; and *pig*, $F(2,34)=3.09$; $p=0.058$. Only one set of stimuli with syllable-offset transitions appropriate for voiceless final stops showed a significant (or close to significant) age effect: *feet*, $F(2,30)=4.01$, $p=0.029$. Labeling functions for these stimuli are presented in Fig. 6. Adults showed a slightly stronger inclination to label stimuli with *feet* offset transitions as *feed* at longer vocalic durations than children did. However, this is the only set of stimuli with syllable-offset transitions appropriate for voiceless final stops in which this trend was found, and adults still did not label more than 50% of these stimuli as *feed*, even at the longest durations.

C. Discussion

This second experiment was undertaken to see if children and adults would demonstrate the labeling pattern observed for stimuli created from natural tokens in Experiment 1: That is, would listeners of all ages weight formant offsets greatly and vocalic duration much less so? Four sets of stimuli were created from natural tokens varying in place of constriction for both the stop and the vowel. For all sets of stimuli, the partial *r*s revealed that adults, 7-year-olds, and 5-year-olds alike weighted formant offsets greatly, and vocalic duration much less so. For all sets of stimuli, adults, 7-year-olds, and 5-year-olds alike showed large separations between functions, depending on whether stimuli were created from words with voiceless or voiced final stops. Because earlier studies have attributed this effect to syllable-offset transitions, we conclude that these transitions are the most likely source of the effect here, as well. For all sets of stimuli, adults, 7-year-olds, and 5-year-olds failed to show particularly steep functions. However, as with the one set of stimuli created from natural tokens in Experiment 1, age-related differences in slope were observed for labeling functions of all stimuli created from words with voiced final stops. Only one of the functions for stimuli created from words with voiceless final stops showed this age effect. In summary, the general pattern of developmental increase in the weighting of vocalic duration was somewhat attenuated for these natural stimuli, but it certainly was not eradicated.

IV. GENERAL DISCUSSION

The original goal of this study was to examine whether children learning English as their first language would demonstrate the same weighting strategies for words ending in voiced and voiceless stops as non-native English-speaking

adults have demonstrated in experiments by others. Regarding weighting strategies for native English-speaking adults, it was assumed at the outset that the results of other investigators showing that vocalic duration is an important, if not primary, cue to voicing for syllable-final consonants would be easily replicated. Accordingly, the synthetic stimuli of Crowther and Mann (1992, 1994) were used, and a second set of synthetic stimuli created with the same general design. Vocalic durations and F1-offset transitions in these stimuli matched what is found in natural speech samples, but other aspects of acoustic structure were held constant across stimuli. Adults' results for both sets of stimuli did indeed match those obtained by Crowther and Mann from native English-speaking adults: Vocalic duration was weighted strongly, with much less weight given to F1-offset transitions. Children's results for vocalic duration fit predictions derived from studies of non-native English-speaking adults: children weighted vocalic duration less than the native English-speaking adults in this study did. At the same time, children's results for the F1-offset transitions fit predictions derived from other developmental studies: children relied on these dynamic signal components more than the adults did.

If this study had stopped there, longstanding views of human speech perception, and of how it develops, would have been perpetuated. However, modified natural stimuli were also included in this study. With those stimuli, adults and children alike mainly made decisions about final-stop voicing based on some property (or properties) of the natural stimuli other than vocalic duration. The most likely candidate for this other property is syllable-offset transitions—the dynamic components of the signal. This perceptual strategy had been predicted for children. It was adults who performed differently from expectations. Adults' apparently strong reliance on dynamic signal components in these phonetic judgments requires a reconsideration of basic principles regarding speech perception. To be sure, a few investigators in the past have suggested that vocalic duration is not a critical cue to adults' voicing decisions for final consonants (Hillenbrand *et al.*, 1984; Wardrip-Fruin, 1982). Nonetheless, vocalic duration has continued to be studied as a critical cue to syllable-final voicing, and so is presumably considered by many investigators to be just such a cue. In fact, Wardrip-Fruin herself investigated the development of children's abilities to use vocalic duration as a cue to final-consonant voicing just two years after concluding that it was not an adequate or necessary cue in adults' speech perception (Wardrip-Fruin and Peach, 1984). In that later study she concluded that adults use both syllable-offset transitions *and* vocalic duration in voicing decisions for final stops. Thus, the role of vocalic duration in these decisions remains equivocal, and any debate on this question reflects general controversy over theories of speech perception. For most of the history of human speech perception research, predominant theories have held that listeners make phonetic judgments by summing the information provided by several acoustic properties (e.g., Hodgson and Miller, 1992; Hogan and Rozsypal, 1980; Kewley-Port, Pisoni, and Studdert-Kennedy, 1983; Massaro and Oden, 1980). However, experiments showing that listeners can recover a phonetic representation from sine wave replicas of speech, which lack most of the static spectral properties of natural speech, have led to the suggestion that human speech perception may actually involve the tracking of dynamic changes in the speech wave form (e.g., Remez *et al.*, 1981). Of course, that conclusion was reached with stimuli that are highly unnatural, and therefore its ability to explain the perception of natural speech signals can be questioned.

Many experiments using modified natural stimuli or synthesized formant speech have demonstrated that adults use acoustic information that is not dynamic in their phonetic decisions. For example, adults use static spectral information (such as fricative noises) in decisions about sibilant place of constriction (e.g., Heinz and Stevens, 1961; Kunisaki and Fujisaki, 1977; Nittrouer and Miller, 1997), and use temporal information in decisions about voicing of initial stops and stops in clusters (e.g., Abramson and Lisker, 1967; Best, Morrongiello, and Robson, 1981; Nittrouer, Crowther, and Miller, 1998). At the same time, however, when phonetic information from static spectral and/or temporal properties is constrained by natural conditions, adults increase the weight they assign to dynamic signal components. For example, natural /f/ and /θ/ noises differ from each other spectrally far less than /s/ and /$\int$/ noises, and so adults weight the fricative-vowel formant transitions more in decisions of fricative place for /f/-vowel and /θ/-vowel sequences than for /s/-vowel and /$\int$/-vowel sequences (Harris, 1958; Nittrouer, 2002). The most novel finding reported in the current study is that adults weighted formant transitions at voicing offset greatly in these voicing decisions, even though a temporal property was readily available. This result suggests that experienced language users, as well as less-experienced children, may actually track the dynamic changes of the vocal tract, as suggested by investigators such as Remez *et al.* (1981). While this suggestion is not new, this study elegantly demonstrates the principle for natural signals.

There is, however, one caveat to this suggestion. The conclusion that adults use dynamic signal properties in the perception of natural speech was reached because of structural differences in the synthetic and natural stimuli used in these two experiments. Both sets of synthetic stimuli lacked formant transitions higher than F1, but all five sets of natural stimuli included those higher transitions. As a result of this difference between synthetic and natural stimuli, the conclusion was reached that it must have been the syllable-offset transitions in the natural stimuli that evoked the voiced or voiceless percepts for listeners. But, in fact, there were other attributes of the natural signals that could have accounted for the phonetic decisions. For example, F1 frequency is higher at syllable center in syllables with voiceless, rather than voiced, stops (Nittrouer *et al.*, submitted; Summers, 1987), and this property has some effect on voicing decisions for adults (Summers, 1988). However, the weight assigned to this property compared to other acoustic properties has not been thoroughly studied, and no study has been conducted of the effects of F1 at syllable center on children's voicing decisions. Also, intensity decays more rapidly at syllable offset for voiceless final stops (Hillenbrand *et al.*, 1984). The perceptual weight assigned to this property has not been in-

vestigated independently of spectral changes at syllable offset. Investigations specifically manipulating these other properties must be completed before it can be unequivocally concluded that formant transitions at the syllable's offset in natural tokens account largely for voicing decisions by adults and children.

In the same vein, the current study did not manipulate all potential cues to syllable-final stop voicing. In particular, release bursts and voicing during closure were removed from the natural stimuli, and were simply not present in the synthetic stimuli. Although Hillenbrand *et al.* (1984) suggested that the release burst and voicing during closure contribute little to voicing decisions for final stops, those investigators only examined adults' perception. Of course, noise spectra would be predicted to contribute mostly to decisions of constriction place, rather than of voicing. In addition, it has been consistently shown that children do not weight noise spectra greatly in place decisions for syllable-initial voiceless stops (e.g., Parnell and Amerman, 1978) or fricatives (e.g., Nittrouer, 1992; Nittrouer and Miller, 1997). Thus, there is little reason to suspect that release bursts would influence listeners' (especially children's) decisions about the voicing of syllable-final obstruents. Nonetheless, the influence specifically of burst releases, as well as of voicing during closure, on voicing decisions for final obstruents warrants examination.

In summary, this study was undertaken to examine whether young children's weighting of vocalic duration and syllable-offset transitions in decisions of voicing for final stops would match predictions derived from developmental and cross-linguistic studies. As predicted by other developmental studies, as well as by cross-linguistic studies with adults, children were found to weight vocalic duration less and syllable-offset transitions more than adults when synthetic stimuli were used. When edited, natural stimuli were presented, listeners of all ages based voicing decisions for final stops largely on some acoustic property (or properties) other than vocalic duration, although evidence of a developmental increase in the weighting of vocalic duration remained. Most likely, the other acoustic properties weighted heavily were the dynamic syllable-offset transitions.

ACKNOWLEDGMENTS

This work was supported by Research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health. The author thanks Kathi Bodily, Sandy Estee, Kathy Shapley, and Melanie Wilhelmsen for help with data collection. The comments of Van Summers regarding an earlier draft of this manuscript are gratefully acknowledged.

[1]Unfortunately, Ratner and Luberoff did not precisely define what they meant by the ''syllable-final consonant,'' and so we do not know whether that means only the voicing during closure and the release burst, or if syllable-offset transitions are included in their definition, as well. But it does not really matter because their study is important mainly because it informs us that vocalic duration is exaggerated in child-directed speech.

Abramson, A. S., and Lisker, L. (**1967**). ''Discriminability along the voicing continuum: Cross-language tests,'' *Proceedings of the 6th International Congress of Phonetic Sciences*, 569–573.

Best, C. T., Morrongiello, B., and Robson, R. (**1981**). ''Perceptual equivalence of acoustic cues in speech and nonspeech perception,'' Percept. Psychophys. **29**, 191–211.

Browman, C. P., and Goldstein, L. (**1990**). ''Gestural specification using dynamically-defined articulatory structures,'' J. Phonetics **18**, 299–320.

Chen, M. (**1970**). ''Vowel length variation as a function of the voicing of the consonant environment,'' Phonetica **22**, 129–159.

Crowther, C. S., and Mann, V. (**1992**). ''Native language factors affecting use of vocalic cues to final consonant voicing in English,'' J. Acoust. Soc. Am. **92**, 711–722.

Crowther, C. S., and Mann, V. (**1994**). ''Use of vocalic cues to consonant voicing and native language background: The influence of experimental design,'' Percept. Psychophys. **55**, 513–525.

Denes, P. (**1955**). ''Effect of duration on the perception of voicing,'' J. Acoust. Soc. Am. **27**, 761–764.

Finney, D. J. (**1964**). *Probit Analysis* (Cambridge University Press, Cambridge, England).

Fischer, R. M., and Ohde, R. N. (**1990**). ''Spectral and duration properties of front vowels as cues to final stop-consonant voicing,'' J. Acoust. Soc. Am. **88**, 1250–1259.

Flege, J. E., and Port, R. (**1981**). ''Cross-language phonetic interference: Arabic to English,'' Lang Speech **24**, 125–146.

Flege, J. E., and Wang, C. (**1989**). ''Native-language phonotactic constraints affect how well Chinese subjects perceive the world-final English /t/-/d/ contrast,'' J. Phonetics **17**, 299–315.

Goldman, R., and Fristoe, M. (**1986**). *Goldman Fristoe Test of Articulation* (American Guidance Service, Circle Pinces, MN).

Gracco, V. L. (**1994**). ''Some organizational characteristics of speech movement control,'' J. Speech Hear. Res. **37**, 4–27.

Greenlee, M. (**1980**). ''Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception,'' J. Child Lang **7**, 459–468.

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

Heinz, J. M., and Stevens, K. N. (**1961**). ''On the properties of voiceless fricative consonants,'' J. Acoust. Soc. Am. **33**, 589–593.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (**1984**). ''Perception of the voiced–voiceless contrast in syllable-final stops,'' J. Acoust. Soc. Am. **76**, 18–26.

Hodgson, P., and Miller, J. L. (**1992**). ''Phonetic category structure depends on multipule acoustic properties: Evidence for within-category trading relations,'' J. Acoust. Soc. Am. **92**, 2464.

Hogan, J. T., and Rozsypal, A. J. (**1980**). ''Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant,'' J. Acoust. Soc. Am. **67**, 1764–1771.

House, A. S., and Fairbanks, G. (**1953**). ''The influence of consonant environment upon the secondary acoustical characteristics of vowels,'' J. Acoust. Soc. Am. **25**, 105–113.

Jastak, S., and Wilkinson, G. S. (**1984**). *The Wide Range Achievement Test—Revised* (Jastak Associates, Wilmington, DE).

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (The MIT Press, Cambridge, MA).

Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (**1983**). ''Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants,'' J. Acoust. Soc. Am. **73**, 1779–1793.

Klatt, D. H. (**1976**). ''Linguistic uses of segmental duration in English: Acoustic and perceptual evidence,'' J. Acoust. Soc. Am. **59**, 1208–1221.

Krause, S. E. (**1982**). ''Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults,'' J. Acoust. Soc. Am. **71**, 990–995.

Kunisaki, O., and Fujisaki, H. (**1977**). ''On the influence of context upon perception of voiceless fricative consonants,'' Annual Bulletin of the Research Institute for Logopedics and Phoniatrics **11**, 85–91.

Lehman, M. E., and Sharf, D. J. (**1989**). ''Perception/production relationships in the development of the vowel duration cue to final consonant voicing,'' J. Speech Hear. Res. **32**, 803–815.

Malécot, A. (**1970**). ''The lenis-fortis opposition: Its physiological parameters,'' J. Acoust. Soc. Am. **47**, 1588–1592.

Massaro, D. W., and Oden, G. C. (**1980**). ''Evaluation and integration of acoustic features in speech perception,'' J. Acoust. Soc. Am. **67**, 996–1013.

Miranda, S., and Strange, W. (**1989**). ''The role of spectral, temporal, and dynamic cues in the perception on English vowels by native and non-native speakers,'' J. Acoust. Soc. Am. **85**, S53.
Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). ''Trading relations in the perception of speech by 5-year-old children,'' J. Exp. Child Psychol. **37**, 231–250.
Nittrouer, S. (**1992**). ''Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries,'' J. Phonetics **20**, 351–382.
Nittrouer, S., Manning, C., and Meyer, G. (**1993**). ''The perceptual weighting of acoustic cues changes with linguistic experience,'' J. Acoust. Soc. Am. **94**, S1865.
Nittrouer, S., and Miller, M. E. (**1997**). ''Developmental weighting shifts for noise components of fricative-vowel syllables,'' J. Acoust. Soc. Am. **102**, 572–580.
Nittrouer, S., Crowther, C. S., and Miller, M. E. (**1998**). ''The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults,'' Percept. Psychophys. **60**, 51–64.
Nittrouer, S. (**2002**). ''Learning to perceive speech: how fricative perception changes, and how it stays the same,'' J. Acoust. Soc. Am. **112**, 711–719.
Nittrouer, S., Estee, S, Lowenstein, J. H., and Smith, J. (submitted). ''The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops,'' J. Acoust. Soc. Am.
O'Kane, D. (**1978**). ''Manner of vowel termination as a perceptual cue to the voicing status of postvocalic stop consonants,'' J. Phonetics **6**, 311–318.
Parnell, M. M., and Amerman, J. D. (**1978**). ''Maturational influences on perception of coarticulatory effects,'' J. Speech Hear. Res. **21**, 682–701.
Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.
Raphael, L. J. (**1972**). ''Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English,'' J. Acoust. Soc. Am. **51**, 1296–1303.
Raphael, L. J., Dorman, M. F., Freeman, F., and Tobin, C. (**1975**). ''Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies,'' J. Speech Hear. Res. **18**, 389–400.
Raphael, L. J., Dorman, M. F., and Liberman, A. M. (**1980**). ''On defining the vowel duration that cues voicing in final position,'' Lang Speech **23**, 297–307.
Ratner, N. B., and Luberoff, A. (**1984**). ''Cues to post-vocalic voicing in mother-child speech,'' J. Phonetics **12**, 285–289.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). ''Speech perception without traditional speech cues,'' Science **212**, 947–949.
Strange, W. (**1989**). ''Evolving theories of vowel perception,'' J. Acoust. Soc. Am. **85**, 2081–2087.
Summers, W. V. (**1987**). ''Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses,'' J. Acoust. Soc. Am. **82**, 847–863.
Summers, W. V. (**1988**). ''F1 structure provides information for final-consonant voicing,'' J. Acoust. Soc. Am. **84**, 485–492.
Sussman, H. M., MacNeilage, P. F., and Hanson, R. J. (**1973**). ''Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations,'' J. Speech Hear. Res. **16**, 397–420.
Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. A. (**1998**). ''Frequency-weighting functions for broadband speech as estimated by a correlational method,'' J. Acoust. Soc. Am. **104**, 1580–1585.
Wardrip-Fruin, C. (**1982**). ''On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants,'' J. Acoust. Soc. Am. **71**, 187–195.
Wardrip-Fruin, C., and Peach, S. (**1984**). ''Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants,'' Lang Speech **27**, 367–379.

Acoustic and perceptual similarity of North German and American English vowels

Winifred Strange[a)]
Ph.D. Program in Speech and Hearing Sciences, The City University of New York—Graduate School and University Center, 365 Fifth Avenue, New York, New York, 10016-4309

Ocke-Schwen Bohn
English Department, Aarhus University, DK-8000 Aarhus C, Denmark

Sonja A. Trent and Kanae Nishi
Department of Psychology, University of South Florida, 4202 Fowler Avenue, Tampa, Florida 33620

(Received 30 September 2003; revised 26 January 2004; accepted 26 January 2004)

Current theories of cross-language speech perception claim that patterns of perceptual assimilation of non-native segments to native categories predict relative difficulties in learning to perceive (and produce) non-native phones. Cross-language spectral similarity of North German (NG) and American English (AE) vowels produced in isolated hVC(a) (di)syllables (study 1) and in hVC syllables embedded in a short sentence (study 2) was determined by discriminant analyses, to examine the extent to which acoustic similarity was predictive of perceptual similarity patterns. The perceptual assimilation of NG vowels to native AE vowel categories by AE listeners with no German language experience was then assessed directly. Both studies showed that acoustic similarity of AE and NG vowels did not always predict perceptual similarity, especially for ''new'' NG front rounded vowels and for ''similar'' NG front and back mid and mid-low vowels. Both acoustic and perceptual similarity of NG and AE vowels varied as a function of the prosodic context, although vowel duration differences did not affect perceptual assimilation patterns. When duration and spectral similarity were in conflict, AE listeners assimilated vowels on the basis of spectral similarity in both prosodic contexts.
[DOI: 10.1121/1.1687832]

PACS numbers: 43.71.Hw, 43.70.Kv, 43.71.Es, 43.70.Fq [RD] Pages: 1791–1807

I. INTRODUCTION

In recent years, there has been increased interest in cross-language comparisons of phonetic categories, growing out of the well-documented problems that adult second language (L2) learners have in acquiring a new phonological system. In his Speech Learning Model (SLM), Flege (1995) claims that continuing problems with ''accented'' production of phonetic segments can be attributed in large part to L2 learners' representation of the L2 segments as equivalent to ''similar'' segments in the native language (L1). That is, if the L2 phones are sufficiently similar to L1 phones, they will be perceptually assimilated to those native categories, with the result that both L1 and L2 segments are produced differently from native monolingual speakers' utterances. If, however, L2 phones are sufficiently dissimilar from any L1 category (i.e., ''new''), the L2 learner will (eventually) establish distinct L1 and L2 phonetic categories, and production of the L2 segments will become more native-like.

In her Perceptual Assimilation Model (PAM), Best (1994, 1995) also invokes the concept of cross-language phonetic similarity to predict the relative difficulties that listeners will have in perceptual differentiation of non-native segmental contrasts. She describes several patterns of perceptual assimilation of L2 segments to L1 phonological categories, which are determined by the *perceived* phonetic similarity of L1 and L2 segments. Two L2 segments which are judged as equally ''good'' instances of a single L1 category (Single-Category pattern) will be most difficult to differentiate, while two L2 segments that are assimilated to two different L1 categories (Two-Category pattern) will be very easy to discriminate. In addition, contrasting L2 segments that differ in their judged goodness as instances of a single L1 category (Category-Goodness pattern) will yield intermediate levels of perceptual difficulty. Finally, if an L2 segment is sufficiently dissimilar from any L1 category, it may be considered an ''uncategorizable'' speech sound. When paired with another L2 phone that is phonetically similar enough to be categorized as an instance of an L1 category (i.e., it is categorizable), the two phones (Uncategorizable versus Categorizable) will be relatively easily discriminated.

According to both these models, then, the perceived similarity of segments in L1 and L2 is an important determiner of the pattern of initial perceptual problems and persistent learning difficulties adult L2 learners have in mastering the L2 phonological system. It is critical, therefore, that cross-language perceptual similarity be established, *independent of identification or discrimination performance*, in order to predict L2 learning difficulties more accurately. In the work of Flege and Best, as well as other researchers in the field, perceptual similarity has been inferred from (1) a comparison of impressionistic descriptions of the phonetic seg-

[a)]Electronic mail: strangepin@aol.com

ments (e.g., Best and Strange, 1992), (2) transcriptions or reports from listeners about similarities between native and non-native segments (e.g., Best, Faber, and Levitt, 1996) or (3) cross-language comparisons of the acoustic structure of the non-native segments (e.g., Flege, 1987; Bohn and Flege, 1990). In more recent studies, perceptual similarity has been assessed directly, using a perceptual assimilation task in which listeners are presented non-native segments and asked to categorize them with respect to which native category they are most similar and to rate their ''category goodness'' as exemplars of the chosen categories (e.g., Bohn and Flege, 1990; Guion, Flege, Akahane-Yamada, and Pruitt, 2000; Strange, Akahane-Yamada, Kubo, Trent, Nishi, and Jenkins, 1998; Strange, Akahane-Yamada, Kubo, Trent, and Nishi, 2001).

In the study reported here, the phonetic similarity of North German (NG) and American English (AE) vowels was investigated. Results of acoustical analysis of a corpus of NG vowels were compared with data from a similar corpus of AE vowels. Perceptual similarity was assessed directly using the perceptual assimilation task in which native speakers of AE with no previous experience with German were asked to categorize and rate NG vowels as exemplars of AE vowel categories. Acoustic and perceptual similarity patterns were then compared to determine the extent to which perceptual similarity of sets of NG vowels were predictable from their context-specific similarity in spectral and temporal structure to AE vowels. Of additional interest was the extent to which acoustic and perceptual similarity varied as a function of prosodic context (citation-form lists vs sentences spoken at continuous speech rates).

The German vowel inventory consists of 14 distinctive monophthongs that include 7 tense–lax (long–short or close–open) pairs: the front unrounded high [iː-ɪ] and mid [eː-ɛ] vowels, and back rounded high [uː-ʊ] and mid [oː-ɔː] vowels, the front rounded high [yː-ʏ] and mid [ø-œ] vowels and the low vowels [ɑː-a]. In North German (NG) dialects, high and mid tense versus lax vowel pairs are differentiated phonetically by both vocalic duration and tongue/jaw position, with the lax vowel of each pair lower (more open) and more centralized than the tense vowel; the low vowels are differentiated almost entirely by duration.[1] The standard American English (AE) inventory includes nine so-called monophthongs and two diphthongized nonrhotic vowels: a front-unrounded series [iː, ɪ, e͡ɪ, ɛ, æː], a back-rounded series [uː, ʊ, o͡ʊ, ɔː], and the low and mid-low back vowels [ɑː, ʌ]. While AE vowels differ phonetically in intrinsic duration, as indicated in their transcription above ([ː]=long), length is considered a secondary or redundant feature of vowel height or closeness (Peterson and Lehiste, 1960; Crystal and House, 1988a, 1988b; Hillenbrand, Clark, and Houde, 2000). Thus, North German and American English are similar in that vowels are distinguished phonetically by five levels of vowel height/openness and by backness. The vowel inventories differ in that lip rounding is redundant with backness in English, whereas it distinguishes unrounded and rounded front vowels in German. Finally, in North German, long and short vowels are realized in careful speech with larger relative duration differences than are long and short vowels in American English (Strange and Bohn, 1998). In both languages, vocalic duration also varies with vowel height such that high vowels are shorter than low vowels, with mid vowels intermediate in duration (Strange and Bohn, 1998; Strange *et al.*, 1998).

According to traditional articulatory-phonetic descriptions, then, NG front rounded vowels do not have counterparts in the AE vowel inventory (they are ''new'' in Flege's terminology). NG [a] and AE[ʌ] can be considered ''similar'' in tongue height and position and are both short monophthongs, while NG[ɔ] and AE [ɔː] are ''similar'' in tongue height, but differ in intrinsic duration. Finally, NG [eː, oː] are monophthongal long vowels, while their AE counterparts are diphthongized (and long) in most phonetic and prosodic contexts.

It is well known that speakers of languages that do not contrast front rounded vowels with either front unrounded or back rounded vowels have difficulty learning to perceive and produce these vowels. However, previous research has produced conflicting results with respect to native AE speakers' ability to perceptually differentiate such contrasts in German and French. Polka (1995) reported native-like categorial discrimination by AE listeners of NG back versus front rounded high tense/u–y/, but less than native-like discrimination of the mid-high lax pair/ʊ-ʏ/; vowels were produced and presented in citation-form dVt syllables. In contrast, Gottfried (1984), and later Levy and Strange (2002), reported that [u–y], as spoken by Parisian French speakers, was very difficult for native English speakers to perceptually differentiate, even after years of experience speaking French (see also Rochet, 1995). Best *et al.* (1996) reported relatively poor discrimination by inexperienced English listeners of the Norwegian high front unrounded versus outrounded [i–y] in citation-form bV syllables, while Levy and Strange's subjects did relatively well on French unrounded versus (out)-rounded [i–y] in dVt syllables. Finally, both Gottfried and Levy and Strange reported significant differences in perceptual performance as a function of the syllabic and consonantal context in which the French front rounded vowels were produced and presented. Levy and Strange reported that inexperienced listeners had difficulty discriminating [i–y] in bVp syllables but not in dVt syllables, whereas the opposite pattern was true for the [u–y] contrast. (Contextual effects on perceptual assimilation will be addressed in a forthcoming paper.)

These discrepant results within and across studies and languages clearly indicate that the prediction of difficulty in the perception of non-native vowels by adult L2 learners, based on contrastive analyses at the level of abstract phonological features or impressionistic descriptions of phonetic segments, is doomed to failure (cf. Kohler, 1981). Across languages, vowels that are transcribed as ''the same'' can be very different acoustically (e.g., French front rounded [y] is more front acoustically than is German front rounded [y] [Strange *et al.*, 2002]). Thus, it is necessary to establish cross-language similarities and differences of vowels at a level more closely associated with their phonetic realization in the particular languages under comparison, if we are to

gain a better understanding of the perceptual problems facing second-language learners. Furthermore, the extent to which *perceived* phonetic similarity may be predicted from a comparison of cross-language context-specific acoustic similarity must be established. For instance, Flege and Hillenbrand (1984) reported that AE learners of French *produced* French [tu] more ''authentically'' than [ty], as judged by native French listeners. However, acoustic analysis showed that their [y] tokens were more similar to native French tokens than were [u] tokens. They inferred from these data that for AE speakers, French [u] was perceptually similar to its AE counterpart, while French [y] was perceived as a ''new'' vowel (i.e., was perceptually dissimilar from any AE vowel). However, no direct assessment of perceptual similarity was made, and no comparison of native French front and back rounded vowels with AE back and front vowels was included.

In the present study, a primary goal was the systematic evaluation of the acoustic and perceptual similarity of NG front rounded vowels relative to AE front unrounded and back rounded vowels. However, rather than investigating only these four NG vowels, both acoustical and perceptual comparisons included the entire inventory of 14 NG monophthongs in comparison with the 11 AE vowels. We reasoned that since vowels are perceived with respect to their relative positions in a speaker's ''vowel space,'' it was important to assess cross-language acoustic and perceptual similarity of the complete vowel inventories of both languages. In acoustic analyses, questions about the relative spacing of vowels in the vowel space could be more accurately determined across languages, and issues about speaker normalization could be addressed. For the investigation of the perceptual assimilation of NG vowels by AE listeners, presenting the entire vowel inventory of each speaker would allow the listeners to ''normalize'' the speaker's utterances. In addition, questions about the perception by AE listeners of ''similar'' and ''identical'' NG vowels could be answered.

In this study, cross-language acoustic and perceptual similarity of NG and AE vowels was determined for productions that were relatively uninfluenced by the surrounding consonantal context; i.e., with vowels considered to reflect ''canonical vowel targets'' for each language. The NG vowels were produced and presented in hVp syllables spoken in lists (citation-form syllables) (study 1) and in the same syllables produced and presented in short carrier sentences (study 2).[2] Because short vowels do not occur in open syllables in English or in German, the CVC structure was chosen. The initial /h/ and voiceless labial stop /p/ are phonologically appropriate in the word-initial and word-final position, respectively, in both German and English and pronounced similarly in both languages. Thus, it was assumed that judgements of cross-language perceptual similarity of these syllables would be attributable to the vowels themselves, with little effect of the surrounding consonants on either their production or their perception.

Previous studies of the perception of German vowels (Polka, 1995; Polka and Bohn, 1996) and French vowels (Gottfried, 1984; Rochet, 1995) by AE listeners using materials in which the vowels were produced and presented in the coronal consonant context lead to the prediction that AE listeners will assimilate NG front rounded vowels to back rounded AE vowels. However, in many dialects of American English, back rounded high and mid-high vowels in coronal contexts have higher F2 frequencies than in noncoronal contexts; that is, they are allophonically ''fronted'' in coarticulation with coronal consonants (Stevens and House, 1963; Strange, 1989; Hillenbrand *et al.*, 2001). Thus, NG front rounded vowels may be perceived as more similar to AE so-called back rounded vowels in alveolar contexts than when preceded or followed by noncoronal consonants.

Our second goal in this study was to examine the acoustic and perceptual similarity of the ''similar'' mid and mid-low vowels, NG [eː, ɛ, oː, ɔ] and AE [e͡ɪ, ɛ, o͡ʊ, ɔː]. Gottfried (1984) reported that French [i–e] and [e–ɛ] were difficult for native English speakers to discriminate when presented in isolation and in a tVt context (Gottfried, 1984). Previous acoustical studies of AE and NG vowels (Hillenbrand *et al.*, 2001; Strange and Bohn, 1998; see also Steinlen and Bohn, 1999; Steinlen, 2002, for Southern British English and German vowels) suggest that there are differences in the realization of mid and mid-low vowels (i.e., their location in F1/F2 vowel space relative to high and low vowels), in addition to the differences in diphthongization (for [eː, oː]) and duration (for [ɔ]). If acoustical analyses revealed that NG mid long and mid-low short vowels were produced ''higher'' in vowel space relative to their AE counterparts, we might expect that AE listeners would assimilate mid and mid low vowels to high and mid-high AE categories, respectively, at least some of the time.

Our final goal in this study was to investigate the relative contribution of vocalic duration to the perceptual similarity of NG vowels for AE listeners. Bohn (1995) has claimed that when non-native vowels have no spectral counterpart in the native language, listeners may attend more to temporal cues to distinguish them, even when vowel length is not phonologically contrastive in their native vowel system. In a previous study of Japanese listeners' perceptual assimilation of AE vowels, Strange *et al.* (1998) reported that temporal assimilation patterns varied systematically with prosodic context. Japanese listeners heard long versus short AE vowels as more similar to long (two-mora) versus short (one-mora) Japanese vowels, respectively, when hVb syllables were produced and presented in sentence context, despite the fact that the acoustic durations of long and short AE vowels were very similar in both sets of materials (and the differences were smaller than for Japanese vowels). It was concluded that the rhythmic structure of the sentences allowed Japanese listeners to better ''interpret'' intrinsic duration differences of the AE vowels in relating them to their own vowel inventory. In the present study, it was hypothesized that long and short NG vowels would be perceptually assimilated to long and short AE vowels more consistently when presented in the sentence context than in the isolated syllable context.

II. ACOUSTIC AND PERCEPTUAL SIMILARITY OF NG AND AE VOWELS PRODUCED IN CITATION-FORM SYLLABLES

Study 1 examined a corpus of vowels produced by four adult male native speakers of NG in citation-form hVp syllables. An acoustical analysis was performed to provide descriptive data about average formant frequencies and relative durations of these ''canonical'' NG vowels. Discriminant analyses were then used to establish the extent to which the 14 vowels were differentiated spectrally (using formant frequencies as input parameters) and the contribution of relative duration to acoustic differentiation (adding duration as an input parameter). Acoustic similarity to an existing corpus of AE vowels produced in citation-form lists of hVba disyllables was then established using discriminant analyses in which AE vowels served as the input corpus and NG vowels as the test corpus. We were especially interested in the acoustic similarity of NG front rounded vowels to AE front and back vowels. Although the consonantal and syllable context differed somewhat from the NG corpus, it was presumed that the influence of final consonants (/p/ *vs* /b/) and syllable structure (monosyllable versus disyllable) across corpora on the ''target'' formant frequencies of the vowels would be comparable (and minimal), and that within-language *relative* vowel durations would be interpretable, despite the possible effects of final consonant voicing and syllable structure differences on absolute vocalic durations. On the basis of these cross-language acoustical comparisons, predictions were made with respect to how native AE listeners with no German-language experience would perceptually assimilate NG vowels to their native AE vowel categories.

A. Acoustic similarity of NG and AE vowels

1. Speakers and stimulus materials

Four male speakers of North German produced the NG stimulus corpus; all four had lived in northern Germany (Kiel, West Holstein, Hamburg, Lubeck) all their lives and had not resided in an English speaking country for more than a few months. They were enrolled as students at Kiel University at the time of recording. The speakers ranged in age from 25 to 29 years old, and (as is typical of college students in Germany) all had studied English for at least nine years. They had also studied French in high school (2–4 years).

Recordings were made in a quiet room at Kiel University by the second author (a native speaker of German), with the first author in attendance, using a Sony ECM-939LT (stereo electret condenser microphone) connected to a Marantz PMD 420 cassette tape recorder (as a preamplifier) fed to a Sony DTC-P7 DAT tape deck (48 KHz sampling rate). The 14 nonsense syllables were printed individually on index cards in German orthography with an identifying number printed before each syllable. The speakers were given instructions and practice as needed to produce the number and the syllable, with a pause between, and falling intonation on the syllable. During practice, clear errors in vowel production were corrected, but minor dialectal variations were not commented on. Each speaker recorded three randomizations of 15 stimuli in which the last stimulus was the same as the first. This last utterance was discarded to eliminate any list-final differences in production. The first randomization was used primarily for practice and was not included in the final stimulus corpus unless a token in randomization 2 or 3 was deemed less acceptable as an exemplar of that vowel. After recording, a phonetically trained NG speaker listened to all of the tokens produced by each of the four speakers. The selected exemplars were all considered good tokens of the NG vowels by this listener. The final stimulus corpus included 28 stimuli (2 tokens$\times$14 vowels) for each of four speakers for a total of 112 stimuli. The stimuli were digitally transferred from a DAT recorder (Tascam DA-30 MK II) to a Power Macintosh 8100/100 computer via an Audiomedia II digital I/O sound card. Sound files were then downsampled to 22.01 kHz with 16-bit amplitude resolution and the identifying numbers were deleted from each file.

Four adult male native speakers of AE produced multiple instances of the 11 AE vowels in citation-form hVba context. (This corpus was the same as that used in a study of perceptual similarity of Japanese and AE vowels, Strange *et al.*, 1998). Three of the speakers were residing at the University of South Florida at the time of testing; the remaining speaker was residing in Japan at the time of recording. For the three Florida participants, stimuli were recorded in an IAC chamber using a dynamic microphone (Panasonic WM-1325) fed to a DAT recorder (SONY TCD-D10). The speaker residing in Japan was recorded in an anechoic chamber, using a condenser microphone (SONY ECM-77) fed to a DAT recorder (SONY PCM-2500A,B). Speakers produced the 11 AE vowels in lists of hVba disyllables, each preceded by an identifying number. Each speaker produced 4 randomizations of 12 syllables, where the 12th utterance was identical to the first; this utterance was not used. Utterances from randomizations 2, 3, and 4 were used as stimulus tokens. Thus, for each speaker, there were 33 stimuli (3 instances $\times$11 vowels). The stimuli were recorded at 48 kHz, digitally transferred to a Power Macintosh 8100/100 computer via an Audiomedia II digital I/O soundcard, downsampled to 22.01 kHz with 16 bit amplitude resolution, and then identifying numbers were deleted from the files.

2. Acoustic analysis

Temporal and spectral measurements were performed using SoundScope/16 1.44 (ppc)™ speech analysis software designed by GW Instruments, Inc. (© Copyright 1992 GWI, Somerville, MA 02143). Temporal measurements were made from an inspection of time-synchronized wave form and wideband spectrogram displays (300 Hz filter, 6 dB preemphasis) with LPC formant tracks superimposed. Vocalic duration was defined as beginning at the onset of voicing after the /h/ (the beginning of the first pitch period) and terminating at the end of the final pitch period before consonant closure (end of upper formant energy in the case of AE /b/, silence in the case of NG /p/). Spectral measures were derived from analysis windows at three relative temporal locations—25%, 50%, and 75%—within the vocalic nucleus; only data from the 50% windows (temporal midpoint of the vocalic nuclei) are reported here. LPC spectra (28 coefficients) were computed using a 25 ms Hamming window cen-

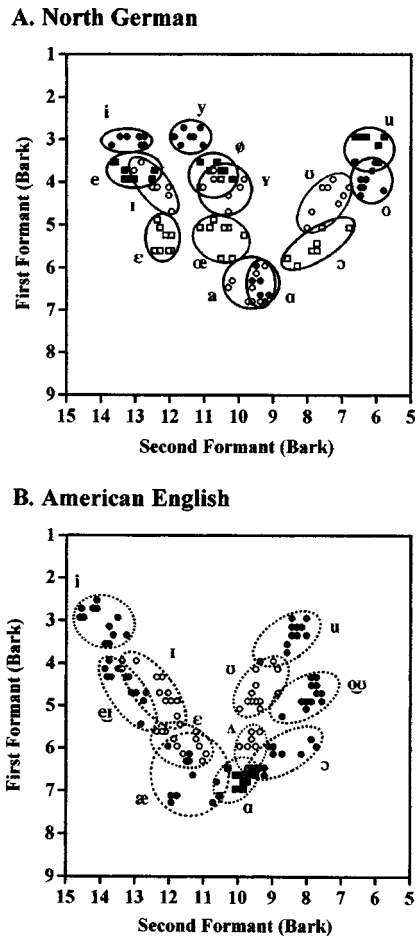


FIG. 1. Formant 1/formant 2 (Bark) plots of North German (A) and American English (B) vowel corpora in Study 1. Citation-form (di)syllables.

tered around the 50% location and values for the first three spectral peaks were tabulated. When LPC peak values reflected spurious peaks or missed formants (based on an inspection of the spectrogram and formant history) manual placement of the marker on wideband FFT spectra or, in rare cases, narrow band FFT were used to estimate formant values.

3. Results

Figure 1(A) displays the 112 NG stimuli as points in F1/F2 Bark space; ellipses surround all 8 instances of each vowel (2 instances×4 speakers). Figure 1(B) displays the 132 stimuli of the AE corpus for comparison. In both these plots, long vowels are depicted by closed symbols; short vowels are depicted by open symbols. Table I presents the frequencies of the first three formants and the vocalic durations of each vowel, averaged over all speakers' tokens (8 for NG; 12 for AE). German vowels are shown on the left; AE vowels on the right. Identical or similar vowels across languages are shown in the same rows to facilitate comparison.

As Fig. 1 shows, the NG front rounded vowels appear to be spectrally more similar to NG front unrounded vowels, than to NG back rounded vowels. It is also apparent that the so-called mid long vowels [eː, øː, oː] were realized with relatively low F1 values, i.e., they were relatively "high" (and close to high long vowels) in the acoustic vowel space. In fact, F1 values for short so-called mid-high vowels [ɪ, ʏ, ʊ] were actually higher on average than the mid long vowels, although categories overlapped somewhat in formant values. Finally, NG long and short low vowels overlapped spectrally.

The AE vowel space showed different patterns of spectral similarity. Mid long and high–mid short vowels [e͡ɪ, o͡ʊ, ɪ, ʊ] overlapped in F1 values, but tended to be differentiated in F2 values. Relative to NG vowels, AE [e͡ɪ, o͡ʊ] appeared to be spectrally more differentiated from [iː, uː]. Other spectrally similar long and short vowels overlapped in vowel space [æː-ɛ, ɔː-ʌ, ɑː-ʌ]. Finally, [ɑː-ɔː] long vowels showed spectral overlap typical of many AE dialects.

In comparing the relative duration differences of NG and AE vowels shown in Table I, the short NG vowels were slightly shorter, on average, than the AE short vowels (NG = 80 ms; AE = 90 ms), whereas the long NG vowels were considerably longer than the AE vowels (153 ms *vs* 115 ms). Thus, as expected from phonological descriptions of German and English, the ratio of long to short vowels for NG (1.9) was substantially greater than the long/short ratio for AE vowels (1.3) for these citation-form utterances.

TABLE I. Average formant frequencies and durations of North German (NG) and American English (AE) vowels in citation-form syllables—study 1. NG long/short vowel ratio = 1.9. AE long/short vowel ratio = 1.3.

NG	F1	F2	F3	Duration	AE	F1	F2	F3	Duration
iː	309	1986	2960	137	iː	312	2307	2917	100
ɪ	428	1800	2460	70	ɪ	486	1785	2573	86
eː	393	2010	2651	161	e͡ɪ	472	2062	2660	122
ɛ	573	1738	2454	83	ɛ	633	1588	2553	91
					æː	730	1568	2519	123
yː	301	1569	1934	142					
ʏ	428	1340	2137	85					
øː	393	1388	2045	165					
œ	559	1353	2277	93					
uː	320	689	1978	129	uː	348	995	2374	104
ʊ	457	834	2368	71	ʊ	489	1148	2472	93
oː	415	683	2277	165	o͡ʊ	500	909	2643	112
ɔ	589	893	2497	76	ɔː	678	1062	2678	132
ɑː	718	1146	2508	172	ɑː	753	1250	2596	109
a	710	1200	2409	82	ʌ	635	1189	2619	89

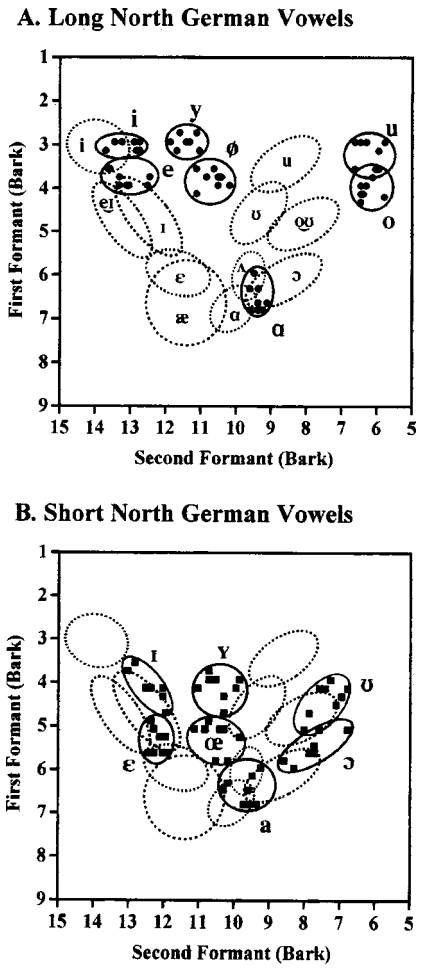


FIG. 2. North German long vowels (A) and short vowels (B) superimposed on ellipses of the 11 AE vowel categories. Study 1.

With respect to cross-language comparisons of spectral patterns, the averages for NG and AE vowels demonstrate considerable cross-language variability in the acoustic similarity of so-called identical and similar vowels. To illustrate this comparison, Fig. 2 displays the distributions, in F1/F2 Bark[3] space, of each NG vowel (filled symbols circumscribed by solid ellipses), superimposed on ellipses surrounding the 12 instances of each of the AE vowels (dashed lines). For clarity, the 7 long NG vowels are shown in Fig. 2(A), while the 7 short vowels are superimposed on the same AE data in Fig 2(B), and the individual AE tokens are not plotted. As the figure and Table I show, the point vowels [iː, ɑː] were quite similar for the two corpora, suggesting that differences in other vowels were probably not due to overall differences in the vocal tract size and shape of these speakers.

In order to quantify within- and cross-language similarities in acoustic structure [including F3 values not shown in Figs. 2(A) and 2(B)], a series of linear discriminant analyses (Klecka, 1980) were performed. First, separate analyses of the NG corpus and the AE corpus were computed to quantify spectral differentiation. For each analysis, F1/F2/F3 frequency values (in Barks) served as input parameters to establish centers of gravity in formant space for the 11 AE vowel categories and 14 NG vowel categories, respectively. A second set of discriminant analyses were performed, in which the vocalic duration was added as an input parameter; again, separate analyses were performed for the NG and AE corpus. Percentages of the ''correct'' classification of the stimuli as the intended vowels, based on optimal parameter weightings established for the input set, were then computed.

For the NG corpus, the overall classification on the basis of spectral parameters alone (79% correct) reflected the considerable overlap of so-called mid long [eː, øː, oː] with so-called mid-high short vowels [ɪ, ʏ, ʊ] and of long and short low vowels [ɑː, a], as can be seen in Fig. 1(A).

Correct classification ranged from 38% to 75% on these 8 vowels; the other 6 vowels were differentiated spectrally with 100% accuracy. When the duration was included as an additional input variable, the overall correct classification rose markedly (overall=93%; range across 14 vowels =63% to 100%); at least 7 of 8 tokens of all vowels except [oː] were correctly classified. The remaining errors included confusions of adjacent ''height'' categories [uː-oː, ʏ-œ, ʊ-ɔ] and adjacent short/long vowels [eː-ɪ, ɑː-a, oː-ʊ].

Discriminant analysis results for the 132-stimulus AE corpus, based on spectral parameters alone, yielded 86% overall correct classification as the speakers' intended vowels (range across 11 vowels=50%–100%). As Fig. 1(B) suggests, most confusions were between spectrally adjacent long/short vowel pairs [e͡ɪ-ɪ, æː-ɛ, ɑː-ʌ, uː or o͡ʊ-ʊ], and between [ɑː-ɔː]. When vocalic duration was added as an additional input parameter, the correct classification improved (overall=92%, range across the 11 vowels=83%–100%), with at least 11 of 12 instances of all vowels except [æː, ɔː] classified correctly. These within-language analyses support the conclusion that vowel duration is somewhat more important in differentiating spectrally-similar vowels in NG than in AE. For both corpora, a few acoustic ambiguities remain due to speaker differences.

Of greatest interest was the *cross-language* discriminant analysis to establish the spectral similarity of AE and NG vowels. (Because the syllable structure and absolute durations differed for NG and AE corpora, the duration was not used in cross-language discriminant analysis.) The 112 stimuli of the NG corpus served as the test corpus and were classified with respect to AE vowel centers of gravity established for the 132-stimulus AE input corpus. This was accomplished in two steps. First, the 10 so-called identical and similar NG vowels [iː, ɪ, eː, ɛ, uː, ʊ, oː ɔ, ɑː, a] were tested against the 11 AE vowel categories. Overall, only 56% of these NG vowels were classified as exemplars of their AE counterparts (range across vowels=0%–100%). Clearly then, NG vowels transcribed with the same phonetic symbols did not necessarily coincide spectrally to their AE counterparts. Second, the 4 front rounded vowels [yː, ʏ, øː, œ] were classified with respect to the 11 AE vowel centers of gravity; no definition of the ''correct'' classification of these vowels was possible.

To specify further the cross-language acoustic similarity of particular NG and AE vowels (including the front rounded vowels of greatest interest), Table II presents classification results for each NG vowel, grouped into sets of front rounded vowels (top), mid and mid-low vowels (middle) and

TABLE II. Acoustic similarity (F1/F2/F3 Bark values) of North German (NG) and American English (AE) vowels: Study 1—syllables.

		Modal classification		Other categories	
	NG vowel	AE vowel	# of stimuli	AE vowel	# of stimuli
Front rounded	yː	iː	7	ɪ	1
	øː	ɪ	4	ʊ	2
				uː	2
	ʏ	ɪ	4	ʊ	3
				uː	1
	œ	ɪ	3	ɛ	1
		ʊ	3	ʌ	1
Mid and mid low	eː	e͡ɪ	4	iː	2
				ɪ	2
	oː	o͡ʊ	4	uː	4
	ɛ	ɛ	4	ɪ	4
	ɔ	o͡ʊ	6	ɔː	2
High	iː	iː	8		
	uː	uː	8		
Mid-high	ɪ	ɪ	6	iː	1
				e͡ɪ	1
	ʊ	o͡ʊ	5	uː	3
Low	ɑː	ɑː-ɔː	4	ʌ	4
	a	ɑː-ɔː	5	ʌ	3

high, mid-high and low vowels (bottom). The AE category to which most tokens of each NG vowel were assigned, is given in the second column with the number of tokens so assigned (third column), while the next two columns show the AE categories to which the remaining tokens were assigned. For this analysis, the AE categories [ɑː-ɔː] were collapsed due to their considerable spectral and temporal overlap in the comparison corpus and their perceptual confusability by many native AE listeners.

Looking first at the front rounded vowels, results revealed that three of the four vowels were, indeed, acoustically intermediate between front unrounded and back rounded AE vowels; half the tokens of [ʏ, øː, œ] were classified as more similar to AE front vowels; half were classified as AE back vowels. All tokens of NG [y] were spectrally more similar to front than back AE vowels. Thus, while front rounded NG vowels are more similar acoustically to front than back NG vowels [see Fig. 1(A)], they are only slightly more similar to front than to back AE vowels. This is because AE back vowels are acoustically ''fronted'' relatively to NG back vowels (see Fig. 2).

Cross-language acoustic classification patterns for the mid and mid-low vowels (second set) verify cross-language differences shown in Fig. 2. Half the tokens of 3 of the NG vowels [eː, oː, ɛ] were classified as more similar to higher AE vowels, while 75% of the NG [ɔ] tokens were classified as most similar to AE [o͡ʊ]. This confirms earlier findings that mid and mid-low German vowels are located ''higher'' in vowel space than are AE mid and mid-low vowels.

With respect to the high, mid-high, and low NG vowels, only [iː, uː] were unanimously assigned to their AE counterparts. For the other 3 vowels, individual tokens were assigned to different AE categories such that the modal assignment percentages varied from 50% to 75%. However, inconsistency in the classification of NG [ɪ] and [ɑː, a] reflected spectral overlap of AE categories [e͡ɪ-ɪ] and [ɑː-ɔː-ʌ], respectively. Finally, NG [ʊ] was never classified as most similar to AE [ʊ], being acoustically higher and less fronted than its AE counterpart.

4. Discussion

Results of the acoustic comparison of NG and AE vowels can be summarized as follows: (a) NG front rounded vowels, while more similar acoustically to front unrounded vowels in German, were found to be acoustically intermediate between front unrounded and back rounded AE vowels. (b) While NG back rounded vowels were classified as most similar spectrally to back AE vowels, F2 values across languages clearly reveal that AE back vowels were ''fronted'' relative to their NG counterparts even in this noncoronal consonantal context. (c) Mid and mid-low NG vowels were acoustically ''higher'' than their AE counterparts. (d) Whereas NG and AE front mid-high [ɪ] were very similar spectrally, back mid-high [ʊ] differed considerably for NG and AE vowels. The spectral overlap across NG and AE corpora for the high front and low back vowels [ɪː- ɑː] suggest that the differences in NG and AE vowels summarized in (a)–(d) above were not due to cross-corpora speaker normalization problems, but rather were due to cross-language differences in relative spacing of vowels in F1/F2/F3 vowel space.

B. Perceptual similarity of NG and AE vowels

Based on the acoustic comparison of the NG stimuli with a set of AE vowels produced in a similar consonantal context, we predicted the following perceptual assimilation patterns:

(1) NG front rounded [yː, ʏ, øː, œ] were expected to be assimilated inconsistently to front and back AE vowel categories except for [yː], which would be considered more similar to AE [iː]. In general, these vowels were predicted to be judged as very poor exemplars of any AE category, reflecting the fact that they are phonologically noncontrastive in English and are acoustically dissimilar from any AE vowel categories in noncoronal contexts.
(2) NG mid and mid-low vowels [eː, oː, ɛ, ɔ] were expected to be inconsistently assimilated to their AE counterparts (and considered relatively ''poor'' exemplars), due to their being produced with higher ''target'' values (lower F1 frequencies), and, for the long vowels, as monophthongs. Similarly, NG [ʊ] was expected not to be assimilated to its AE counterpart, but rather heard as more similar to higher back vowels.
(3) The NG vowels [iː, ɪ, aː, uː] were predicted to be consistently assimilated to their AE counterparts. However, [uː] might be judged as a relatively poorer exemplar of its AE counterpart, due to its being acoustically further ''back'' (lower F2 frequencies). NG [a] was expected to be assimilated consistently to AE [ɑː-ɔː] or [ʌ], depending upon the extent to which listeners attended to relative duration in making cross-language similarity judgments.

(4) Temporal assimilation patterns of other vowels were also of interest in this study. While spectral and temporal acoustic similarity patterns led to the same perceptual assimilation prediction for most NG vowels, [øː, ɔ, ʊ, a] provided a test of whether AE listeners would categorize these vowels as more similar to the spectrally closest AE vowel or to the temporally more similar AE vowel. If vowel duration had a significant effect on assimilation of these vowels, then [øː] might be expected to be assimilated more often to AE long [iː, uː] than to AE [ɪ, ʊ], while the remaining three NG vowels would be assimilated to AE short vowels [ʊ, ʌ]. However, since these vowels were produced and presented as isolated syllables, previous research suggests that the influence of vowel duration might not be significant in this study.

1. Method

a. Stimulus materials. The 112-stimulus corpus described above served as materials for the perceptual assimilation test. Each speaker's productions were arranged into a separate listening test in which the 28 stimuli (14 vowels $\times 2$ tokens) appeared 4 times each (112 trials) in random order in a block, with a 5 s response interval between presentations. The test consisted of two such randomized blocks, for a total of 8 trials on each syllable, 16 trials for each NG vowel produced by that speaker.

b. Listeners and procedures. Twelve native speakers of AE were recruited from introductory classes in Psychology and Communication Sciences. Some had a beginning course in phonetic transcription, but none were trained listeners. None had any experience with German, either in courses as a foreign language, or with individuals who spoke German.

The test stimuli were output via a DAT recorder routed through a power amplifier (Tascam PA-20B) to headphones and presented to individual listeners seated at the computer console placed inside an acoustic chamber. Categorization and rating responses were obtained using an interactive HyperCard program that displayed the 11 AE vowel category choices represented by key words *heed, hid, hayed, head, had, hud, hod, hawed, hoed, hood, who'd*. These response alternatives appeared on the computer screen with the IPA symbol for each vowel category listed above and a response "button" beside each word. A stimulus was presented and listeners clicked on the button beside the AE word that contained the vowel that was most similar to the vowel they heard during a 5 s response interval. Then the same stimulus was repeated and a seven-point horizontal scale appeared below the response alternatives. Listeners rated the goodness of the stimulus as an exemplar of the category they had chosen by clicking on the scale. Instructions were to click on the 1 if the vowel was "very foreign-sounding;" 7 if the vowel was "very English-sounding" and to choose an appropriate number between 1 and 7 if the vowel was somewhere in-between.[4]

The experiment was administered in five sessions, usually on five separate days; a few listeners completed two sessions on the same day with at least a one-hour break to combat possible fatigue effects. Day 1 included a response familiarization procedure to train the listeners to use the response alternatives consistently and to perform the task when listening to AE vowels. Listeners completed 55 trials (five blocks of 11 AE vowels each) and were given feedback following each of the first three blocks. The last two blocks were used as an assessment of task mastery. Listeners who misidentified more than three AE vowel tokens, or misidentified both presentations of any one vowel, (except for [ɑː-ɔː] confusions),[5] were not included in the study.

If the familiarization criterion was met, listeners returned for four test sessions of approximately 90 min each. In each session, they were tested on a different NG speaker's productions; speaker presentation order was counterbalanced across subjects. Before each test, listeners were reminded to use the whole seven-point scale in judging the stimuli. Each test session consisted of (a) 28 "listen only" stimuli for the listener to become familiar with the new speaker; (b) two sample stimuli (one "foreign-sounding" [œ] and one "English-sounding" [iː]); (c) 14 practice trials on which listeners categorized and rated the stimuli, but the responses were not included in the analysis; and then (d) the two blocks of 112 stimuli for that speaker.

2. Results

For each of the 14 NG vowels, the frequency of selection of each response category was tallied across all speaker conditions and all 12 listeners (16 trials/vowel$\times$12 listeners $=$192 trials). Frequencies are reported as percentages of opportunities (frequency/192); the modal (most frequently selected) and the second most frequently selected response alternatives are reported in Table III (columns 2 and 5) with their percentages (columns 3 and 6). Finally, the goodness ratings assigned the modal and second most frequent response choices were tallied (again pooling over all speaker conditions and listeners' responses) and the median ratings are reported (columns 4 and 7). Thus, the modal response percentages can be considered a measure of consistency of perceptual assimilation to a particular native vowel category for the average AE listener, whereas the median rating can be considered a measure of the judged goodness of the NG vowel as a member of that AE category. If the NG vowel is perceived as an excellent exemplar of an AE category, response consistency should be near 100% and goodness ratings should be close to the English-sounding end of the scale (7). If a NG vowel is perceived as a poorer instance of a particular AE vowel, then we would expect modal category consistency to be somewhat lower and the median goodness rating to be closer to the "foreign-sounding" end of the scale (1). Finally, if a particular NG vowel is perceived as very different from any AE vowel (uncategorizable in Best's terms), we might expect low consistency within and across listeners in category assignment and very low goodness ratings.

The NG vowels are organized in Table III in three clusters according to predictions about assimilation, based on spectral similarity patterns: (1) front rounded vowels (acoustically intermediate between front and back AE vowels); (2) mid and mid-low vowels and [ʊ] (which were spectrally dissimilar from their AE counterparts); and (3) high, mid-high,

TABLE III. Perceptual Assimilation of NG Vowels to AE Categories: Study 1—Syllables.

		Most frequent category			2nd most frequent category		
	NG vowel	AE vowel	% chosen	Median rating	AE vowel	% chosen	Median rating
Front rounded	yː	uː	69	2	iː	24	1
	øː	ʊ	37	1	uː	30	2
	ʏ	ʊ	56	3	ʌ	20	2
	œ	ʌ	62	5	ɛ	30	3
Mid	eː	iː	66	5	e͡ɪ	23	4
	oː	o͡ʊ	89	5	ɔː	5	1
Mid-low	ɛ	ɛ	97	6	ɪ	1	5
	ɔ	ɑː-ɔː	90	5	ʌ	8	5
Mid-high back	ʊ	o͡ʊ	42	3	ɑː	22	2
High	iː	iː	97	7	ɪ	1	2
Mid-high front	uː	uː	86	4	o͡ʊ	6	4.5
	ɪ	ɪ	97	7	iː	1	5
Low	ɑː	ɑː-ɔː	98	6	ʌ	1	6.5
	a	ɑː-ɔː	79	5	ʌ	12	6

and low vowels (which were spectrally similar to their AE counterparts).

First, the four front rounded vowels were, in general, not consistently assimilated to any one AE vowel (37%–69%). All four vowels, however, were considered more similar to back AE vowels; pooled over all four vowels and all AE back vowel response alternatives, the overall assimilation to back vowels was 77%. This was true for all four speakers' productions; back vowel responses ranged from 68% to 84% across speakers. Front vowel assimilations were somewhat greater for [yː] (27%) and [œ] (32%) than for the other two vowels (16% for both vowels). Except for [œ], median ratings reflected the fact that listeners judged these vowels to be relatively poor exemplars of either front or back AE vowel categories.

An inspection of individual listeners' perceptual assimilation patterns revealed considerable individual differences in assimilation of NG fronted rounded vowels. Only 2 of the 12 listeners assimilated all 4 front rounded NG vowels primarily to front AE categories (69% and 95%, respectively, pooling over all AE front vowel responses). A third listener assimilated NG [yː, œ] to front AE vowels on a majority of trials. For the remaining 9 listeners, modal assimilation patterns for all 4 vowels were to back AE categories (front vowel assimilations: median=6%, range=0%–17%, pooling over all 4 vowels and all 4 speakers).

Turning to the spectrally dissimilar NG vowels shown in the second cluster, it is apparent that the mid back [oː] was assimilated as a moderately good exemplar of its AE counterpart (89% consistency; median rating=5), while the mid front [eː] was not heard as most similar to its AE counterpart, but rather, was assimilated more often as a moderately good exemplar of AE [iː]. Individual listener data showed that only 3 of 12 listeners assimilated NG [eː] to AE [e͡ɪ] on the majority of trials. The other nine listeners assimilated it to [iː] or sometimes [ɪ]. When [ɑː-ɔː] responses were collapsed,[5] NG [ɔ] was assimilated relatively consistently as a moderately good exemplar of its AE counterpart, while NG [ɛ] was very consistently assimilated as a very good exemplar of its AE counterpart. Finally, NG [ʊ] was not assimilated to any one AE vowel on a majority of trials and was judged to be quite foreign sounding. Indeed, no one speaker's instances of NG [ʊ] were categorized as any one AE vowel on more than 53% of opportunities. Individual listener data also revealed inconsistency across speakers and tokens in the assignment of this NG vowel to AE categories. This NG vowel, then, was considered "uncategorizable" by the listeners in this study.

As was predicted, the spectrally similar NG vowels [iː, ɪ, ɑː] were very consistently categorized as most similar to their AE counterparts and judged to be excellent exemplars of those vowels. Again as predicted, the remaining "point vowel" NG [uː], was assimilated a bit less consistently and rated only a fair exemplar, reflecting its spectral dissimilarity from AE [uː]. Finally, NG [a] was assimilated as a moderately to very good exemplar of three spectrally overlapping AE vowels [ɑː, ɔː, ʌ].

With respect to temporal assimilation patterns, overall, long NG vowels were assimilated to long AE categories 88% of the time (range=46% to 99% across the 7 vowels), while short NG vowels were assimilated to short AE vowels only 62% of the time (range=9% to 99% across the 7 vowels). These large ranges reflect that fact that the four vowels for which spectral and temporal similarity patterns led to different predictions [øː, ʊ, ɔ, a] were all assimilated more often to the spectrally closest AE vowel that differed in intrinsic duration. For the remaining 10 vowels, temporal assimilation to the "correct" AE duration category averaged 95% (range =91% to 99%). Thus, we can conclude that when cross-language spectral similarity is in conflict with temporal similarity, AE listeners' assimilation of non-native vowels is more influenced by spectral similarity.

To summarize so far, 4 NG vowels [iː, ɪ, ɛ, ɑː] were assimilated very consistently as excellent instances of particular native AE vowels (consistency >95%; median ratings 6–7), while another 4 NG vowels [uː, oː, ɔ, a] were consistently assimilated as somewhat less good exemplars of particular AE vowels (consistency 79%–90%; median ratings 4–5). An additional 4 vowels [eː, yː, ʏ, œ] were less consistently assimilated as fair to poor exemplars of particular

AE categories (consistency 56%–73%; median ratings 2–5), while the remaining two vowels [øː, ʊ] were uncategorizable as any one AE vowel (consistency <50%; median ratings =1–3).

C. Discussion

In comparing perceptual assimilation data (Table III) and acoustic data (Table II), it can be seen that perceptual similarity patterns were not well predicted by cross-language spectral similarity patterns for the front rounded vowels. While acoustically intermediate between front and back AE vowels, these vowels were far more often assimilated to AE back vowel categories, albeit judged very poor tokens of those categories. For the mid and mid-low front and back NG vowels, which have ''similar'' counterparts in AE, we can see that, again, acoustic similarity patterns did not always predict perceptual assimilation accurately. NG [oː, ɛ] were quite consistently assimilated to their AE counterparts, despite acoustic differences in ''height,'' while the other vowels were often assimilated to higher AE vowels, as predicted from acoustic similarity patterns. The high, mid-high, and low NG and AE vowels [iː, ɪ, ɑː, uː] that were very similar acoustically in both spectral and temporal structure, were perceptually assimilated highly consistently and stimuli were rated as very good exemplars of native categories. Of interest is that category goodness judgments for NG [uː] appeared to reflect the phonetic differences from the somewhat fronted AE similar vowel. Thus, we can at least tentatively conclude that goodness judgments reflect knowledge about phonetic details in the native language that can be accessed in the perceptual assimilation task.

The results of this study demonstrate that neither abstract (context-independent) phonological descriptions of NG and AE vowels, nor (context-dependent) acoustic comparisons adequately predict perceptual assimilation patterns. One problem with this study, however, was that the citation-form NG and AE vowels that were compared acoustically differed in the syllable structure (monosyllabic versus disyllabic) and consonantal context. Furthermore, the productions of the NG vowels, especially their intrinsic duration differences, may not reflect the acoustic structure of NG vowels produced in the sentence context. Thus, study 2 was conducted to examine the perceptual and acoustic similarity of NG and AE vowels in /hVC/syllables when they were produced and presented in short carrier sentences at a speaking rate more closely resembling continuous (casual) speech.

III. ACOUSTIC AND PERCEPTUAL SIMILARITY OF NG AND AE VOWELS IN SENTENCE CONTEXT

Study 2 was a replication of the first study, except that the NG vowel corpus was hVp syllables produced in the sentence, ''Ich habe hVp gesagt'' by the same set of speakers with instructions to ''speak at the rate you would if you were speaking to a native German listener.'' This corpus was then compared to a corpus of AE vowels produced in hVb syllables in the sentence ''I say the hVb on the tape,'' produced under similar instructions to ''speak as if you were talking to a friend who was a native English speaker.''

As in the first study, cross-language acoustic similarity of three sets of vowels was investigated: (1) the NG front rounded vowels [yː, ʏ øː, œ] relative to front and back AE vowels; (2) the NG mid and mid-low vowels [eː, ɛ, oː, ɔ] and mid-high back [ʊ], which in study 1, were found to be acoustically dissimilar from their AE counterparts; and (3) the remaining high, mid-high front, and low vowels [iː, ɪ, uː, ɑː, a], which were acoustically similar across languages in study 1. Of interest was the extent to which cross-language spectral similarity differed from that established for ''canonical'' vowels spoken in citation-form utterances. Perceptual assimilation tests to assess the extent to which patterns of phonetic similarity differed for vowels in sentence context from those established for citation-form syllables were then completed and results of context-specific acoustic and perceptual similarity patterns were compared.

A. Acoustic Similarity of NG and AE vowels produced in sentence context

1. Speakers and stimulus materials

The same four NG speakers as in study 1 produced the stimulus materials during the same recording session, using the same equipment as in study 1. The sentences containing the hVp syllables were written on index cards and speakers read them at a rate simulating that of continuous speech. After some initial practice, speakers produced three randomizations of 15 sentences (the final sentence was not used). As before, the first randomization was used for further practice and was not included in the final corpus unless an utterance in randomization 2 or 3 was rejected because of a disfluency or difference in sentence prosody. The same phonetically trained German listener transcribed the target vowels; all were considered good instances of the intended vowels.

The same four AE speakers who produced the hVba disyllables for comparison in study 1 produced a set of hVb syllables imbedded in the carrier sentence that was similar in structure to the German sentence in the number of syllables and position of the target vowel in the sentence. The recording was done at the same time as for the citation corpus, using the same equipment and procedures. (This corpus was also used to assess the Japanese perceptual assimilation of AE vowels in Strange *et al.*, 1998.) The final corpus consisted of three instances of each of 11 AE vowels produced by each speaker. Acoustic analysis was performed using the same procedures as in study 1.

2. Results

Figure 3(A) displays the 8 instances of each NG vowel (2 instances×4 speakers) plotted in F1/F2 Bark frequency space, with ellipses surrounding all tokens of each vowel. Figure 3(B) displays the 12 instances of each AE vowel (3 instance×4 speakers). Table IV presents the average formant frequencies and durations for NG vowels (on the left) and AE vowels (on the right). The duration ratio of long to short vowels for each language is given in the heading.

First, it is immediately apparent that the temporal structure of the NG hVp syllables spoken in sentence utterances differed markedly from those produced in isolated syllables.

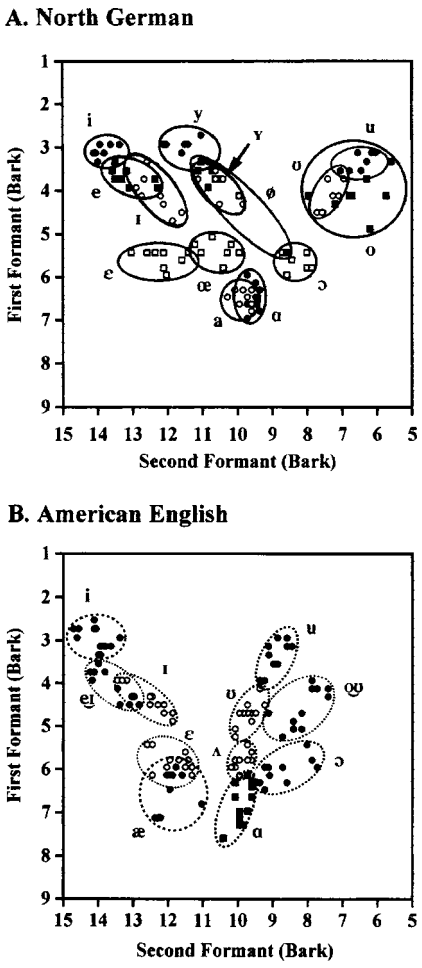


FIG. 3. Formant 1/formant 2 (Bark) plots of North German (A) and American English (B) vowel corpora in study 2. Syllables in sentences.

The long/short duration ratio was smaller (1.5 *vs* 1.9), reflecting the fact that the long vowels were shortened considerably more when produced in sentence context (mean duration = 95 ms *vs* 153 ms in isolated syllables), than were the short vowels (mean duration = 63 ms *vs* 80 ms in isolated syllables). This pattern held for all four NG speakers' productions and all seven long/short vowel pairs. In contrast, the long/short duration ratio for AE vowels was the same in sentence materials as for citation materials and absolute durations were slightly *longer* for vowels in sentences than in disyllables. Despite the reduction in the temporal distinctiveness of NG long and short vowels as a function of prosodic context and the differences across languages in absolute durations (due, in part, to the AE vowels being lengthened preceding voiced consonants), the *relative* duration differences between NG long and short vowels were still reliably greater than for long and short AE vowels. Thus, we can conclude that vowel length is phonetically more salient in NG than in AE, despite the fact that in both languages, long/short vowel pairs (except for NG [ɑː-a]) differ in target formant frequencies as well as in length.

TABLE IV. Average formant frequencies (Hz) and durations (ms) of North German (NG) and American English (AE) vowels in hVp/hVb syllables in sentences—Study 2. NG long/short vowel ratio = 1.5. AE long/short vowel ratio = 1.3.

NG	F1	F2	F3	Duration	AE	F1	F2	F3	Duration
iː	317	1943	2971	84	iː	303	2336	2961	108
ɪ	428	1784	2462	54	ɪ	461	1826	2634	94
eː	382	2008	2697	97	ɐɪ	423	2175	2722	132
ɛ	597	1738	2471	65	ɛ	627	1657	2544	98
					æː	714	1645	2456	147
yː	306	1590	2061	84					
ʏ	406	1348	2104	63					
øː	409	1345	2051	99					
œ	551	1364	2231	74					
uː	344	710	2002	84	uː	342	1064	2422	115
ʊ	441	836	2398	60	ʊ	495	1202	2492	107
oː	427	727	2454	100	ɒʊ	479	933	2571	126
ɔ	610	966	2414	60	ɔː	660	1056	2571	152
ɑː	713	1173	2438	115	ɑː	754	1234	2609	125
a	713	1227	2395	64	ʌ	631	1232	2619	98

TABLE V. Results of within-context and cross-context discriminant analyses. Percentages reflect "correct" classification as the intended vowels, based on spectral parameters alone (F1/F2/F3 in Bark) shown in the left-hand columns, and spectral parameters plus vocalic duration, shown in the right-hand columns.

A. North German Corpora	F1/F2/F3	F1/F2/F3 + duration
Sentence input–sentence test	80%	90%
Citation input–citation test	79%	93%
Citation input–sentence test	73%	78%
B. American English Corpora	**F1/F2/F3**	**F1/F2/F3 + duration**
Sentence input–sentence test	84%	94%
Citation input–citation test	86%	92%
Citation input–sentence test	79%	92%

As in study 1, a series of linear discriminant analyses was performed to quantify within- and cross-language acoustic similarity. First, separate analyses for NG and AE corpora were performed, using F1/F2/F3 Bark frequencies as input parameters to establish within-language spectral distinctiveness for NG and AE vowels. A second set of discriminant analyses were performed using duration as an additional input parameter. Table V (A and B) presents the overall correct classification results for these analyses in the top row of each section (sentence input–sentence test). For comparison, data from study 1 (citation input–citation test) are included in the second row for each language.

For the NG sentence corpus, the overall correct classification on the basis of spectral parameters alone was 80% (50%–100% across the 14 individual vowels). Most confusions were of spectrally similar long/short pairs [eː-ɪ], [ɑː-a], [øː-ʏ], [oː or uː-ʊ]. When the duration was included as an input parameter, overall correct classification rose to 90% (63%–100% across vowels) with at least 7 of 8 tokens of all vowels except [oː, øː, œ] correctly classified. This is slightly lower than for study 1 and reflects greater variability within and across speakers in production of vowels in sentences. When the NG vowels produced in sentences were evaluated against spectral centers of gravity established for the citation utterances (i.e., using "canonical" spectral values), overall correct classification for NG vowels produced in sentences was only 73% (38% to 100% across vowels), as shown in the third row of Table V(A). That is, spectral centers of gravity of vowels varied across prosodic contexts such that 2 or

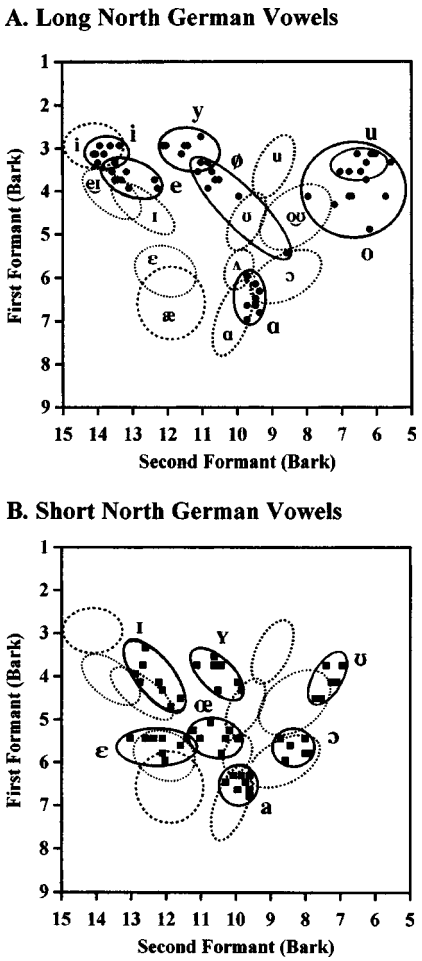


FIG. 4. North German long vowels (A) and short vowels (B) superimposed on ellipses of the 11 AE vowel categories. Study 2.

more tokens of 9 of the 14 vowels produced in sentence context were classified as more similar to spectrally adjacent ''canonical'' vowel targets. Nevertheless, when duration was included as a parameter, some of these confusions were resolved (78% correct classification overall; 38% to 100% across vowels), with at least 7 of 8 tokens of 10 of the vowels correctly classified on the basis of citation-form spectral centers of gravity and duration established on the basis of citation utterances.

These cross-context analyses reaffirm the finding that high, mid-high, and mid NG vowel spectral targets [iː, ɪ, eː] [yː, ʏ, øː], and [uː, ʊ, oː] are located quite close together in vowel space, and may be ambiguous across speakers and prosodic contexts. The marked decrease (90% vs 78%) in the correct classification rate for the cross-context analysis that included vocalic duration as a parameter reveals that durations of NG vowels in citation-form utterances were not representative of the temporal structure of vowels spoken in continuous speech. However, even in continuous speech, relative vowel duration served to acoustically differentiate spectrally similar NG vowels if continuous speech vowel durations were used to establish parameter weightings and centers of gravity (increase from 80% to 90% correct classification).

As shown in Table V(B), 84% of the AE vowel tokens produced in sentences were correctly classified as the intended vowels on the basis of spectral parameters alone (58% to 100% across the 11 vowels). Again, most confusions were of spectrally adjacent long/short pairs [ɐɪ-ɪ, æː-ɛ, ɑː-ʌ] and of [ɑː-ɔː]. When the duration was included as an input parameter to a new discriminant analysis, overall correct classification rose to 94% (83%–100%) with at least 11 of 12 tokens of all vowels except [ɔː, ʌ] correctly classified. When canonical (citation-form) spectral and temporal values were used to assess vowels produced in sentences (citation input–sentence test, overall correct classification decreased only slightly (92% overall), reflecting the fact that absolute and relative durations were similar across prosodic contexts for these corpora. Overall the correct classification of vowels produced in sentences on the basis of canonical spectral centers of gravity (79% overall) showed that AE vowel targets varied somewhat as a function of prosodic context, especially for AE back rounded vowels [ɔː, o͡ʊ, uː].

Of greatest interest was the cross-language discriminant analysis, in which the 112-vowel NG sentence corpus was classified with respect to the parameter weightings and centers of gravity established for the AE sentence corpus, using F1/F2/F3 Bark frequencies as parameters. (Again, no cross-language analysis was performed with duration as an additional input value.)

Figure 4 illustrates the cross-language spectral overlap for long (A) and short (B) NG vowels, superimposed on ellipses, indicating AE vowel categories. Table VI presents the cross-language classification results for individual NG vowels. As in the previous study, NG vowels produced in sentence context varied considerably in their spectral similarity to particular AE vowels (50%–100% classification consistency across the 14 vowels). Overall, the front rounded NG vowels (upper set) were again spectrally intermediate between front and back AE vowels (47% classified as front,

TABLE VI. Acoustic similarity (F1/F2/F3 Bark) of NG and AE vowels: Study 2—Sentences.

		Most frequent		2nd most frequent	
	NG vowel	AE vowel	# of stimuli	AE vowel	# of stimuli
Front rounded	yː	ɪ	6	ɐɪ	2
	øː	ɪ	4	u	2
				ʊ	1
				ɔ	1
	ʏ	uː	4	ʊ	3
				ɪ	1
	œ	ʊ	4	ɛ	2
				ʌ	1
				uː	1
Mid	eː	ɐɪ	5	ɪ	2
				iː	1
	oː	o͡ʊ	8		
Mid-low	ɛ	ɛ	8		
	ɔ	ɑː-ɔː	6	o͡ʊ	2
Mid-high back	ʊ	o͡ʊ	7	uː	1
High	iː	iː	7	ɐɪ	1
	uː	ʊ	6	o͡ʊ	2
Mid-high front	ɪ	ɪ	5	ɐɪ	3
Low	ɑː	ɑː-ɔː	5	ʌ	3
	a	ɑː-ɔː	4	ʌ	4

53% classified as back), with [yː] consistently assigned to front AE vowels, [ʏ, œ] more similar to back AE vowels, and [øː] evenly split between front and back AE categories.

Turning next to the five vowels that were acoustically dissimilar to their AE counterparts in study 1 ([eː, ɛ, oː, ɔ, ʊ] shown in the middle set of the table), the data reveal some differences. All instances of both NG [ɛ, oː] were classified as acoustically most similar to the equivalent AE category. In addition, more tokens of NG [eː, ɔ] were classified as similar to their AE counterparts than in study 1, while NG [ʊ] was still more similar acoustically to AE [o͡ʊ] than to AE [ʊ]. In general, then, these vowels were better matches to AE vowels when produced in sentence context than in citation-form syllables.

The last set of vowels (the lower set in the table) were acoustically very similar to their AE counterparts in study 1. Except for NG [uː], these vowels were most similar acoustically to the AE vowel transcribed as ''identical'' in sentence context as well, although there was still some acoustic ambiguity among the front vowels [iː, eː, ɪ] and the back vowels [uː, oː, ʊ], reflecting the relative closeness of NG high and mid long vowels on the one hand, and the acoustic overlap of AE [e͡ɪ, ɪ] on the other hand. As in study 1, NG overlapping long and short low vowels [ɑː, a] were acoustically similar to AE overlapping categories [ɑː, ɔː, ʌ].

In summary, then, in comparison with the results of study 1, there were some differences in cross-language acoustic similarity patterns for NG and AE vowels produced in the sentence context. NG [ɛ, ɔ] were somewhat better fits to their AE counterparts in sentence context, and NG [ʊ] was even more similar to AE [o͡ʊ]. In contrast, NG [iː, uː] were somewhat poorer fits to their AE equivalents in sentence utterances. Finally, the NG front rounded vowels tended to be classified as similar to AE back vowels slightly more often in sentence context (53% vs 44% for citation syllables). Despite these minor differences, however, predictions about perceptual assimilation patterns from acoustic similarity were the same as before: (1) front rounded vowels should be assimilated inconsistently to both front and back AE vowels; (2) NG mid long [eː], but not [oː] should be assimilated inconsistently to higher AE vowels, whereas NG [ʊ] should be assimilated to AE [o͡ʊ or uː]; and (3) there might also be some inconsistency in the assimilation of NG vowels transcribed as ''identical,'' especially NG [ɪ, uː].

B. Perceptual similarity of NG and AE vowels produced in sentence context

The NG sentence corpus was presented to native speakers of AE to assess directly how they would be assimilated to native AE vowel categories. Questions of interest were as follows.

(1) To what extent are perceptual assimilation patterns (and differences in assimilation from study 1) predictable from the acoustic similarity of NG and AE vowels produced in similar carrier sentences? Of special interest was whether the perceptual similarity of front rounded NG vowels to front or back AE vowels corresponded more closely with acoustic similarity patterns than in study 1.
(2) Do patterns of perceptual assimilation of NG vowels differ from those found in study 1? Specifically, are any of the eight NG vowels that were considered good to excellent exemplars of AE categories when produced in citation-form syllables, assimilated *less* well when produced in sentence-length utterances? Conversely, are any of the six NG vowels that were considered poor exemplars or uncategorizable as any AE categories in syllables, *better* assimilated when produced in sentence-length utterances?
(3) Are long and short NG vowels assimilated to long and short AE vowels more consistently when they are produced and presented in sentence-length utterances? From our previous research on Japanese listeners' assimilation of AE vowels (Strange *et al.*, 1998), it was expected that the presence of the rhythmic pattern of the sentence might perceptually enhance temporal differences in vowels such that temporal similarity to native categories might influence perceptual judgements to a greater extent.

1. Method

Twelve native speakers of AE, drawn from the same pool of undergraduate students in introductory Psychology and Communication Sciences & Disorders classes, served as listeners. Equipment and procedures were identical to those described in study 1, except that subjects were told to listen to the vowel in the target syllable of the sentence and given familiarization and practice using the sentence materials.

2. Results and discussion

As in study 1, perceptual assimilation responses were tallied across all 12 listeners for each of the 14 vowels and frequencies of selection of each response alternative reported as percentages of opportunities. Median ratings of category goodness for each response alternative were also computed. Table VII reports the most frequently selected response alternative for each vowel, with percentages and median ratings (columns 2–4) and the second most frequently selected alternatives, with percentages and ratings (columns 5–7). For NG [ɑː] and [ɔ], the AE [ɑː] and [ɔː] response categories were collapsed (and percentages combined) because many listeners did not distinguish these vowels in their own dialect. Thus, for these vowels, the response alternatives listed in column 5 are the third most frequently selected alternative. The NG vowels are clustered into the same three sets as for study 1: (a) the front rounded vowel (top); (b) the mid and mid low front and back vowels [eː, ɛ, oː, ɔ] and [ʊ] (middle); and (c) the high, mid high front, and low vowels (bottom).

Two of the front rounded NG vowels [yː, œ] were quite consistently assimilated to AE [uː, ʌ], respectively, but judged as relatively poor exemplars of those categories. The other two front rounded vowels were also assimilated overwhelmingly to AE back vowels, but not consistently to any single AE category. Overall, the percentage of assimilations

TABLE VII. Perceptual assimilation of NG Vowels to AE Categories: Study 2—Sentences.

		Most frequent category			2nd most frequent category		
	NG vowel	AE vowel	% chosen	Median rating	AE vowel	% chosen	Median rating
Front rounded	yː	uː	87	2	ʊ	7	2
	øː	uː	43	3	ʊ	28	2
	ʏ	ʊ	56	2	ʌ	24	3
	œ	ʌ	80	4	ɛ	10	4
Mid	eː	e͡ɪ	53	6	iː	30	6
	oː	o͡ʊ	81	5	uː	6	2
Mid-low	ɛ	ɛ	93	6	iː	1	6.5
	ɔ	ɑː-ɔː	64	3	ʌ	23	3
Mid-high back	ʊ	o͡ʊ	67	5	ʊ	10	2
High	iː	iː	91	7	uː	2	5
	uː	uː	72	3	ʊ	11	2
Mid-high	ɪ	ɪ	87	7	ɛ	3	5
Low	ɑ	ɑː-ɔː	93	6	o͡ʊ	1	6
	a	ɑː-ɔː	74	5	ʌ	17	3

to back AE vowel categories was 96% (89% to 98% across the 4 NG vowels). This pattern reflected performance on all four speakers' utterances. Further, no listener assimilated any of the four NG front rounded vowels to front unrounded AE categories a majority of the time. The assimilation of NG [œ] to AE [ɛ] ranged across listeners from 0% to 36% (median=7%).

The second set of vowels varied considerably with respect to consistency in assimilation and judged goodness to AE counterparts. As in study 1, while NG [oː] was quite consistently perceived as a fair exemplar of its AE counterpart, NG [eː] was assimilated to AE [e͡ɪ] more consistently than in study 1, but was still often perceived as more similar to AE [iː]; note, however, that the [eː] tokens were judged as relative good exemplars of either [e͡ɪ] or [iː]. NG [ɛ] was consistently assimilated as an excellent exemplar of its AE counterpart, while NG [ɔ] was assimilated as a relatively poor exemplar of AE [ɑː-ɔː] or [ʌ]. Finally, NG [ʊ] was most often assimilated as a fair exemplar of AE [o͡ʊ], or as AE [ʊ, ʌ, ɑː].

Of the last set of vowels, NG [iː, ɑː] were very consistently perceived as excellent exemplars of their AE counterparts, while NG [uː] was less consistently assimilated as a poor exemplar of the acoustically more ''fronted'' AE [uː] or [ʊ]. As in study 1, NG [ɪ] was quite consistently perceived as an excellent exemplar of AE [ɪ], and NG [a] was assimilated more often as a fair exemplar of long AE [ɑː-ɔː] than to short AE [ʌ].

The pattern of perceptual assimilation in Table VII differed somewhat from that reported in study 1 (Table III). The first eight NG vowels [iː ɪ, ɛ, ɑː, a, ɔ, oː, uː], which were perceived as good to excellent exemplars of equivalent AE categories in study 1, were somewhat less consistently assimilated to their AE counterparts in sentence context (82%) than in syllable context (92%). This was especially true for NG [ɔ] and [uː] that were perceived as relatively poor exemplars in the sentence context. Of the six NG vowels [eː, ʊ, yː, ʏ, øː, œ], which had yielded inconsistent assimilation patterns in study 1, four [eː, ʊ, yː, œ] were perceived more consistently when produced and presented in sentence context (72% vs 60% in syllable context), and NG [eː] actually changed from being assimilated more often to AE [iː] in syllables to [e͡ɪ] when syllables were imbedded in sentences. However, 4 of these vowels [eː, ʊ, øː, ʏ] were still perceptually ambiguous (<75% categorization consistency) and/or rated as poor exemplars of any AE vowel (Mdn rating<4).

In this study, as before, perceptual assimilation patterns suggested that spectral similarity was more important than temporal similarity in determining to which AE categories NG vowels would be assimilated. NG long vowels were assimilated to AE long vowels on 89% of trials, while NG short vowels were assimilated to short AE vowels on only 61% of trials. The poorer temporal assimilation of short vowels was due to NG short [ʊ, ɔ, a] being assimilated more often to spectrally similar AE long [o͡ʊ, ɔː, ɑː], respectively. Thus, temporal assimilation patterns were not consistently better here than in Study 1.

As in study 1, acoustic similarity patterns were not predictive of perceptual assimilation of front rounded NG vowels to front and back AE vowels (upper sets of Tables VI and VII). In the sentence context, the front rounded vowels were even more consistently categorized as more similar to back rounded AE vowels, although goodness ratings indicated that they were heard as somewhat poorer exemplars of those categories than were NG back rounded vowels. For mid and mid-low NG vowels and [ʊ] (middle sets), acoustic similarity patterns predicted modal perceptual assimilation categories for all five vowels, although the minority acoustic and perceptual classifications sometimes differed. The fit between acoustic and perceptual similarity for these vowels was, in general, better when produced and presented in sentences than when produced and presented in citation-form syllables. Finally, for high, mid-high front, and low vowels (bottom sets), acoustic similarity patterns predicted modal perceptual assimilation categories for all vowels except NG [uː], although there were again some discrepancies between minority acoustic and perceptual classifications. The categorization of NG [uː] to its AE counterpart was better than was

predicted from spectral similarity patterns. However, poor goodness ratings may have reflected the listeners' perception of this vowel as quite different from their native [uː].

IV. GENERAL DISCUSSION

The results of these two studies suggest that neither context-dependent comparisons of vowel production (as specified by spectral similarity patterns) nor context-independent impressionistic descriptions of vowel inventories capture all the relevant information necessary to account for perceived similarities of vowels by non-native listeners. One weakness of this study was that the phonetic and syllabic contexts did vary somewhat across languages such that durational information could not be compared. However, research currently being completed in our laboratory which compares the acoustic and perceptual similarity of NG and AE vowels in exactly the same citation context (hVba) is yielding very similar results (Strange *et al.*, 2002, 2004). Thus, we are confident that the discrepancies between cross-language acoustic similarity and perceived similarity patterns for front rounded vowels and for mid and mid-low vowels reported here were not due to these minor differences in consonantal/syllabic context.

Given these differences in acoustic and perceived similarity, the *direct* assessment of perceptual assimilation patterns will be necessary if we are to anticipate perceptual learning problems of L2 learners. Previous research on Japanese listeners' perceptual assimilation of AE vowels (Strange *et al.*, 1998, 2001) showed that the perceived similarity of vowels may be influenced significantly by the prosodic context and the immediate phonetic context in which the vowels are produced and presented. A comparison of results of studies 1 and 2 also suggest some differences in the perceptual similarity of NG and AE vowels as a function of the prosodic context: (1) NG front rounded vowels were assimilated more consistently within and across listeners to back rounded AE vowels when produced and heard in a sentence context, despite their acoustic ambiguity in both contexts. (2) Except for [eː], the mid and mid-low NG vowels were perceived as less similar to their AE counterparts in the sentence context, despite the fact that they were more similar acoustically to AE vowels in that context. (3) The vowels [iː, ɑː] and [ɪ] were consistently perceived as similar to their AE counterparts in both prosodic contexts, while NG [uː] was less consistently assimilated to AE [uː] and judged a relatively poor exemplar, especially in the sentence context. These differences as a function of the prosodic context point to the importance of examining cross-language perceptual similarity with materials that more closely resemble continuous speech.

On the basis of the findings reported here, we can predict that native speakers of AE who are beginning to learn German as an L2 will have considerable difficulty with several NG vowel contrasts. As previous research has suggested (Polka, 1995; Polka and Bohn, 1996; see also Gottfried, 1984; Rochet, 1995 for French), NG front rounded vowels will be confused with back rounded vowels, even in non-coronal contexts, because both are assimilated perceptually to back rounded AE vowel categories. In this study, the NG vowels [yː, øː, uː] were all assimilated most often to AE [uː] or [ʊ], although [øː] was very poor fit to both categories. According to the predictions of Best's PAM, we would expect that [yː-uː] would present the most difficulty (the single category type in sentences, category-goodness type in syllables), while [øː-uː] and [øː-yː] might be discriminated rather better (uncategorizable–categorizable type). The NG mid-high short vowels [ʏ, œ, ʊ] might show a similar pattern, although modal responses differed across these vowels. Thus, we might predict that these vowels would be easier to discriminate than the long vowels. Polka (1995) reported the opposite result; that is, AE listeners' discrimination of the short front versus back rounded pair was poorer than for the long pair. Note, however, that the speaker in Polka's study spoke a Southern German dialect. Since Northern and Southern dialects are considerably different in both the spectral and temporal structure of vowels, we cannot generalize across studies. We might also expect perceptual confusions among NG [øː-ʏ] because of their spectral overlap and the apparent lack of attention to vowel duration by AE listeners.

The NG low vowel pair [ɑː-a] would also be expected to cause perceptual difficulties for AE learners of German. These vowels are distinguished acoustically almost entirely by duration, which AE listeners apparently ignored in making perceptual assimilation judgments. In many contexts, these vowels would be moderately difficult to discriminate (Category Goodness Type) unless AE learners were to become aware of their distinctive temporal differences. Finally, the NG vowels [eː-iː] might be expected to cause moderate discrimination problems, even though these vowels are phonologically distinctive in both languages. Gottfried (1984) reported that AE learners of French had difficulty with this "native" contrast. In both German and French, [eː] is not diphthongized, and the present study replicated previous findings that NG [eː] and [iː] are acoustically more similar in the "target" formant structure than their AE counterparts. Perceptual assimilation results confirmed this pair as a Category Goodness type. It is interesting to note that while the NG German pair [oː-uː] also overlap spectrally and are both monophthongal, they were more often perceptually assimilated to different AE categories (Two-Category Type). On this basis, we would predict that the front vowel contrast would be more difficult to learn than the back vowel contrast for AE learners of German. We would predict few problems with the perception of NG [iː, ɪ, ɛ, ɑː, uː]. However, NG [ɛ, uː] might be produced with an "accent" since target formant frequencies differ across languages.

In conclusion, the studies reported here document the cross-language phonetic similarity of North German and American English vowels, using both direct measures of perceptual similarity, and acoustic comparisons of AE and NG vowels produced in both citation-form (careful) and sentence (continuous, more casual) contexts. Perceptual assimilation results by AE speakers with no previous experience with German yielded patterns of cross-language similarity that differed from those predicted either from abstract impressionistic descriptions of the phonetic inventories *or* from context-dependent spectral similarity patterns. Perceptual assimilation of some NG vowels varied considerably with pro-

sodic context, suggesting that claims about cross-language perceptual similarity of vowels (and perhaps consonants as well) based on citation-form syllables may be limited in their generalizability to continuous speech materials. From these comparisons of acoustic and perceptual similarity of NG and AE vowels, we can conclude that listeners are able to make fine-grained phonetic judgments about the cross-language similarity of vowels if they are presented in citation-form materials or in sentences in which the vowels occur in a fixed consonantal context. Judgments of category goodness reflected listeners' awareness of the allophonic inappropriateness of the non-native vowels.

Future studies in which the immediate phonetic context varies unpredictably need to be performed to address several remaining issues: (1) how acoustic and perceptual similarity of NG and AE vowels vary as a function of consonantal context, and (2) whether category goodness judgments reflect an awareness of context-specific allophonic differences in cross-language similarity when the phonetic context varies from trial to trial (more similar to natural speaking situations). Answers to these questions will begin to address the nature of the underlying representations of native vowel categories and how those categories are accessed when listeners are making judgements about non-native vowels.

ACKNOWLEDGMENTS

This research was completed on a research grant to the first author from the NIH (NIDCD—00323). The authors wish to acknowledge the contributions of students and colleagues who helped with the evaluation of stimuli and analysis of results: Katherine Bielic, William Clarke III, Saratha Kumarasamy, Thorsten Piske, Melissa Sedda, David Thornton, and James J. Jenkins.

[1]Tense and lax (close–open) German vowels will hereafter be referred to as long and short, respectively. The diacritic [ː] will be used to designate the long vowels.

[2]Many studies of AE vowels have used hVd syllables, primarily because most of the vowels in this context form real English words. However, since the final /d/ may influence vowel tongue position, it was decided to use a labial final consonant that has minimal effects on ''target'' formant positions (Stevens and House, 1963; Hillenbrand, Clark, and Nearey, 2001; Strange *et al.*, 2002).

[3]The formula used to compute Barks was as follows: $13\times\text{Arctan}(0.76\times\text{Hz}/1000)+3.5\times\text{Arctan}((\text{Hz}/1000/7.5)^2)$.

[4]Listeners were instructed that they could change their categorization response after the second presentation of each stimulus before judging it as native or foreign sounding. However, participants almost never chose this option. Thus, we can assume that their first categorization response reflected similarity of the second presentation as well as the first. Two identical presentations of each stimulus were included so that goodness judgments could be made immediately after hearing the stimulus rather than at a delay. In this way, we hoped that listeners' responses would more accurately reflect detailed phonetic comparison with ''stored'' representations of native vowel categories.

[5]In many dialects of American English, the /ɑ-ɔ/ contrast is partially or completely neutralized. Thus, many of our listeners found it very difficult to differentiate these vowels in the familiarization materials and when using the key words to indicate their perceptual assimilation responses. We thus collapsed these response categories for all perceptual assimilation data analysis.

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Word*, edited by J. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA), pp. 167–224.

Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Timonium, MD), pp. 171–204.

Best, C. T., Faber, A., and Levitt, A. (**1996**). ''Assimilation of non-native vowel contrasts to the American English vowel system,'' J. Acoust. Soc. Am. **99**, 2602.

Best, C. T., and Strange, W. (**1992**). ''Effects of language-specific phonological and phonetic factors on cross-language perception of approximants,'' J. Phonetics **20**, 305–330.

Bohn, O.-S. (**1995**). ''Cross-language speech perception in adult: First language transfer doesn't tell it all,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Timonium, MD), pp. 279–304.

Bohn, O.-S., and Flege, J. E. (**1990**). ''Interlingual identification and the role of foreign language experience in L2 vowel perception,'' Appl. Psycholing. **11**, 303–328.

Crystal, T. H., and House, A. S. (**1988a**). ''Segmental duration in connected-speech signals: Current results,'' J. Acoust. Soc. Am. **83**, 1553–1573.

Crystal, T. H., and House, A. S. (**1988b**). ''Segmental duration in connected-speech signals: Syllabic stress,'' J. Acoust. Soc. Am. **83**, 1574–1585.

Flege, J. E. (**1987**). ''The production of ''new'' and ''similar'' phones in a foreign language: Evidence for the effect of equivalence classification,'' J. Phonetics **15**, 47–65.

Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings, and problems,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Timonium, MD) pp. 233–277.

Flege, J. E., and Hillenbrand, J. (**1984**). ''Limits of phonetic accuracy in foreign language speech production,'' J. Acoust. Soc. Am. **76**, 708–721.

Gottfried, T. L. (**1984**). ''Effects of consonantal context on the perception of French vowels,'' J. Phonetics **12**, 91–114.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). ''An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants,'' J. Acoust. Soc. Am. **107**, 2711–2724.

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

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). ''Effects of consonant environment on vowel formant patterns,'' J. Acoust. Soc. Am. **109**, 748–763.

Klecka, W. R. (**1980**). *Discriminant Analysis* (Sage Publication, Newbury Park, CA).

Kohler, K. J. (**1981**). ''Contrastive phonology and the acquisition of phonetic skills,'' Phonetica **38**, 213–226.

Levy, E., and Strange, W. (**2002**). ''Effects of consonantal context on perception of French rounded vowels by American English adults with and without French language experience,'' J. Acoust. Soc. Am. **111**, 2361.

Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.

Polka, L. (**1995**). ''Linguistic influences in adult perception of non-native vowel contrasts,'' J. Acoust. Soc. Am. **97**, 1286–1296.

Polka, L., and Bohn, O.-S. (**1996**). ''A cross-language comparison of vowel perception in English-learning and German-learning infants,'' J. Acoust. Soc. Am. **100**, 577–592.

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 Press, Timonium, MD), pp. 379–410.

Steinlen, A. (**2002**). ''A cross-language comparison of the effects of consonantal context on vowels producted by native and non-native speakers,'' unpublished doctoral dissertation, Aarhus University, Denmark.

Steinlen, A., and Bohn, O.-S. (**1999**). ''Acoustic studies comparing Danish vowels, British English vowels, and Danish-accented British English vowels,'' J. Acoust. Soc. Am. **105**, 1097.

Stevens, K. N., and House, A. S. (**1963**). ''Perturbation of vowel articulations by consonantal context: An acoustical study,'' J. Speech Hear. Res. **6**, 111–128.

Strange, W. (**1989**). ''Dynamic specification of coarticulated vowels spoken

in sentence context,'' J. Acoust. Soc. Am. **85**, 2135–2153.
Strange, W., and Bohn, O.-S. (**1998**). ''Dynamic specification of coarticulated German vowels: perceptual and acoustical studies,'' J. Acoust. Soc. Am. **104**, 488–504.
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (**1998**). ''Perceptual assimilation of American English vowels by Japanese listeners,'' J. Phonetics **26**, 311–344.
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). ''Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners,'' J. Acoust. Soc. Am. **109**, 1691–1704.
Strange, W., Weber, A., Levy, E., Shafiro, V., and Nishi, K. (**2002**). ''Within- and across-language acoustic variability of vowels spoken in different phonetic and prosodic contexts: American English, North German, and Parisian French,'' J. Acoust. Soc. Am. **112**, 2384.
Strange, W., Levy, E., and Lehnhoff, R. (**2004**). ''Perceptual assimilation of French and German vowels by American English listeners: Acoustic similarity does not predict perceptual similarity,'' Journal of the Acoustical Society of America (abstract forthcoming).

Effects of stimulus and noise rate variability on speech perception by younger and older adults

Sandra Gordon-Salant
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

Peter J. Fitzgibbons
Department of Audiology and Speech, Gallaudet University, Washington, DC 20002

(Received 8 November 2002; accepted for publication 11 December 2003)

The present experiments examine the effects of listener age and hearing sensitivity on the ability to understand temporally altered speech in quiet when the proportion of a sentence processed by time compression is varied. Additional conditions in noise investigate whether or not listeners are affected by alterations in the presentation rate of background speech babble, relative to the presentation rate of the target speech signal. Younger and older adults with normal hearing and with mild-to-moderate sensorineural hearing losses served as listeners. Speech stimuli included sentences, syntactic sets, and random-order words. Presentation rate was altered via time compression applied to the entire stimulus or to selected phrases within the stimulus. Older listeners performed more poorly than younger listeners in most conditions involving time compression, and their performance decreased progressively with the proportion of the stimulus that was processed with time compression. Older listeners also performed more poorly than younger listeners in all noise conditions, but both age groups demonstrated better performance in conditions incorporating a mismatch in the presentation rate between target signal and background babble compared to conditions with matched rates. The age effects in quiet are consistent with the generalized slowing hypothesis of aging. Performance patterns in noise tentatively support the notion that altered rates of speech signal and background babble may provide a cue to enhance auditory figure–ground perception by both younger and older listeners.
[DOI: 10.1121/1.1645249]

PACS numbers: 43.71.Lz, 43.71.Es, 43.71.Ky [DOS] Pages: 1808–1817

I. INTRODUCTION

Aging listeners have difficulty understanding speech, particularly when it is degraded by some form of temporal waveform distortion. The presence of hearing loss among many elderly listeners imposes an additional deficit in speech recognition that is evident in quiet, noise, and most forms of speech degradation. Nevertheless, there is evidence for age-related deterioration in speech recognition performance that exceeds that which can be attributed to hearing loss alone. Some stimulus and task factors appear to contribute to this age-related speech recognition deficit, including the type and degree of waveform distortion, the number of stimulus distortions, the length of the recall task, and the availability of contextual cues (Dubno *et al.*, 1984; Gordon-Salant and Fitzgibbons, 1993, 1997; Pichora-Fuller *et al.*, 1995). An important finding from previous research is that aging effects are prominent on speeded speech tasks, which suggests that older people exhibit a temporal processing deficit.

Age-related difficulties in understanding temporally distorted speech could arise as a consequence of temporal processing deficits associated with peripheral sensory mechanisms, central timing mechanisms, and cognitive capacities. Moreover, each of these mechanisms could contribute with varying import to a listener's performance in a specific set of stimulus/task conditions. Peripheral effects could result from sensory coding problems arising from sensorineural hearing loss and/or suprathreshold processing of short-term spectral, durational, and intensive attributes of phonetic elements. Problems associated with central timing mechanisms are more likely to influence judgments about stimulus duration and/or the perception of suprasegmental prosodic attributes of speech tempo and rhythm. Age-related cognitive decline associated with limitations in executive functions, speed of processing, and memory may also affect perception of speech signals that have a rapid presentation rate, multiple or unpredictable forms of distortion, or limited contextual cues. The goal of the present investigation is to define stimulus and processing factors that may account for the elderly listener's difficulty with temporally distorted speech.

At present, the predominant source(s) of the age-related deficit for rapid speech remains unclear. In a recent study (Gordon-Salant and Fitzgibbons, 2001), elderly listeners with and without hearing loss showed difficulty recognizing rapid speech created with selective time compression of consonants, and minimal difficulty recognizing rapid speech created with selective time compression of vowels or pauses. These results suggest that older listeners experience a deficit in processing the brief, impoverished acoustic cues for consonants in time-compressed speech. Although performance on the task involving selective time compression of consonants was the principal factor accounting for the variance in recognition of speech that was uniformly time compressed, performance on this task accounted for only 53.3% of the performance deficit for speech that was uniformly time compressed. This finding indicates that other factors contribute to

 0001-4966/2004/115(4)/1808/10/$20.00

the age-related difficulty for understanding speeded speech.

Two hypotheses may be useful in understanding the detrimental effects of rapid speech on elderly listener's performance. The first is the cognitive slowing hypothesis (Birren, 1965; Salthouse, 1982), which states that a generalized slowing accompanies the aging process and affects every event in the nervous system. The slowing model predicts that reductions in available processing time have a dramatically disproportionate effect on the performance of elderly participants; it is often cited to account for elderly listeners' poor scores on tasks that increase the stimulus presentation rate (Wingfield *et al.*, 1985). The second hypothesis is that older listeners have difficulty adjusting to novel stimuli or switching attention from one stimulus to another (e.g., Bryan *et al.*, 1999). One requirement for recognition of ongoing speech is perceptual normalization, which is the process of converting wide variations in the acoustic characteristics of speech sounds to standard phonetic representations. Sommers *et al.* (1994) manipulated the acoustic variability in speech signals using alterations in talkers and speech rate across stimuli. Young listeners showed poorer speech recognition scores in these variable conditions than in conditions with more uniform acoustic/phonetic speech signals (i.e., single talker, or uniform rate), suggesting that increasing demands on cognitive functions required for perceptual normalization can limit speech identification accuracy. In a follow-up study (Sommers, 1997), older listeners exhibited poorer recognition performance than younger listeners for speech signals with acoustic–phonetic variations (multiple talkers) compared to uniform acoustic–phonetic composition (single talker) within a stimulus list. Moreover, elderly hearing-impaired listeners showed poorer performance than elderly normal-hearing listeners in conditions with mixed speech rates. Overall, these results suggest that older listeners may experience decline in cognitive abilities related to perceptual normalization, including accurate analysis and recognition of novel stimuli, adjusting to variation in stimulus speed, and switching attention from one stimulus to another. The slowing hypothesis and the hypothesis that listeners may exhibit difficulty in the ability to adapt to a change in stimulus speed or novelty, while perhaps not mutually exclusive, could be used to account for older listeners' difficulty in recognizing rapid speech.

In experiment 1, rapid sentence-length speech stimuli, created through time-compression techniques, were presented to young and elderly listeners with normal hearing and with hearing loss. Speech-rate conditions included normal-rate speech, time-compressed speech with uniform time compression throughout the stimulus, and time-compressed speech in which one segment of the stimulus was time compressed. The slowing hypothesis predicts that performance of elderly listeners should be progressively poorer with increments in the overall presentation rate: performance should be best for normal-rate speech, poorer for speech with a single time-compressed segment (regardless of its location), and poorest for speech with uniform time compression throughout the entire stimulus. The hypothesis related to a decline in the ability to adjust to stimulus change (exclusive of any age-related slowed processing) predicts that a variation in speech rate within the sentence would place added demands on cognitive processes required to normalize the acoustic–phonetic composition of the speech signal, resulting in poorest performance in the condition with a single time-compressed segment. Conditions in which the location of the time-compressed segment is randomized rather than fixed should place additional cognitive demands because of greater novelty or unpredictability of the stimulus, creating a further deterioration in performance by elderly listeners.

The present study also investigated the effect of rate variations in background noise composed of multiple talkers on recognition of rapid speech. The addition of background noise to rapid speech is particularly difficult for older listeners (Gordon-Salant and Fitzgibbons, 1995; Tun, 1998). One possible source of this difficulty is an age-related decline in executive control. Executive functions are high-level processes that supervise the operation of other cognitive processes, for example, planning and implementing a sequence of behaviors or inhibiting task-irrelevant information. Tasks that involve divided attention, such as listening to speech in a competing message, are thought to place a large demand on executive control (Tun *et al.*, 2002). The literature on cognition and aging suggests that aging is accompanied by a decline in executive control (Bryan *et al.*, 1999; MacPherson *et al.*, 2002), and in particular, evidence has shown that older listeners exhibit a reduced ability to inhibit the processing of irrelevant information (Hasher and Zacks, 1988). The detrimental effects of background talkers on a speeded speech task also may be attributed to acoustic–phonetic masking (energetic masking), or to informational masking (Brungart *et al.*, 2001). The temporal characteristics of noise appear to influence its masking effectiveness in normal-rate speech tasks. Carhart *et al.* (1969) and Takahashi and Bacon (1992) showed that young normal-hearing listeners were able to take advantage of differences in the temporal characteristics of signal and noise to improve speech perception performance. The ability of older listeners to take advantage of temporal fluctuations in noise to mitigate its effects on speech recognition is unclear: one study (Stuart and Phillips, 1996) has shown that older listeners are less able than younger listeners to take advantage of temporal fluctuations in noise, two studies have shown no age effects (Takahashi and Bacon, 1992; Souza and Turner, 1994), and a recent study (Dubno *et al.*, 2002) has shown that older listeners benefited from interrupted noise compared to steady-state noise for syllable recognition, although the magnitude of benefit was less than that observed for younger listeners. Thus, prior research suggests that young listeners, and possibly elderly listeners, benefit from temporal modulations in noise compared to steady-state noise. One form of temporal variation in noise that has not been examined previously is the speed of a background noise composed of multiple talkers, relative to the speed of the spoken message. It is possible that listeners may be able to take advantage of differences between the speed of a target message and that of a background masker, to enhance figure–ground separation. In this case, it could be predicted that varying speeds of speech and noise would form a basis for listeners to separate the target

TABLE I. Thresholds (dB HL) (*re*: ANSI, 1996) and standard deviations (shown in parentheses) of young normal hearing (Yng Norm Hrg), elderly normal hearing (Eld Norm Hrg), young hearing-impaired (Yng Hrg Imp), and elderly hearing-impaired (Eld Hrg Imp) listeners for octave frequencies from 250 through 4000 Hz.

	Frequency (Hz)									
	250		500		1000		2000		4000	
Yng Norm Hrg	5.7	(4.6)	1.7	(3.1)	2.0	(3.2)	2.0	(5.6)	2.3	(5.3)
Eld Norm Hrg	12.7	(7.5)	7.7	(7.5)	7.3	(4.7)	5.45	(5.68)	11.82	(7.51)
Yng Hrg Imp	9.5	(17.4)	25.0	(19.7)	30.5	(18.6)	32.0	(17.5)	43.5	(14.9)
Eld Hrg Imp	19.0	(9.49)	19.3	(11.63)	25.3	(10.26)	36.0	(10.89)	50.67	(8.84)

and background and perform better than in conditions with matched speech and babble speeds. The theory that speech perception problems in noise backgrounds by older people is a manifestation of limited executive control (i.e., reduced ability to inhibit processing of irrelevant information) might predict that older people will exhibit poorer speech recognition performance than younger people in background babble regardless of its speed because any noise represents a source of irrelevant information that older listeners must inhibit in order to attend to and process relevant information. Experiment 2 therefore examined younger and older listeners' recognition of time-compressed speech in three noise conditions that varied the rate of the babble, including matched and unmatched rates relative to the target speech signal.

The current experiments were designed to investigate possible age differences in the ability to recognize rapid speech with and without variations in speech rate, and with variations in noise rate. Because many older listeners have age-related hearing loss and because hearing impairment reduces audibility of critical speech cues (Dubno *et al.*, 1989), it is important to distinguish the detrimental effects attributed to age from those attributed to hearing impairment. The experimental design employed four listener groups, young normal-hearing, older normal-hearing, young hearing-impaired, and older hearing-impaired, which permitted an assessment of the separate effects of age and hearing impairment, as well as possible interactive effects between them.

The current investigation also assessed the influence of speech contextual cues on speech understanding performance. Limitation in the availability of linguistic cues represents another cognitive demand that could exert a greater influence on the performance of elderly listeners compared to younger listeners. Prior studies (Wingfield *et al.*, 1985; Gordon-Salant and Fitzgibbons, 2001) showed that age-related deficits on speeded speech tasks were strongly influenced by the linguistic structure of the speech materials. In the present experiments, it was anticipated that older listeners would experience greater difficulty than younger listeners in temporally challenging conditions when the availability of linguistic cues was reduced (Wingfield *et al.*, 1985). To that end, the speech stimuli included three forms of sentence-length materials: original sentences, syntactic sets, and random-order words.

II. EXPERIMENT 1: VARIATIONS IN SPEECH RATE

The purpose of experiment 1 was to examine the influence of variable-rate speech on younger and older listeners' speech recognition performance.

A. Methods

1. Subjects

A total of 51 adults participated in the study. Subjects assigned to the young normal-hearing group (YNH; $n=15$) were recruited from the student population at the University of Maryland. These subjects were 18–40 years of age and had pure-tone thresholds $\leqslant$15 dB HL (*re*: ANSI, 1996) from 250–4000 Hz. Subjects assigned to the elderly normal-hearing group (ENH; $n=11$) were community-dwelling individuals aged 65–76 years who met the same audiometric criteria as the YNH subjects. The young hearing-impaired listeners (YHI; $n=10$) were 18–40 years, with mild-to-moderate sloping sensorineural hearing losses. The etiology of the hearing losses was heredity or unknown. The elderly hearing-impaired listeners (EHI; $n=15$) were 65–76 years with mild-to-moderate sloping sensorineural hearing losses. These subjects had a negative history of otologic disease, noise exposure, and family history of hearing loss. The gradual progression of the hearing loss coupled with an absence of a known cause suggested that the etiology of the hearing losses of the older listeners was presbycusis. Table I presents the mean audiograms of the four listener groups.

Additional criteria for subject selection included monosyllabic word recognition scores in quiet (Northwestern University Test No. 6) exceeding 80%, normal tympanograms, and acoustic reflex thresholds for contralateral pure-tone stimuli (500–2000 Hz) elicited at levels below the 90th percentile for individuals with comparable hearing thresholds (Silman and Gelfand, 1981). All listeners were native speakers of English and had not participated in listening experiments previously. Older listeners also passed a brief screening test for general cognitive awareness (the Mini-Mental Status Questionnaire, Pfeiffer, 1975), and were required to have sufficient motor skills to provide a legible written response.

2. Stimulus materials

The stimuli were the eight lists of the low-probability (LP) sentences of the Revised Speech Perception in Noise Test (R-SPIN; Bilger *et al.*, 1984). The sentences were digitized onto a PC and edited to create three forms: original sentences, syntactic sets, and random-order words. The original, LP sentences contain no semantic cues to predict other words in the sentence (e.g., ''Mr. White should have considered the sleeve.''). The syntactic sets preserved each subject, verb, and object phrase within each sentence, but presented these phrases in random order (e.g., ''Should have consid-

ered Mr. White the sleeve.''). The random-order words consisted of presentation of each word in the sentence in completely randomized order (e.g., ''Have White the should Mr. sleeve considered.''). The three speech forms were implemented for all 200 LP-SPIN sentences.

A computer algorithm for time compression was applied to all speech stimuli (WEDW software, Global Option; www.castle@asel.udel.edu). This algorithm extracted quasialternate epochs throughout the designated portion of the speech waveform; these extracted segments correspond roughly to alternate single pitch periods (averaging 10 ms) in the signal. Following removal of the alternating pitch periods, the algorithm applied a weighting function to overlapped points between the extracted segment and adjacent remaining segments to produce a gradual rise–fall time between sequential speech segments (T. Bunnell, 1993, personal communication). As a result, the software technique created a rapid signal of specified duration without audible clicks or discontinuities, while preserving the pitch of the original signal. The time-compression algorithm was applied uniformly throughout each sentence-length signal, or selectively to the initial, middle, or final segment of the signal. Time compression of a single segment of the original waveforms or syntactic sets corresponded to the initial, middle, or final phrase of the signal. In the case of random-order words, time compression of a single segment was applied to a set of two or three contiguous words (depending on sentence length) appearing in the beginning, middle, or end of the stimulus. The speech rate for the original sentences was approximately 200 words per minute (wpm), and was reasonably constant ($\pm$10%) across all eight sentence lists and across the three speech forms. Removal of the quasialternate epochs throughout each sentence-length signal effectively removed 50% of the acoustic signal, and concatenated the remaining segments to create the 50% time compression. Thus, the speech rate of the uniform time-compressed signals was approximately 400 wpm for each stimulus in the three speech forms. For signals with the single-phrase time compressed, the speech rate was approximately 265 wpm.

The rms level of each stimulus was calculated, and the amplitude of each of these waveforms was scaled in dB relative to the rms level of the waveform for a calibration tone, to equate all stimuli. The stimuli were converted into analog form and recorded onto a digital audio tape recorder (Sony PCM 2500) with a 16-s interstimulus interval (ISI). The order of the sentences was randomized for each recording of each list. The 16-s ISI has been shown in previous experiments to be sufficient for older listeners to provide a written response for these sentence-length stimuli (Gordon-Salant and Fitzgibbons, 1997). A calibration tone was recorded at the beginning of each tape that was equivalent in overall rms level to each of the speech stimuli.

3. Procedures

There were six speech rate conditions applied to the three stimulus forms, for a total of 18 listening conditions. The two uniform speech rate conditions were normal-rate speech and speech that was time compressed at a 50% time-compression ratio (TCR) throughout the signal. Three conditions applied a 50% TCR to a single segment of each speech signal that was fixed in location at the first, second, or third segment. The final condition applied a 50% TCR to a single segment of each speech signal, but the location of the time-compressed segment varied from trial to trial. The 18 listening conditions were presented in completely random order across listeners, with random assignment of stimulus list to listening condition.

During the experiment, listeners were seated in a double-walled sound-attenuating booth. The stimuli were played back from the digital audio tape recorder, routed to an amplifier (Crown D75), attenuator (Hewlett-Packard 350D), audio mixer-amplifier (Colbourn S82-24), second attenuator (HP 350D), and delivered to the listener through a monaural insert earphone (Etymotic ER3A). The stimuli were presented at a level of 90 dB SPL. This level was selected to insure that the speech signals were presented at a suprathreshold level across frequency (0.25 kHz–4 kHz) to each hearing-impaired listener, and to equate absolute stimulus level for presentation across all subject groups. The ear with better hearing sensitivity or better speech recognition (if thresholds were bilaterally symmetrical) was the test ear for hearing-impaired listeners. The test ear was alternated for listeners with normal hearing. All conditions for experiment 1 were presented in quiet.

The listener's task was to write the entire speech stimulus perceived on an answer sheet. Listeners were encouraged to guess if they were unsure of the spoken message. While the written response may have added a short-term memory component to the task, it insures accuracy of scoring the listener's responses. Additionally, none of the listeners demonstrated any difficulty in providing written responses within the allotted 16-s ISI. Prior to participating in the experimental conditions, listeners heard and responded to a practice tape that included samples of each stimulus form and speech rate manipulation. Testing for each listener was completed in approximately 3 h.

B. Results

Subjects' responses were scored for percentage of correct content words (nouns, verbs, prepositions) recalled within each stimulus list for each condition. Figure 1, panels (a), (b), and (c), presents the mean scores of the four listener groups for sentences, syntactic sets, and random-order words, respectively, in the six speech rate conditions presented in quiet. The scores were analyzed using analysis of variance (ANOVA) with a split-plot factorial design (two between-subjects factors: age, hearing status; two within-subjects factors: stimulus form and time-compression condition) following arc-sine transformation. The results revealed a significant main effect of hearing group ($p<0.05$), with hearing-impaired listeners performing more poorly than listeners with normal hearing. There were also significant main effects of age ($p<0.01$), stimulus form ($p<0.01$), and time-compression condition ($p<0.01$), as well as significant interactions between age and condition ($p<0.01$) and between stimulus form and condition ($p<0.01$).

Simple main effects analyses and multiple comparison tests were conducted to identify the sources of the interaction

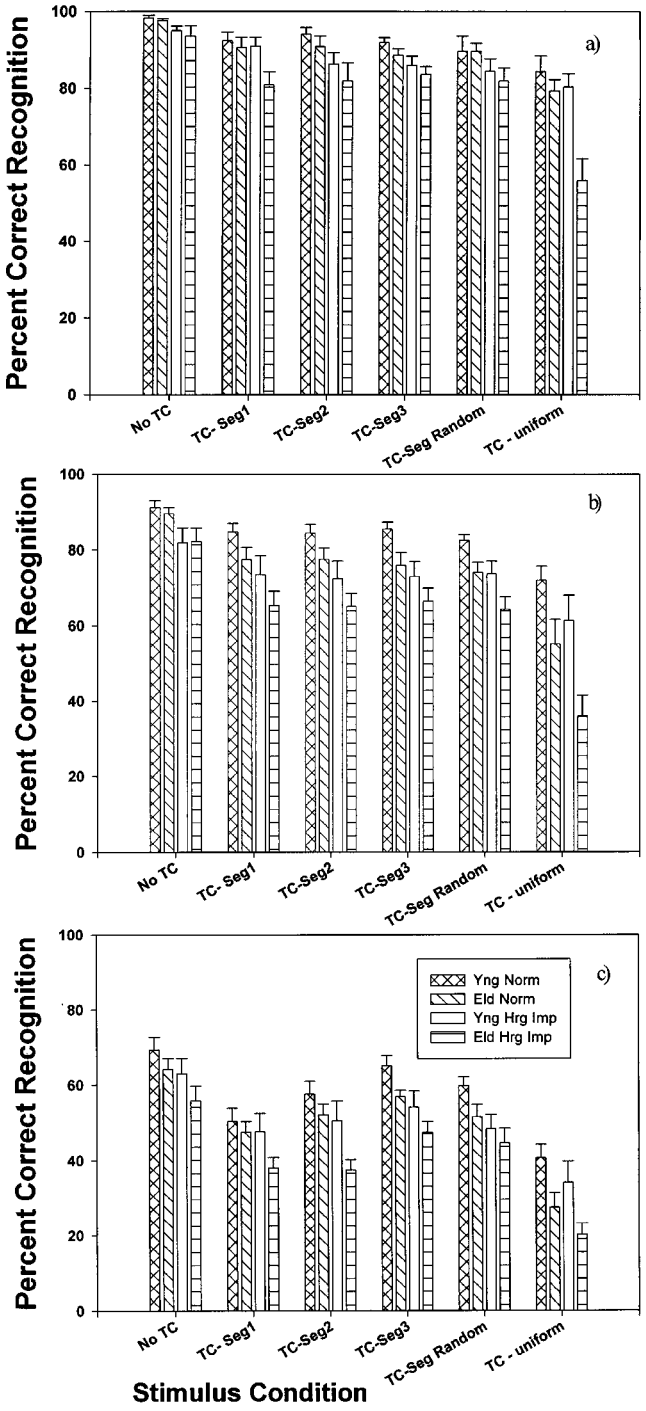


FIG. 1. Mean percent-correct speech recognition scores of the four listener groups (Yng Norm=young normal hearing, Eld Norm=Elderly normal hearing, Yng HI=Young hearing-impaired, Eld HI=Elderly hearing-impaired) for six speech-rate conditions in quiet (No TC=no time compression, TC-Seg1=time compression for segment 1, TC-Seg2=time compression for segment 2, TC-Seg3=time compression for segment 3, TC-Seg Random=time compression for one segment of random location, TC-uniform=time compression implemented uniformly throughout the stimulus), for three stimulus forms [Panel(a)=sentences, panel(b)=syntactic sets, panel(c)=random order words]. Error bars represent the standard error of the mean.

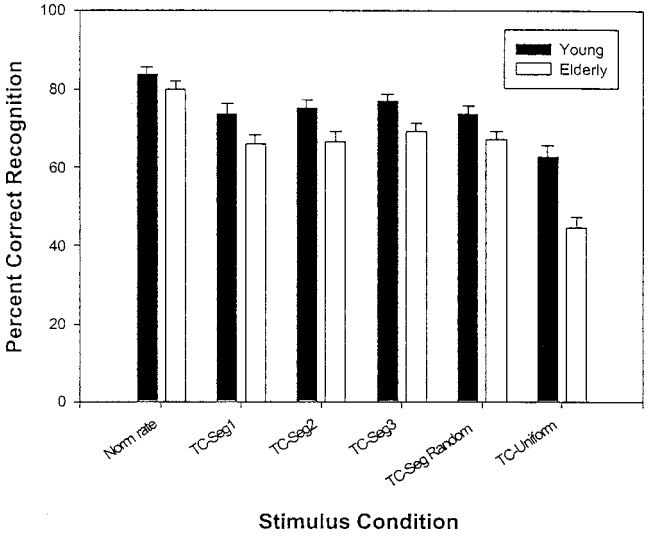


FIG. 2. Mean percent-correct speech recognition scores in the six listening conditions for young and elderly listeners. The data are collapsed across all three stimulus forms, and across the normal-hearing and hearing-impaired groups. Error bars represent the standard error of the mean.

effects. Bonferroni corrections were applied for each set of multiple comparisons to avoid type I errors and set the alpha level at $p<0.01$. To examine the age×condition interaction, average data for the two age groups in the six time-compression conditions (collapsed across hearing loss categories and stimulus forms) are shown in Fig. 2. The statistical analyses revealed that older listeners performed more poorly than younger listeners in all conditions involving time compression, but not in the normal-rate speech condition. Performance in the six time-compression conditions was somewhat different for younger and older listeners. Younger listeners obtained poorer scores in the uniform time-compression condition than in the normal-rate speech condition; none of the other differences was statistically significant for this group. Thus, performance of these listeners in the single-segment time-compression conditions was equivalent across these conditions and equivalent to performance for normal-rate speech and for uniformly time-compressed speech. Older listeners obtained significantly poorer scores in the single segment time-compressed conditions (TC-Seg1, TC-Seg2, TC-Seg3, TC-Seg Random) compared to the normal-rate condition, and significantly poorer scores in the condition in which the entire speech stimulus is time compressed (TC-uniform) compared to all other conditions. The older listeners did not show significant performance differences between the four single-segment time-compression conditions.

Figure 3 presents data averaged across the four groups for the three speech forms and six time-compression conditions, to explore further the form×condition interaction. Simple main effects analyses and multiple comparison tests revealed that for each condition, recognition scores were poorer for the syntactic sets ($p<0.01$) than the original sentences, and poorer for the random-order words ($p<0.01$) than for the original sentences and syntactic sets. The pattern of performance in the different conditions was similar for sentences and syntactic sets: scores were highest in the normal-rate speech conditions, poorer but equivalent in the single-segment time-compression conditions, and poorest in the uniform time-compression condition. For random-order words, a different performance pattern emerged: perfor-

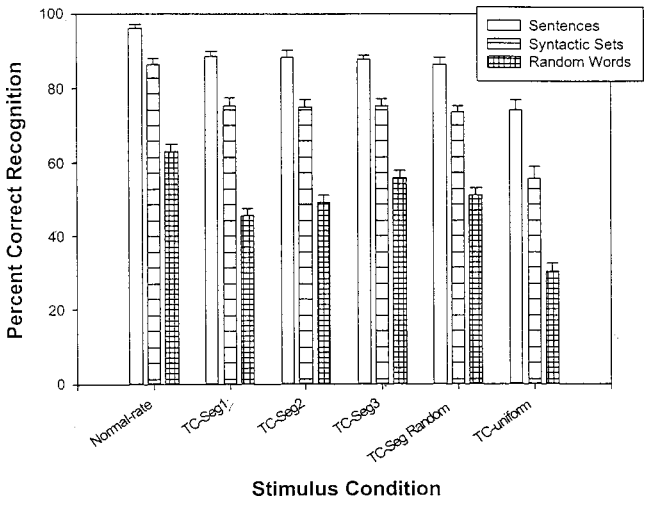


FIG. 3. Mean percent-correct recognition scores in the six listening conditions across the three stimulus forms. Scores are collapsed across the four listener groups.

mance was significantly poorer in the uniform time-compression condition compared to all other conditions, but scores for the normal-rate speech condition were higher than those observed for conditions with time compression of segments 1 or 2 only.

C. Discussion

The main effect of hearing status indicates that hearing-impaired listeners showed significantly reduced scores compared to normal-hearing listeners in all conditions. Although significantly reduced relative to the scores of normal-hearing listeners, the mean speech recognition scores of hearing-impaired listeners exceeded 90% in the normal-rate condition for sentences, reflecting the selection of subjects with good-to-excellent monosyllabic speech recognition scores and the utility of the high speech presentation level. The effect of hearing impairment did not interact with time-compression condition, indicating that the reduced audibility and/or distortion imposed by mild-to-moderate sensorineural hearing loss influenced performance regardless of the presentation rate of the signal. It is also noteworthy that the hearing loss effect did not interact with age effects, suggesting that hearing impairment and age are independent sources contributing to many older listeners' difficulties with understanding speech in time-compression conditions. These findings generally agree with our previous work (Gordon-Salant and Fitzgibbons, 1993).

Age-related deficits were prominent in many of the speech tasks. In particular, older listeners, both normal-hearing and hearing-impaired, performed more poorly than younger listeners in all conditions involving time compression. These findings are consistent with previous reports of older listeners' difficulty with time-compressed speech when the time-compression algorithm is applied throughout the speech stimulus (e.g., Vaughan and Letowski, 1997). The present results also extend these findings to communication situations in which only one segment of a spoken message is incremented in rate. Thus, older listeners experience more difficulty than younger listeners in understanding the message when only one phrase of the message is rapid. Because conversational speech is rarely spoken "recitation style" with uniformity throughout the message, it is likely that everyday speech is characterized by similar fluctuations in presentation rate. Rapidly articulated everyday speech may include additional modifications to the acoustic characteristics of speech (e.g., alterations in closure duration, transition duration, incomplete closure) that can further degrade the signal and reduce speech intelligibility. Nevertheless, the current results indicate that an increment in speech rate, even for a small proportion of a message containing fully articulated phonemes, may be one source of age-related difficulty in understanding everyday speech.

The condition effect varied with subject age. The principal significant effect for young subjects was better recognition of normal-rate speech than uniformly time-compressed speech, suggesting that the overall rapid speech rate condition was challenging for these listeners. However, younger listeners did not show significant decrements in scores for the single-segment time-compression conditions compared to the normal-rate speech condition, nor significant decrements in scores for uniform time-compression compared to the single-segment conditions, suggesting that these listeners do not show a progressive decline in performance with each increment in presentation rate. Performance across conditions was somewhat different for older listeners. The elderly listeners showed best performance for normal-rate speech, poorer performance for single-phrase time-compressed speech, and poorest performance for uniform time-compressed speech. Moreover, there were no differences in performance for the different single-segment time-compression conditions, including the random segment condition. This condition effect suggests that older listeners' performance for time-compressed speech is influenced primarily by the duration of the sentence that is processed by time compression. These results are consistent with the slowing hypothesis (Birren, 1965; Salthouse, 1982; Wingfield *et al.*, 1985), which predicts that the performance of older listeners becomes disproportionately poorer as available processing time is reduced. The results do not support an interpretation based on reduced cognitive function related to adaptation to novel stimuli, as described in the Introduction. If a decline in the ability to adjust to changes in the stimulus speed associated with perceptual normalization were the principal source of the deficit, then poorer performance for a single time-compressed phrase compared to either uniform speech rate condition (normal rate or 50% TCR) would have been observed, as greater adjustment to speech rate would be required for analysis of the more widely disparate time-varying signals. Moreover, poorer performance would be predicted for a single, random time-compressed phrase than a single time-compressed phrase in a fixed location, as additional cognitive abilities would be required to switch attention and process the unpredictable stimulus represented by the random condition. This clearly wasn't the pattern of performance observed. Rather, speech recognition performance of elderly listeners became progressively poorer as the proportion of the sentence affected by time compression was increased, suggesting that it is the overall speed of the sen-

tence that dominates the temporal effects of time-compressed speech for older listeners. Conversely, the present results also indicate that older listeners recognize speech better in conditions with a single speeded phrase, regardless of the location of the single phrase, in comparison to a uniformly speeded message. This indicates that slowing down only a portion of a message may be beneficial to elderly listeners.

The effect of stimulus form was similar across all listener groups and all time-compression conditions. Listeners showed progressively poorer scores as the amount of linguistic information in the stimuli was reduced: recognition scores were highest for sentences in their original word order, scores were reduced for syntactic sets, and scores were poorest for random-order words. Thus, the availability of linguistic contextual cues aided sentence recognition in normal-rate speech conditions and under the adverse listening conditions of accelerated speech rates. The original sentences used in these experiments were the low-probability SPIN sentences, which do not contain semantic contextual cues. The effect of stimulus form in the present experiments suggests that the grammatical rules of sentence structure and word order are important in aiding overall speech recognition, and underscore the value of any type of contextual information for facilitating speech recognition. Younger and older listeners showed the same effect of stimulus form in the various time compression conditions, contrary to previous results (Wingfield *et al.*, 1985). In this earlier study, older listeners exhibited greater detrimental effects of time-compressed speech compared to younger listeners, as the linguistic redundancy in the speech materials was reduced. The disparate results between the two studies may be associated with the amount of semantic contextual information available in the sentence stimuli, the specific time-compression method, or the audiological characteristics of the older listeners. In the present study, the effect of time-compression condition was somewhat different for random-order words compared to the other stimulus forms, with poorer performance observed for recognition of stimuli with time compression of segments 1 or 2, compared to normal-rate speech. The source of this observation is not readily apparent, but may be attributed in part to the difficult nature of recalling a list of unrelated random words and the importance of the initial information in a spoken utterance.

III. EXPERIMENT 2: VARIATIONS IN BABBLE RATE

Experiment 2 was conducted to investigate the hypothesis that temporally mismatched background babble, relative to a target speech signal, would produce less masking than temporally matched babble. The second hypothesis investigated was that older listeners would be less able to take advantage of temporal differences between target signal and background babble than younger listeners.

A. Methods

1. Subjects

The same subjects who participated in experiment 1 also participated in experiment 2.

2. Stimulus materials and background noise

Speech materials consisted of the uniformly time-compressed (50% TCR) sentences, syntactic sets, and random-order words, derived from the LP-SPIN sentences, as described in experiment 1.

A background of multitalker babble, recorded with the R-SPIN test, was presented in all of the conditions. Three babble rates were created by digitizing the 12-talker babble from the original R-SPIN tapes to a PC, and using either the natural rate babble or applying the time-compression algorithm (25% TCR and 50% TCR) to the sampled babble. For each stimulus on each list, brief waveforms of each form of the babble were created so that they were aligned -20 ms with the onset of the target stimulus and $+20$ ms with the offset of the target stimulus (i.e., the onset of the babble preceded the onset of the speech stimulus by 20 ms and the offset of the babble trailed the offset of the speech stimulus by 20 ms). The rms level of each sentence-length babble sample (compressed and uncompressed) was calculated, and the amplitude of each of these waveforms was scaled in dB relative to the rms level of the waveform for a calibration tone. The rms level of each sample remained constant before and after the time-compression processing. Additionally, the amplitude density functions of each babble sample were examined to determine if they were altered systematically by time compression. Kurtosis values were calculated as a measurement of the peakedness of these functions. The mean kurtosis values (and standard deviations) for the normal-rate babble, 25% time-compressed, and 50% time-compressed babble samples were 0.88 (0.64), 0.21 (0.82), and 0.26 (0.79), respectively, and were not significantly different ($p > 0.05$). As noted in experiment 1, the target stimuli were previously scaled in level to be equivalent in rms to that of a calibration tone. The stimuli and background babble (with their respective calibration tones) were played back from separate channels of the PC, converted to analog form, and recorded on separate channels of the DAT recorder.

3. Procedures

The nine new conditions presented in experiment 2 were the three forms of uniformly time-compressed speech presented in three rates of babble: normal-rate babble, time-compressed babble with a 25% TCR, and time-compressed babble with a 50% TCR. The stimuli and noise were played back on separate channels of the DAT, routed to separate attenuators (Hewlett-Packard 350D), mixed and amplified (Colbourn audio mixer amplifier model S82-24), and presented through a single insert earphone. The speech presentation level was 90 dB SPL, and the signal-to-noise ratio was $+12$ dB at the earphone. The order of conditions was randomized across listeners. Listeners were instructed to ignore the noise and write down the entire sentence-length stimulus. The time required for completion of experiment 2 was approximately 2 h.

B. Results

Figure 4, panels (a), (b), and (c), presents the mean recognition scores for uniformly time-compressed sentences,

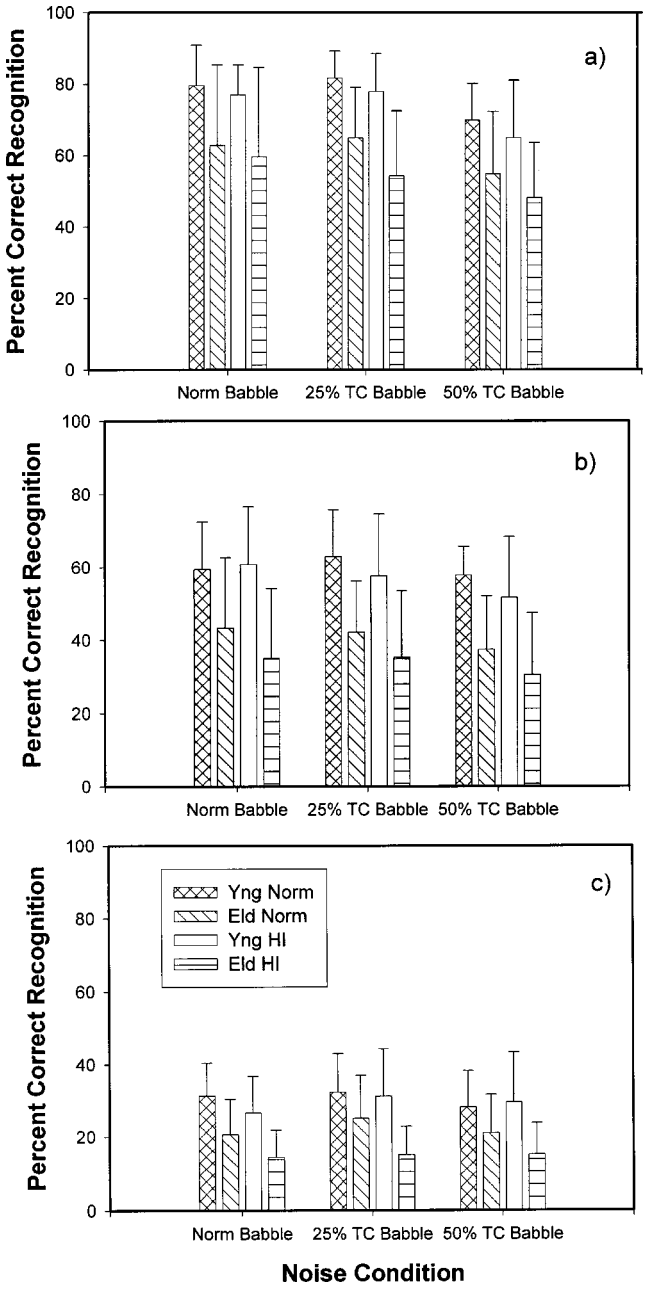


FIG. 4. Mean percent-correct speech recognition scores of the four listener groups for uniform time-compressed speech presented in three noise conditions (normal-rate babble, babble compressed at a 25% time-compression ratio, and babble compressed at a 50% time-compression ratio), for three stimulus forms [Panel(a)=sentences, panel(b)=syntactic sets, panel(c) =random-order words]. Error bars represent the standard error of the mean.

syntactic sets, and random-order words, respectively, in the various noise conditions. ANOVAs were conducted on the arc-sine transformed recognition scores separately for each stimulus form using a split-plot randomized factorial design with two between-subjects factors (age and hearing status) and one within-subjects factor (noise condition). The results showed a significant main effect of age ($p<0.01$) for all three stimulus forms, and a significant main effect of noise condition ($p<0.01$) for the sentences and syntactic sets. The finding that the main effect of age was significant and not involved in any interactions substantiated the observation that older listeners performed more poorly than younger listeners in all conditions. Multiple comparison testing was conducted to analyze further the noise condition effect, and revealed that recognition of rapid speech in time-compressed babble with 50% TCR was significantly poorer than recognition of rapid speech in either normal-rate babble or time-compressed babble with 25% TCR. The effect of hearing status was not significant for any of the stimulus forms.

C. Discussion

Recognition of time-compressed speech in noise appears to be affected by the temporal characteristics of the noise relative to those of the speech signal, particularly for rapid speech with linguistic contextual cues. In the present experiment, listeners performed better in conditions in which the rate of the target speech signal and background babble were mismatched compared to when they were matched. The general pattern of performance, where condition effects were observed, was poor but equal scores in the normal-rate babble and 25% TCR babble conditions, and poorer scores in the 50% TCR babble condition. This pattern of results indicates that listeners compared the overall rates of a target speech signal and the background noise and were able to take advantage of differences in these rates to improve speech recognition. Moreover, the results suggest that listeners were apparently able to extract word-rate variations in a series of babble backgrounds that were all unintelligible. These findings support the predictions based on figure–ground separation, which postulated that listeners would perform better when the presentation rates of speech and noise were different relative to conditions in which the rates were the same. The current experiments provide an initial test of this supposition with one speech rate and several babble rates. Further tests of this hypothesis would require additional speech rates, and background babble rates that are both faster and slower than the target speech rates. An alternative explanation for the condition effect is that increasing the rate of the speech babble created in increase in energetic masking, resulting in poorest performance in the noise condition with 50% time compression of the babble. However, if an increase in energetic masking was the principal underlying factor to account for the results, then the results should show a decrease in recognition performance in the 25% TC babble condition relative to the normal-rate babble condition, assuming a fairly linear relationship between modulation rate and performance. This finding was not observed. Although the samples of normal-rate 12-talker babble contained random dips in the waveform envelope during which there is a temporary increase in the signal-to-noise ratio, we observed no consistent changes in the temporal structure following time compression. However, these few conditions don't provide a systematic examination of the energetic masking hypothesis; further testing with additional, faster babble rates and controlled modulation rates would be needed to provide a stronger test of energetic masking effects.

Younger and older listeners exhibited the same general performance pattern, indicating that both age groups were able to take advantage of the temporal mismatch between target and background. These findings generally agree with those reported by others (Takahashi and Bacon, 1992; Souza

and Turner, 1994; Dubno *et al.*, 2002), who used different types of speech stimuli, noise, and methods to alter the temporal fluctuations of speech and noise. However, it is also noted that older listeners in the present study performed more poorly than younger listeners in all conditions. Thus, while older listeners derived about the same improvement as younger listeners as a result of the temporal alterations in the noise, their overall performance level was depressed relative to that of younger listeners. The source of the age effect in all noise conditions is unknown at present. One possibility relates to the hypothesis that performance on tasks under executive control is reduced in older people. The ability to inhibit the processing of a competing voice in the background while trying to listen to a target speaker is considered an executive function (Tun *et al.*, 2002). The executive control theory predicted that age-related differences would be observed in all conditions involving a background of noise, regardless of its rate. The age effect observed in all noise conditions and for all stimulus forms, with no interaction effects, was consistent with the predictions of the executive control theory. Tun *et al.* (2002) also observed that older listeners performed more poorly than younger listeners in a series of speech recognition tasks involving distracting verbal stimuli. The experimental variable in this previous study was the semantic content of the target signals and competing talkers, rather than presentation rates of target and background. Nevertheless, the impaired ability of older people to inhibit the processing of a speech background, regardless of its semantic content, supported an age-related decline in divided attention and selective listening, two high-level cognitive abilities under central executive control. The findings of the current study are not consistent with the notion that older listeners are less sensitive to changes in the overall presentation rate of speech stimuli than younger listeners. If age-related differences in rate discrimination were a key factor to account for the results, then older listeners would have benefited less from the temporally mismatched conditions relative to the matched conditions, compared to younger listeners.

The effects of noise background varied somewhat with the availability of linguistic cues. Although a noise condition effect was observed for sentences and syntactic sets, this effect was not observed for random-order words, the stimulus form that is devoid of linguistic information. One possible explanation for this finding is that recognition performance for speeded, random words was considerably diminished, and that any words that listeners were able to retrieve among these impoverished stimuli were sufficiently robust to be minimally affected by noise, regardless of its speed. The absence of a condition effect for random-order words is consistent with the notion that all listeners have difficulty inhibiting the processing of irrelevant information when the target signal itself is difficult to recognize.

An effect of hearing status was not observed in the noise conditions, although it was observed consistently in quiet. This difference is probably associated with more extensive differences in the audibility of the speech signals between the normal and hearing-impaired groups in quiet than in noise. The presence of the noise appeared to alter the audibility of the signal more extensively for normal-hearing listeners than for hearing-impaired listeners. For example, for the normal-rate babble only, a *post hoc* analysis indicated a significantly greater decline in performance from the quiet condition to the normal-rate babble condition for normal-hearing listeners compared to hearing-impaired listeners ($p < 0.01$). Thus, the presence of noise minimized the difference in signal audibility (i.e., effective band signal-to-noise ratios necessary for speech recognition) between normal-hearing and hearing-impaired listeners.

IV. SUMMARY AND CONCLUSIONS

The principal findings of the current experiments were as follows.

(1) Older listeners perform more poorly than younger listeners on nearly all speech recognition tasks involving speeded speech, with poorest performance observed for uniform time compression of speech.
(2) Increasing the speed of only one phrase in a sentence is detrimental for older listeners, compared to normal-rate speech.
(3) Time compression of a single phrase presented randomly does not produce poorer scores than time compression of a fixed phrase in the sentence.
(4) Both younger and older listeners show better recognition of time-compressed sentences when the temporal characteristics of target speech signal and background babble are mismatched than when temporal characteristics are matched.
(5) Hearing-impaired listeners perform more poorly than normal-hearing listeners in all normal-rate and speeded speech conditions presented in quiet.

The findings generally support the hypothesis that younger and older listeners are differentially affected by speeded speech tasks, with increased speed of any phrase of a spoken message having a particularly detrimental effect on the performance of older listeners. Moreover, age-related deficits increase with faster presentation rates, which correspond to reduced stimulus durations. These results agree with the slowing hypothesis, which states that reductions in available processing time have an excessive impact on performance of older listeners. Additionally, performance is consistently poorer for older listeners than for younger listeners during speeded speech tasks in a background of speech babble presented at several rates, including a normal-rate and two speeded speech rates. The age-related deficit in the ability to ignore a background of distracting information appears to be consistent with a cognitive decline in executive function. For both younger and older listeners, the interference of background noise is greatest when the temporal composition of the speech and noise is matched, and less so when it is dissimilar, at least for speech stimuli presented at a rapid rate and containing contextual linguistic information. Thus, it appears that both younger and older listeners compare overall rates of the target and background, and are better able to resolve the target signal when its rate is distinct from that of a speech background. This example of auditory

figure–ground separation suggests that overall speech rate relative to background speech rate is a possible cue for improving speech recognition in noise. Generally, the results suggest that older listeners are less adept than younger listeners at adjusting to moment-to-moment fluctuations in speaking rate, but have a comparable ability to younger listeners at taking advantage of a temporal mismatch between signal and noise to resolve the target message. Further research examining the range of speeds of target and noise over which this advantage operates in older listeners would be useful for possible refinement of hypotheses intended to unravel the source of age-related performance deficits in degraded listening conditions.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute on Aging of the NIH (R01AG 09191). The authors are grateful to Jennifer Lantz and Claudia Pastorelli for their assistance in data collection, and to Saul Strieb for his invaluable contributions to stimulus creation. We also thank Tim Bunnell, E.I. Dupont Labs, for his support in our effective utilization of the WEDW software, and four anonymous reviewers for their valuable suggestions on an earlier draft of this manuscript.

ANSI (**1996**). ANSI S3.6-1996, ''American National Standard Specification for Audiometers'' (American National Standards Institute, New York).

Bilger, R. C., Neutzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). ''Standardization of a test of speech perception in noise,'' J. Speech Hear. Res. **27**, 32–48.

Birren, J. E. (**1965**). ''Age changes in speed of behavior: Its central nature and physiological correlates,'' in *Behavior, Aging, and the Nervous System*, edited by A. T. Welford and J. E. Birren (Thomas, Springfield, IL), pp. 191–216.

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

Bryan, J., Luszdz, M. A., and Pointer, S. (**1999**). ''Executive function and processing resources as predictors of adult age differences in the implementation of encoding strategies,'' Aging, Neuropsychol. Cognition **6**, 273–287.

Carhart, R., Tilliam, T., and Greetis, E. (**1969**). ''Perceptual masking in multiple sound backgrounds,'' J. Acoust. Soc. Am. **45**, 694–703.

Dubno, J. R., Dirks, D., and Morgan, D. (**1984**). ''Effects of age and mild hearing loss on speech recognition in noise,'' J. Acoust. Soc. Am. **76**, 87–96.

Dubno, J. R., Dirks, D., and Schaefer, A. (**1989**). ''Stop consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. II. Articulation index predictions,'' J. Acoust. Soc. Am. **85**, 355–364.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). ''Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing,'' J. Acoust. Soc. Am. **111**, 2897–2907.

Gordon-Salant, S., and Fitzgibbons, P. (**1993**). ''Temporal factors and speech recognition performance in young and elderly listeners,'' J. Speech Hear. Res. **36**, 1276–1285.

Gordon-Salant, S., and Fitzgibbons, P. (**1995**). ''Comparing recognition of distorted speech using an equivalent signal-to-noise ratio index,'' J. Speech Hear. Res. **38**, 706–713.

Gordon-Salant, S., and Fitzgibbons, P. (**1997**). ''Selected cognitive factors and speech recognition performance among young and elderly listeners,'' J. Speech Lang. Hear. Res. **40**, 423–431.

Gordon-Salant, S., and Fitzgibbons, P. (**2001**). ''Sources of age-related recognition difficulty for time-compressed speech,'' J. Speech Lang. Hear. Res. **44**, 709–719.

Hasher, L., and Zacks, R. T. (**1988**). ''Working memory, comprehension, and aging: A review and a new view,'' Psych. Learning and Motivation **22**, 193–225.

MacPherson, S. E., Phillips, L. H., and Sala, S. D. (**2002**). ''Age, executive function, and social decision making: a dorsolateral prefrontal theory of cognitive aging,'' Psychol. Aging **17**, 598–609.

Pfeiffer, E. (**1975**). ''A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients,'' J. Am. Geriatr. Soc. **23**, 443–441.

Pichora-Fuller, K., Schneider, B. A., and Daneman, M. (**1995**). ''How young and old adults listen to and remember speech in noise,'' J. Acoust. Soc. Am. **97**, 593–608.

Salthouse, T. A. (**1982**). *Adult Cognition, An Experimental Psychology of Human Aging* (Springer, New York).

Silman, S., and Gelfand, S. (**1981**). ''The relationship between magnitude of hearing loss and acoustic reflex thresholds,'' J. Speech Hear. Res. **46**, 312–316.

Sommers, M. S. (**1997**). ''Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment,'' J. Acoust. Soc. Am. **101**, 2278–2288.

Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1994**). ''Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude,'' J. Acoust. Soc. Am. **96**, 1314–1324.

Souza, P. E., and Turner, C. W. (**1994**). ''Masking of speech in young and elderly listeners with hearing loss,'' J. Speech Hear. Res. **37**, 655–661.

Stuart, A., and Phillips, D. P. (**1996**). ''Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbyacusic listeners,'' Ear Hear. **17**, 478–489.

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.

Tun, P. A. (**1998**). ''Fast noisy speech: Age differences in processing rapid speech with background noise,'' Psychol. Aging **13**, 424–434.

Tun, P. A., O'Kane, G., and Wingfield, A. (**2002**). ''Distraction by competing speech in young and older adult listeners,'' Psychol. Aging **17**, 453–467.

Vaughan, N. E., and Letowski, T. (**1997**). ''Effects of age, speech rate, and type of test on temporal auditory processing,'' J. Speech Lang. Hear. Res. **40**, 1192–1200.

Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (**1985**). ''Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time,'' J. Gerontol. **40**, 579–585.

Quantification of Optison bubble size and lifetime during sonication dominant role of secondary cavitation bubbles causing acoustic bioeffects

Pavel P. Kamaev, Joshua D. Hutcheson, Michelle L. Wilson, and Mark R. Prausnitz[a)]
School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, 311 Ferst Drive, Atlanta, Georgia 30332-0100

(Received 11 June 2003; accepted for publication 8 September 2003)

Acoustic cavitation has been shown to deliver molecules into viable cells, which is of interest for drug and gene delivery applications. To address mechanisms of these acoustic bioeffects, this work measured the lifetime of albumin-stabilized cavitation bubbles (Optison) and correlated it with desirable (intracellular uptake of molecules) and undesirable (loss of cell viability) bioeffects. Optison was exposed to 500 kHz ultrasound (acoustic pressures of 0.6–3.0 MPa and energy exposures of 0.2–200 J/cm^2) either with or without the presence of DU145 prostate cancer cells (10^6 cells/ml) bathed in calcein, a cell-impermeant tracer molecule. Bubble lifetime was determined using a Coulter counter and flow cytometer, while bioeffects were evaluated by flow cytometry. The lifetime of Optison cavitation nuclei was found to decrease and bioeffects (molecular uptake and loss of cell viability) were found to increase with increasing acoustic energy exposure. These bioeffects correlated well with the disappearance of bubbles, suggesting that contrast agent destruction either directly or indirectly affected cells, probably involving unstabilized cavitation nuclei created upon the destruction of Optison. Because Optison solutions presonicated to destroy all detectable bubbles also caused significant bioeffects, the indirect mechanism involving secondary cavitation bubbles is more likely. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1624073]

PACS numbers: 43.80.Gx, 43.80.Sh [FD] Pages: 1818–1825

I. INTRODUCTION

In addition to ultrasound's widespread role as a diagnostic medical tool (Kremkau, 1998), ultrasound is being studied for a number of new therapeutic applications. For example, ultrasound has been shown to increase DNA transfection and protein delivery in cell suspensions and animals (Fechheimer *et al.*, 1987; Miller *et al.*, 1999; Ward *et al.*, 2000; Guzman *et al.*, 2002). Ultrasound has also been shown to induce changes in adhesion, migration and proliferation of sonicated cells that may have use in arterial recanalization (Rosenschein *et al.*, 1995; Alter *et al.*, 1998), enhance antibiotic action against some bacteria (Rediske *et al.*, 1999), selectively trigger apoptosis of malignant cells in blood (Lagneaux *et al.*, 2002), accelerate bone repair (Heybeli *et al.*, 2002), and increase transport across the blood–brain barrier (Mesiwala *et al.*, 2002). Many of these effects are hypothesized to be caused by reparable sonoporation, which involves a temporary disruption of cell membrane structure that subsequently reseals (Ward *et al.*, 2000; Wu *et al.*, 2002). Many of the acoustic bioeffects are believed to be caused primarily by ultrasound-induced cavitation (Leighton, 1994; Barnett *et al.*, 1994).

Acoustic bioeffects have been utilized for ultrasound-mediated drug delivery based on reparable sonoporation. This approach has received increasing attention as a novel strategy to target transport of drugs and genes into cells and tissue (Mitragotri *et al.*, 1995; Keyhani *et al.*, 2001; Pruitt and Pitt, 2002; Taniyama *et al.*, 2002; Zderic *et al.*, 2002). In contrast to systemic procedures, delivery enhanced by ultrasound can be focused onto target tissues. This can reduce the whole body doses of a drug and allow delivery of molecules to precise locations in the body, which may reduce side effects and treatment costs (Miller, 2000).

Ultrasound-mediated delivery can also transport molecules intracellularly (Bao *et al.*, 1997; Guzman *et al.*, 2001a, b), which is often important for gene- and protein-based therapies. Although exciting applications of ultrasound-mediated therapy have been demonstrated, reproducible bioeffects may be difficult to achieve if only acoustic conditions are controlled. Medical applications may require directly controlling cavitation, which depends in a complex time-dependent manner on the acoustic and physical environment. This is because ultrasound's bioeffects are known to be largely mediated by cavitation that takes place even at moderate acoustic pressures in the presence of contrast agents (Ward *et al.*, 2000; Heybeli *et al.*, 2002).

Contrast agents like Optison or Albunex provide nuclei for cavitation in the forms of stabilized gas spheres or pockets of gas and shell fragments released from the disruption of stabilized spheres (Brayman and Miller, 1997; Dalecki *et al.*, 1997; Miller, 1998; Ward *et al.*, 2000; Wu, 2002). Application of ultrasound to these nuclei can cause oscillations and violent collapse of bubbles in the surrounding liquid, microstreaming, shock waves and microjets (Leighton, 1994; Marmottant and Hilgenfeldt, 2003). The availability of cavitation nuclei for interaction with ultrasound determines the ability

[a)]Author to whom correspondence should be addressed. Electronic mail: mark.prausnitz@chbe.gatech.edu

 0001-4966/2004/115(4)/1818/8/$20.00

of sonication to generate inertial cavitation (Chen *et al.*, 2003a). A study with human erythrocytes and 1 MHz ultrasound revealed that gas bodies capable of nucleating violent cavitation activity persist even after the rapid destruction of albumin-covered contrast agent Albunex (Brayman and Miller, 1997). Theoretical considerations predict that free bubbles liberated from contrast-agent gas bodies are much more active than the stabilized shell-encapsulated gas bubbles (Miller, 1998; Wu, 2002). Such bubbles, over a very wide range of bubble sizes, can be highly inertially active even during just one acoustic cycle (Flynn and Church, 1988). Other studies have addressed bubble destruction mechanisms and fragmentation thresholds (Dayton *et al.*, 1999; Chomas *et al.*, 2001; Chen *et al.*, 2003b).

Motivated by the need to characterize and control cavitation activity and, we hypothesize, thereby control bioeffects, the aims of this study were to measure the size and lifetime of cavitation and correlate them with bioeffects on cell suspensions *in vitro*. Bubble lifetime is defined as the duration of sonication after which Optison bubbles can no longer be detected. An additional task was to test the hypothesis that bioeffects caused by ultrasound in the presence of Optison are mediated by secondary cavitation bubbles generated from Optison nucleation sites.

II. MATERALS AND METHODS

A. Cell sample preparation

For experiments involving cells, we used the methods described by Guzman *et al.* (2001a). Briefly, DU145 human prostate cancer cells (ATCC HTB 81, American Type Culture Collection, Rockville, MD) were cultured as monolayers in a humidified atmosphere of 95% air and 5% CO_2 at 37 °C in RPMI-1640 media supplemented with 10% (v/v) heat inactivated fetal bovine serum and 100 μg/ml penicillin-streptomycin (Cellgro, Mediatech, Herndon, VA). Cells were harvested by trypsin/EDTA (Cellgro) digestion, washed, and resuspended at a concentration of 10^6 cells/ml (average cell-to-cell distance of $\sim 80\,\mu$m) in RPMI containing 10 μM calcein (623 Da, radius = 0.6 nm; Molecular Probes, Eugene, OR). Calcein is a green-fluorescent, normally cell-impermeant molecule that was used to quantify transport of molecules into viable cells.

B. Ultrasound apparatus

Ultrasound at 500-kHz frequency, 60-ms pulse length and 6% duty cycle (i.e., 1 pulse per second) was generated by a submersible, focused, piezoceramic transducer with 7.5-cm diameter (Techno Scientific, Woodbridge, Ontario, Canada) housed in a polycarbonate tank containing an acoustic absorber (SC-501 Acoustic Rubber, Sonic Concepts, Woodinville, WA) and filled with de-ionized, distilled and partially degassed water (O_2 levels below 75% saturation) (U.S. Filter, Palm Desert, CA) at 23–24 °C. The transducer had a 9 cm focal length and −6 dB beam width of 3 mm at the focal beam point.

A 500-kHz sinusoid generated by a programmable wave form generator (Stanford Research Instruments, Sunnyvale, CA) and amplified by a custom tone burst amplifier (Techno Scientific) powered and controlled the response of the transducer. The acoustic field generated by the transducer was calibrated using a 0.2-mm aperture PVDF membrane hydrophone (NTR Systems, Seattle, WA). Energy exposure, E, delivered to the focal beam point during the time, t, was calculated as $E = P^2 t/\rho c$, where P is rms pressure, ρ is density of water, and c is speed of sound in water (Leighton, 1994). Spatial-peak-temporal-peak negative pressure was varied between 0.6 and 3.0 MPa.

C. Ultrasound exposure and quantification of bioeffects

Cells in RPMI that contained calcein and 1.7% v/v albumin-stabilized gas bubbles [$\sim 10^7$ bubbles/ml, i.e., cell-to-bubble ratio of 0.1, cell-to-bubble distance of 40 μm (Ward *et al.*, 2000; Guzman *et al.*, 2003)] (Optison, Amersham Health, Princeton, NJ) were slowly injected (to mitigate shear-flow effects on cell membranes and Optison bubbles) into a 1.2-ml polyethylene transfer pipette (Samco, San Fernando, CA, catalog no. 241) and exposed in the focal beam point of the transducer at known acoustic conditions. ''Sham'' exposures to ultrasound (i.e., no ultrasound applied) were conducted as negative controls.

Because the pipette sample cross section (1 by 2.5 cm) was larger than the acoustic beam width, there was a nonuniform exposure inside the pipette. This effect was compensated for by vigorous mixing generated by ultrasound inside the pipette, so that cells and bubbles were continuously cycled through the high pressure region. This mixing was evident by visual observation even on the time scale of a single 60-ms burst. Additional efforts to promote mixing, such as rotation of the pipette (Brayman and Miller, 1999), were not required.

After sonication, samples were immediately transferred to 1.5-ml microcentrifuge tubes (Brinkmann Instruments, Westbury, NY) and left to incubate for 5 min at room temperature to permit cells to ''recover'' (Keyhani *et al.*, 1998). Cells were washed to remove calcein present in the extracellular fluid and subsequently incubated in phosphate-buffered saline containing 0.1 mg/ml propidium iodide (Molecular Probes) to stain nonviable cells. Fluorescent reference beads (Molecular Probes) were added to facilitate cell viability analysis, as described previously by Guzman *et al.* (2001a).

Molecular uptake (i.e., amount of calcein delivered into a cell) and cell viability were quantified by a FACS Vantage SE flow cytometer (Becton Dickinson, San Jose, CA) as described by Guzman *et al.* (2001a). Fluorescence measurements were used to distinguish viable from nonviable cells (PI fluorescence, 665- to 685-nm bandpass filter) and to measure calcein uptake (calcein fluorescence, 515- to 545-nm bandpass filter).

D. Measurement of bubble size, concentration, and lifetime

Before sonication, the concentration of Optison provided by the manufacturer is $5.0–8.0\times 10^8$ bubbles/ml (Amersham Health). Optison bubbles are reported to range in diameter from approximately 1 to 20 μm, with 92.5% smaller than 10

μm (Jablonski *et al.*, 1998), which is consistent with Coulter counter measurements in our study (see below).

To determine the size, concentration, and lifetime of bubbles during and after sonication, we used three different methods. Coulter counting and flow cytometry were used for postsonication measurement and acoustic scattering was used for measurements made during sonication. For Coulter counter and flow cytometer measurements, Optison bubbles were added to RPMI media to achieve a final concentration of 1.7% v/v (unless otherwise noted) and exposed to acoustic conditions over the same range as in the experiments with cells. Samples were then diluted 200-fold with Isoton II solution (Beckman Coulter) (Sontum and Christiansen, 1994) and immediately examined by electrical zone sensing with a Coulter Multisizer II (Beckman Coulter, Fullerton, CA) to determine the size distribution and concentration of bubbles as small as 2 μm, which was the instrument's detection limit. The Coulter apparatus was fitted with a 100-μm orifice, the aperture current was fixed at 1600 μA, the gain factor was set at 2, and the siphon volume was fixed at 500 μl. Each experiment was performed in triplicate. Measurements were made within 5 min after sonication.

To count bubbles with diameters less than 2 μm, a flow cytometer (Becton-Dickinson, LSR I, San Jose, CA) was employed. Flow cytometry light scattering measurements have been used in previous studies to determine the size of particles, and have been shown to yield results similar to electrically based measurements obtained by Coulter counting (Yeng *et al.*, 2002). After exposing Optison to ultrasound as described above, flow cytometry forward scatter measurements were made and compared to polymer beads (Polysciences, Warrington, PA) with known diameters for size calibration.

These experiments to characterize bubble populations were performed in the absence of cells to eliminate difficulty in distinguishing cells and cell debris from bubbles in the Coulter counter. To verify that removal of cells did not alter bubble dynamics, bubble size distribution and concentration were compared using three methods to account for the effects of cells: (i) data from Optison bubbles exposed to ultrasound without cells, (ii) data from Optison exposed to ultrasound with cells, modified by subtracting out data from cells exposed to ultrasound without Optison and (iii) data from Optison exposed to ultrasound with cells, modified by subtracting out data from cells exposed to ultrasound with the ''remnants'' of Optison after exposure to vacuum (30 min at 0.6 atm) to remove bubbles. Bubble concentrations and size distributions were approximately the same for each of these scenarios both before and after sonication at 5 J/cm^2 (data not shown), suggesting that cells did not have a protective (or destructive) effect on bubbles during sonication.

We were concerned that during the 5 min between sonication and analysis, bubble populations could change. In a control experiment, we found that both flow cytometry and Coulter counting showed that Optison bubbles did not change their concentration and size distribution during at least 10 min in mixture with RPMI (data not shown). This suggests that albumin-stabilized Optison bubbles present at the end of sonication were probably unchanged at the time of subsequent analysis. However, we do not have information about the fate of other bubbles that may not have been stabilized by an albumin shell and possibly changed or even disappeared between sonication and analysis.

Due to these limitations of post-exposure analysis of bubble populations, we also used real-time acoustic scattering measurements to provide information about bubble concentration and lifetime (but not size). During sonication, a custom, broadband, piezopolymer hydrophone with active diameter 13 mm (Sonic Concepts) was placed approximately 2 mm behind the sample pipette in line with the transducer. Placing the hydrophone behind the sample pipette could cause acoustic reflections. However, rates of Optison destruction were the same either with or without the hydrophone placed behind the sample, suggesting that the acoustic field was not significantly changed.

The signal from the hydrophone was amplified and acquired with an NI 5911 high-speed digitizer and LabVIEW software (National Instruments, Austin, TX) operating on a personal computer. Average intensity at the fundamental frequency (500 kHz) of every ultrasound burst was estimated using fast Fourier transformation (FFT) of the obtained data. This intensity was used as a relative, inverse measure of bubble concentration when normalized relative to intensity measured using samples without Optison at the same conditions. This approach is based on the expectation that Optison bubbles scatter acoustic energy and thereby reduce the signal received by the hydrophone. We further calibrated by assuming that scattering intensities during the first 100 μs of sonication corresponded to the signal with no bubbles destroyed.

As an additional method to detect bubbles, acoustic signals scattered or emitted from the exposure pipette were collected by a hydrophone (Sonic Concepts) approximately 2.5 cm adjacent to the sample pipette (i.e., at 90° off axis from the transducer) at the location where the maximum signal intensity was recorded. Collected signals were transformed to the frequency domain using FFT as described before.

The broadband noise was used as a measure of inertial cavitation activity (Chen *et al.*, 2003a, 2003b) by measuring the root-mean-squared (rms) intensity of the frequency spectra between 750 and 1000 kHz. This range was large enough to provide meaningful averaging, but sufficiently narrow to avoid the base of the harmonic peaks (Chen *et al.*, 2003a). Broadband noise measurements were normalized relative to the broadband intensity recorded at the same conditions without Optison to account for background noise from the solution without Optison, chamber walls, side-lobes of the transducer, and reflected waves.

E. Statistical analysis

At each condition tested, a minimum of three replicate samples was measured, from which the mean and standard error of the mean (SEM) was calculated. A Student's t-test was used when comparing two experimental conditions. One-way analysis of variance (ANOVA, $\alpha = 0.05$) was performed when comparing one factor with three or more experimental conditions and two-way analysis when comparing two factors. A p-value < 0.05 was considering statistically significant. Data are expressed as mean $\pm$ SEM in the figures.

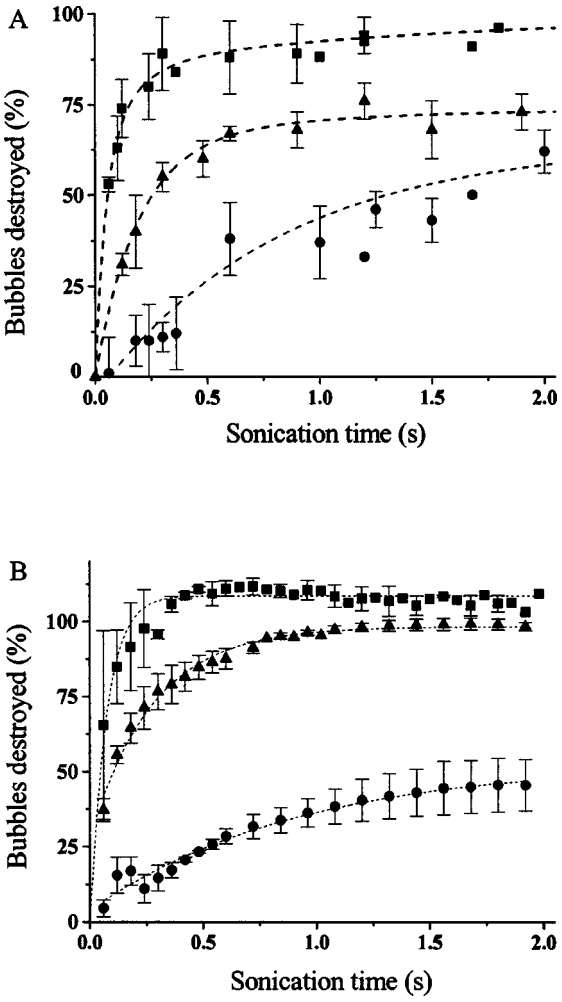


FIG. 1. Optison bubbles destroyed as a function of sonication time at different acoustic pressures: 0.3 (●), 0.5 (▲), 1.8 (■) MPa. Data obtained from (a) Coulter counter 3–5 min after sonication and (b) real-time acoustic scattering measurements during sonication. Data expressed as mean ±SEM. Least-square curve fits are shown to aid viewing.

To identify trends in experimental data, regression models based on restricted cubic splines (S-Plus, MATHSOFT, Seattle, WA) were used. ''Goodness'' of fit for each trend was determined using the multiple R^2 statistic, which represents the amount or variability in the response variable that is explained by the fitted variable. A multiple R^2 of 1 indicates a perfect relationship between the fit and response variables, while a multiple R^2 of 0 indicates no relationship.

III. RESULTS AND DISCUSSION

A. Optison bubble destruction as a function of time and pressure

Previous studies have observed destruction of Optison bubbles due to ultrasound (Dayton *et al.*, 1999; Chomas *et al.*, 2001; Chen *et al.*, 2003b), but have not quantified the kinetics of bubble destruction or its dependence on acoustic parameters. As shown in Fig. 1(a), the application of ultrasound was found to destroy Optison bubbles as a function of time and pressure. Bubble concentrations were determined in Fig. 1(a) by electrical zone sensing using a Coulter counter. At the highest pressure examined (1.8 MPa), bubbles were rapidly destroyed during the first few hundred milliseconds and then the level of destruction plateaued at a value close to 100% destruction (one-way ANOVA, $p>0.10$, for sonication times greater than 120 ms at 1.8 MPa). At 0.5 MPa, and at other intermediate pressures (0.6, 0.8, 1.2 MPa; data not shown to avoid clutter), bubbles were destroyed more slowly and less completely, but the rate of destruction still reached an apparent plateau within a few hundred milliseconds (one-way ANOVA, $p>0.20$, for sonication times greater than 600 ms at 0.5 MPa). In these cases, the plateau value generally increased with pressure and was between approximately 50% and 100% destruction. At the lowest pressure examined, bubble destruction steadily increased and began to plateau near 2 s of sonication (one-way ANOVA, $p>0.15$ for sonication times greater than 1.7 s at 0.3 MPa). When no ultrasound was applied, statistically insignificant numbers of bubbles were destroyed (one-way ANOVA, $p=0.80$; data not shown).

For the data shown in Fig. 1(a), the Coulter counter was capable of measuring the presence of bubbles greater than 2 μm in diameter. Thus, bubbles larger than 2 μm were counted, but bubbles smaller than 2 μm were not detected. To determine if smaller bubbles were present, we used flow cytometry to detect bubbles as small as 0.2 μm in diameter. Using light scattering measurements in a flow cytometer, we found bubble destruction rates similar to those determined by Coulter counting (data not shown), which gives additional confidence in the bubble destruction data shown in Fig. 1(a).

A limitation of measuring bubble concentrations by Coulter counting or flow cytometry is that they required a delay of approximately 3–5 min after sonication before the sample could be carried to the Coulter counter or flow cytometer and the measurement completed. Thus, bubbles that were detected were only those that survived sonication and remained intact afterwards, i.e., intact Optison bubbles and any other bubbles sufficiently stable to survive for a few minutes.

To address this limitation, we also determined rates of bubble destruction using a real-time assay based on acoustic scattering by Optison bubbles. Intensity of the ultrasound signal at the fundamental frequency (500 kHz) transmitted through the sample increases during sonication because the sample becomes more acoustically transparent as Optison bubbles are destroyed. This intensity was used as an inverse measure of bubble concentration when normalized relative to intensity measured using samples without Optison. Figure 1(b) shows bubble destruction kinetics determined by acoustic scattering. Comparison with measurements by Coulter counting [Fig. 1(a)] and flow cytometry (data not shown) indicate that all measurement methods give similar results.

B. Optison bubble destruction as a function of energy

Because rates of bubble destruction depend on exposure time and even more strongly on pressure, we hypothesized that the family of curves shown in Fig. 1 could be collapsed down into a single curve when plotted as a function of acoustic energy exposure. To test this hypothesis, the extent of bubble destruction [measured by Coulter counting, Fig. 1(a)]

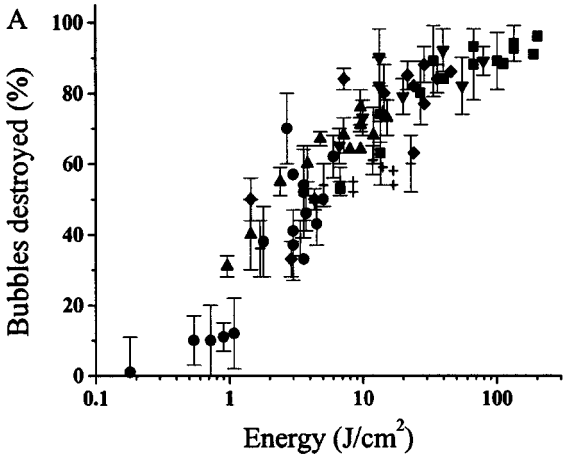


FIG. 2. Optison bubbles destroyed as a function of acoustic energy exposure following sonication at different pressures and exposure times. (a) Coulter counter measurements following sonication of solutions with an initial Optison concentration of 1.7% v.v. and acoustic pressures of 0.3 (●), 0.5 (▲), 0.6 (+), 0.8 (♦), 1.3 (▼), and 1.8 (■) MPa for exposure times ranging from 0.1 to 2.0 s. (b) Flow cytometer measurements following sonication of solutions with initial Optison concentrations of 0.25 (●), 1.7 (■), and 14.4 (▲) % v/v. Acoustic pressures ranged from 0.3 to 1.8 MPa over exposure times ranging from 0.1 to 2.0 s. Data expressed as mean±SEM.

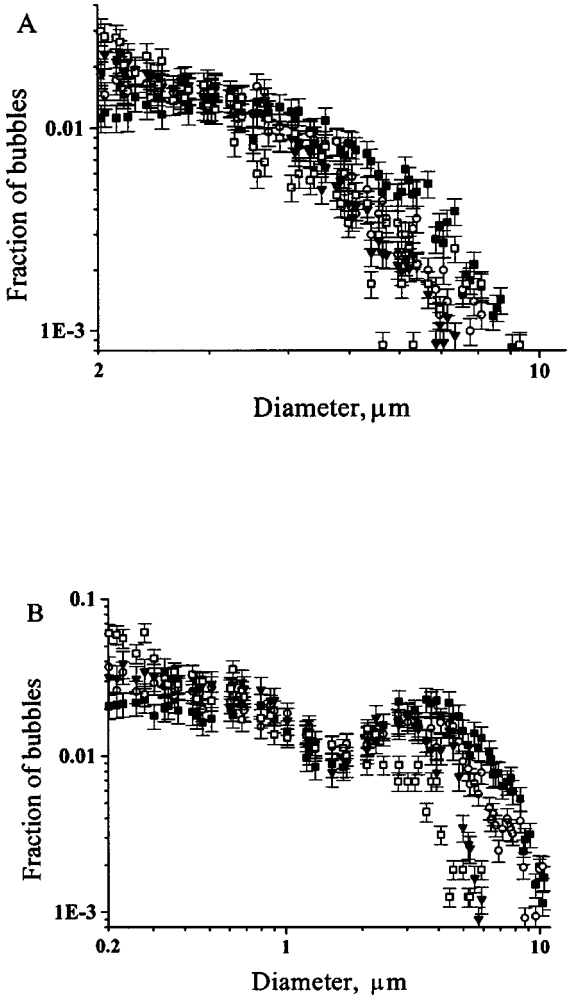


FIG. 3. Histograms of Optison bubble size distribution as determined by (a) Coulter counting and (b) flow cytometry at different energy exposures: 0 (■), 2 (○), 10 (▼), and 25 (□) J/cm². Data expressed as mean±SEM. The visual difference between (a) and (b) is caused primarily by the different scales on the *x* axes.

was replotted as a function of acoustic energy exposure [Fig. 2(a)] and shown to monotonically increase with increasing energy (one-way ANOVA, $p<0.001$). Analysis of data collected by acoustic scattering measurements [Fig. 1(b)] give a similar result (data not shown).

As a further test of correlation with energy, we plotted bubble destruction at three different initial Optison concentrations as measured by flow cytometry as a function of acoustic energy exposure [Fig. 2(b)]. This analysis further showed that the extent of bubble destruction increased with increasing energy (two-way ANOVA, $p<0.001$) and that it also increased with increasing Optison concentration (two-way ANOVA, $p<0.05$).

It is interesting to note that bubble destruction depended on acoustic energy exposure, where almost all bubbles were destroyed at energies of at least 30–50 J/cm² (Fig. 2). This energy level for destruction of the total bubble population is notably close to the optimal exposure conditions observed by Guzman *et al.* (2001a) for good molecular uptake by viable prostate cancer cells using the same apparatus.

C. Change in Optison bubble size distribution

The analysis presented in Figs. 1 and 2 shows that bubble concentration decreases during sonication, but does not address possible changes in bubble size distribution. To our knowledge, changes in Optison bubble size during sonication have not been measured before. As shown in Fig. 3(a) by Coulter counting (size range 2–10 μm) and Fig. 3(b) by flow cytometry (size range: 0.2–10 μm), bubble size distribution also changes with acoustic energy, where the fraction of large bubbles (i.e., >3–5-μm diameter) are preferentially destroyed (two-way ANOVA, $p<0.005$ for bubbles smaller than 3 μm). The visual difference between Figs. 3(a) and (b) is caused primarily by the different scales on the *x* axes. When comparing the size distributions measured over the 2–10-μm range shared by both techniques, the similarity in readings from both methods is more apparent.

For reference, octafluoropropane bubbles suspended in water have a resonant diameter of 11.4 μm at 500 kHz, as

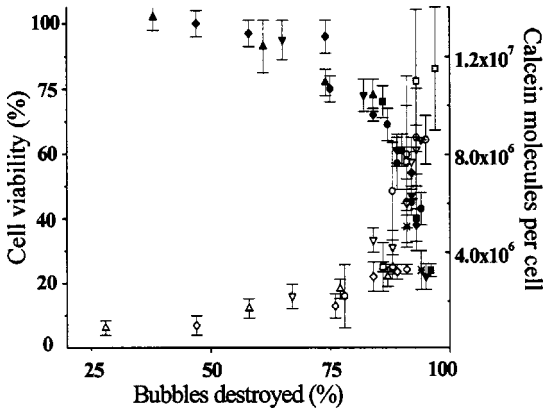


FIG. 4. Cell viability (black symbols) and calcein uptake (white symbols) as a function of Optison bubbles destroyed during sonication (determined by Coulter counting). Acoustic pressures were: (▲, △) 0.6, (◆, ◇) 1.2, (▼, ▽) 1.6, (●, ○) 2, (■, □) 2.4, (∗) 2.8 MPa. Exposure times ranged from 0.1 to 2.0 s. Bioeffect data are from Guzman *et al.* (2001a) and expressed as mean± SEM.

determined using the Minnaert resonant frequency equation (Leighton, 1994) with a constant pressure heat capacity of 114.71 J/mol K (Praxair, Danbury, CT). Bubbles of this size would oscillate more vigorously, thereby breaking their albumin shells and destroying those bubbles preferentially. The resonant size of an Optison bubble might be somewhat larger due to its albumin shell, which probably increases resonant diameter by a factor of about 2 (Church, 1995), Examining the distribution histograms indicates that these large bubbles are replaced to a large extent by ones with diameters just above 2 μm [Coulter counter, Fig. 3(a)] or 0.2 μm [flow cytometry, Fig. 3(b)]. This discrepancy between the two measurements may be explained by noting that these increases in small bubble populations occur just above the detection limits of each device and therefore may represent noise from debris produced by bubble fragments, as opposed to bubbles of those diameters.

D. Correlation of Optison bubble destruction with bioeffects

It has been proposed that destruction of a contrast agent can serve as a nucleation event for secondary bubbles that cause bioeffects (Brayman and Miller, 1997). We therefore tested this hypothesis using the quantitative data set on bubble destruction collected in this study. To relate bioeffects to changes in bubble concentration, we used measurements of intracellular uptake and viability of prostate cancer suspensions collected by Guzman *et al.* (2001a) using the same apparatus and range of ultrasound conditions examined in this study. This dataset shows that uptake of calcein increased and cell viability decreased with acoustic pressure, time, and energy exposure (Guzman *et al.*, 2001a). When these bioeffect data are plotted versus the bubble destruction data measured in the present study, a good correlation (restricted cubic spline multiple $R^2=0.92$ and 0.94 for intracellular uptake and cell viability, respectively) is seen between bioeffects and bubbles destroyed (Fig. 4). Viability initially decreased as a weak function of bubble destruction and then dropped precipitously once most bubbles were destroyed.

TABLE I. Bioeffects produced by ultrasound on cells in solution containing no Optison, presonicated Optison, or fresh Optison.

	Calcein molecules/cell	Cell viability (%)
No Optison	$(1.5\pm0.5)\times10^3$	95± 5
Presonicated Optison	$(3.0\pm0.6)\times10^6$	66± 7
Fresh Optison	$(5.8\pm0.6)\times10^6$	28± 4

Similarly, uptake increased as a nonlinear function of bubble destruction, with a sharp increase once most bubbles were destroyed.

It is interesting to note that when 60%–70% of the bubbles are destroyed, there are significant levels of molecular uptake at high viability. After more than 80% of bubbles are destroyed, viability steeply decreases. This result suggests that destruction of a critical number of nucleation sites is sufficient to obtain desirable bioeffects and further destruction causes large losses in cell viability. Careful regulation of bubble activity could help maintain both molecular uptake and cell viability at high levels during ultrasound exposure.

Bioeffects observed at large acoustic energies, after most detectable Optison bubbles had been destroyed, could arise as a consequence of ''secondary'' microbubbles created when the stabilizing shells surrounding Optison are broken and the encapsulated gas escapes.

To further test the hypothesis that secondary bubbles could cause these bioeffects, we presonicated a suspension of Optison bubbles at 130 J/cm^2 to remove essentially all detectable bubbles. Then, cells were immediately added and the mixture was sonicated again at 130 J/cm^2. Despite the removal of almost all detectable bubbles, cell viability decreased to 66% and calcein uptake reached 3×10^6 molecules/cell (Table I). Sonication at the same ultrasound conditions with ''fresh'' Optison bubble had similar, but more pronounced, effects. In contrast, sonication of cells using media that never contained Optison bubbles at the same ultrasound conditions caused insignificant uptake or loss of viability (Guzman *et al.*, 2001a). Altogether, these observations mean that sonication of cells in the presence of secondary bubbles generated by presonication causes bioeffects greater than when no bubbles were used, but less than when fresh Optison was used. The reduced effect relative to fresh Optison is probably due to the short lifetime of secondary bubbles which can quickly disappear without continuous ultrasound stimulation (Chomas *et al.*, 2001). This experiment indicates that residual material (i.e., secondary bubbles) present after Optison bubbles are destroyed plays an important role in the observed bioeffects, which is consistent with our hypothesis.

To further examine the role of secondary bubbles, we measured broadband noise generated during sonication. Broadband noise is characteristically emitted during collapse of a bubble, which could be during Optison bubble destruction or inertial collapse of secondary bubbles (Chen *et al.*, 2003a, b). Figure 5 shows that even at high pressures and long times, when essentially all Optison bubbles have been destroyed [Fig. 1(a)], broadband noise levels can be high. This indicates that cavitation activity is occurring, presumably due to inertial collapse of secondary bubbles.

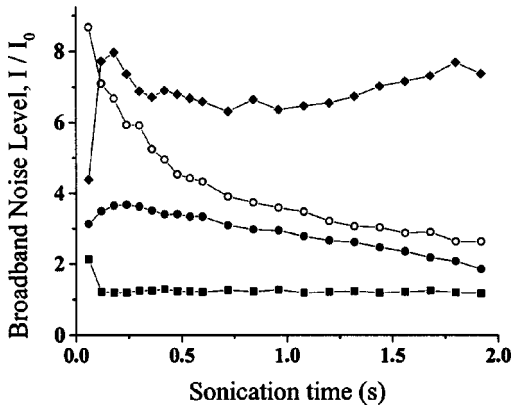


FIG. 5. Relative level of broadband noise emitted during sonication as a function of time at different acoustic pressures: (■) 0.6, (●) 1.6, (○) 2.4, (♦) 2.8 MPa. The reference intensity, I_0, was the broadband noise recorded at the same conditions without Optison. Data expressed as mean ± SEM.

We did not expect, or find, correlation between bubble size distribution and bioeffects, because the bubbles measured in Fig. 3 are not expected to directly affect cells. It is the destruction of these bubbles (i.e., the formation of secondary bubbles) that is bioactive. Secondary bubble size distributions would be more directly relevant to bioeffects, although we were not able to make these measurements.

IV. CONCLUSION

In this study we quantified destruction of Optison bubbles during 500 kHz sonication and correlated those measurements with bioeffects. To facilitate measurements, we showed that flow cytometry could be used to measure Optison bubble concentration and size distribution over a broad range of bubble sizes (0.2–10 μm) and found general agreement with measurements by Coulter counting over the range of overlap (2–10 μm). Using this approach, we determined that Optison bubbles were destroyed as an increasing function of Optison concentration and acoustic pressure and time, as well as acoustic energy exposure, which served as a unifying parameter. Optison bubble size distribution was also shown to change as a function of acoustic energy, where bubbles larger than 3–5 μm were preferentially destroyed.

By comparing to bioeffects data, we found that cell viability and calcein uptake correlated with the lifetime of bubbles, where 60%–70% of bubbles needed to be destroyed before significant bioeffects were observed. Almost all bubbles were destroyed by energies of 30–50 J/cm^2, which is approximately the same energy previously reported to be optimal for producing desirable bioeffects using the same apparatus. These and other measurements suggested that primary Optison bubbles serve as nuclei for free gas microbubbles liberated after rupture of Optison bubbles and these secondary bubbles are responsible for cellular bioeffects. Carefully regulating the kinetics of bubble destruction during sonication should be important to control cell viability and molecular uptake.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Hector Guzman for his helpful discussions. This work was supported in part by the National Science Foundation and the National Institutes of Health.

Alter, A., Roznszajn, L. A., Miller, H. I., and Rosenschein, U. (**1998**). "Ultrasound inhibits the adhesion and migration of smooth muscle cells *in vitro*," Ultrasound Med. Biol. **24**, 711–721.

Barnett, S. B., ter Haar, G. R., Ziskin, M. C., Nyborg, N. L., Maeda, K., and Bang, J. (**1994**). "Current status of research on biophysical effects of ultrasound," Ultrasound Med. Biol. **20**, 205–218.

Bao, S., Thrall, B. D., and Miller, D. L. (**1997**). "Transfection of a reporter plasmid into cultured cells by sonoporation *in vitro*," Ultrasound Med. Biol. **23**, 953–959.

Brayman, A. A., and Miller, M. W. (**1997**). "Acoustic cavitation nuclei survive the apparent ultrasonic destruction of Albunex® microspheres," Ultrasound Med. Biol. **23**, 793–796.

Brayman, A. A., and Miller, M. W. (**1999**). "Sonolysis of albunex supplemented, 40% hematocrit human erythrocytes by pulsed 1-MHZ ultrasound: pulse number, pulse duration and exposure vessel rotation dependence," Ultrasound Med. Biol. **25**, 307–314.

Chen, W.-S., Brayman, A. A., Matula, T. J., Crum, L. A., and Miller, M. W. (**2003a**). "The pulse length-dependence of inertial cavitation dose and hemolysis," Ultrasound Med. Biol. **29**, 739–748.

Chen, W.-S., Matula, T. J., Brayman, A. A., and Crum, L. A. (**2003b**). "A comparison of the fragmentation thresholds and inertial cavitation doses of different ultrasound contrast agents," J. Acoust. Soc. Am. **113**, 643–651.

Chomas, J. E., Dayton, P., Allen, J., Morgan, K., and Ferrara, K. W. (**2001**). "Mechanisms of contrast agent destruction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 232–248.

Church, C. C. (**1995**). "The effect of an elastic solid surface layer on the radial pulsations of gas bubbles," J. Acoust. Soc. Am. **97**, 1510–1521.

Dayton, P. A., Morgan, K. E., Klibanov, A. L., Brandenburger, G. H., and Ferrara, K. W. (**1999**). "Optical and acoustical observations of the effects of ultrasound on contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 220–232.

Dalecki, D., Raeman, C. H., Child, S. Z., Penney, D. P., and Carstensen, E. L. (**1997**). "Remnants of Albunex® nucleate acoustic cavitation," Ultrasound Med. Biol. **23**, 1405–1412.

Fechheimer, M., Boylan, J. F., Parker, S., Sisken, J. E., Patel, G. L., and Zimmer, S. G. (**1987**). "Transfection of mammalian cells with plasmid DNA by scrape loading and sonication loading," Proc. Natl. Acad. Sci. U.S.A. **84**, 8463–8467.

Flynn, H. G., and Church, C. C. (**1988**). "Transient pulsations of small gas bubbles in water," J. Acoust. Soc. Am. **84**, 985–998.

Guzman, H. R., Nguyen, D. X., Khan, S., and Prausnitz, M. R. (**2001a**). "Ultrasound-mediated disruption of cell membranes. I. Quantification of molecular uptake and cell viability," J. Acoust. Soc. Am. **110**, 588–596.

Guzman, H. R., Nguyen, D. X., Khan, S., and Prausnitz, M. R. (**2001b**). "Ultrasound-mediated disruption of cell membranes. II. Heterogeneous effects on cells," J. Acoust. Soc. Am. **110**, 597–606.

Guzman, H. R., Nguyen, D. X., McNamara, A. J., and Prausnitz, M. R. (**2002**). "Equilibrium loading of cells with macromolecules by ultrasound: effects of molecular size and acoustic energy," J. Pharm. Sci. **91**, 1693–1701.

Guzman, H. R., McNamara, A. J., Nguyen, D. X., and Prausnitz, M. R. (**2003**). "Bioeffects caused by changes in acoustic cavitation bubble density and cell concentration: A unified explanation based on cell-to-bubble ratio and blast radius," Ultrasound Med. Biol. **29**, 1211–1222.

Heybeli, N., Yesildag, A., Oyar, O., Gulsoy, U. K., Tekinsoy, M. A., and Mumcu, E. F. (**2002**). "Diagnostic ultrasound treatment increases the bone fracture-healing rate in an internally fixed rat femoral osteotomy model," J. Ultrason. Med. **21**, 1357–1363.

Jablonski, E. G., Dittrich, H. C., Bartlett, J. M., and Podell, S. B. (**1998**). "Ultrasound contrast agents: the advantage of albumin microsphere technology," Rev. Prog. Quant. Nondestr. Eval. **17**, 15–22.

Keyhani, K., Guzman, H. R., Parsons, A., Lewis, T. N., and Prausnitz, M. R. (**2001**). "Intracellular drug delivery using low-frequency ultrasound: quantification of molecular uptake and cell viability," Pharm. Res. **18**, 1514–1520.

Keyhani, K., Parsons, A., Siddeeq, D., Lewis, T. N., and Prausnitz, M. R. (**1998**). ''Size, lifetime, and permeability of ultrasound-mediated cell membrane disruptions,'' Proc. Int. Symp. Contr. Release Bioact. Mat. **25**, 695–696.

Kremkau, F. W., (**1998**). *Diagnosing Ultrasound: Principles and Instruments* (W. B. Saunders, Philadelphia, PA).

Lagneaux, L., de Meulenaer, E. C., Delforge, A., Dejeneffe, M., Massy, M., Moerman, C., Hannecart, B., Canivet, Y., Lepeltier, M. F., and Bron, D. (**2002**). ''Ultrasonic low-energy treatment: A novel approach to induce apoptosis in human leukemic cells,'' Exp. Hematol. **30**, 1293–1301.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic, London).

Marmottant, P., and Hilgenfeldt, S. (**2003**). ''Controlled vesicle deformation and lysis by single oscillating bubbles,'' Nature (London) **423**, 153–156.

Mesiwala, A. H., Farrell, L., Wenzel, H. J., Crum, L. A., Silbergeld, D. L., Winn, H. R., and Mourad, P. D. (**2002**). ''High-intensity focused ultrasound selectively disrupts the blood-brain barrier *in vivo*,'' Ultrasound Med. Biol. **28**, 389–400.

Miller, D. L. (**1998**). ''Frequency relationships for ultrasonic activation of free microbubbles, encapsulated microbubbles and gas-filled micropores,'' J. Acoust. Soc. Am. **104**, 2498–2505.

Miller, D. L., Bao, S., Gies, R. A., and Thrall, B. D. (**1999**). ''Ultrasonic enhancement of gene transfection in murine melanoma tumors,'' Ultrasound Med. Biol. **25**, 1425–1430.

Miller, M. W. (**2000**). ''Gene transfection and drug delivery,'' Ultrasound Med. Biol. **26**, Suppl. 1, S59–S62.

Mitragotri, S., Blankschtein, D., and Langer, R. (**1995**). ''Ultrasound-mediated transdermal protein delivery,'' Science **269**, 850–853.

Pruitt, J. D., and Pitt, W. G. (**2002**). ''Sequestration and ultrasound-induced release of doxorubicin from stabilized pluronic p105 micelles,'' Drug Deliv. **9**, 253–258.

Rediske, A. M., Roeder, B. L., Brown, M. K., Nelson, J. L., Robison, R. L., Draper, D. O., Schaalje, G. B., Robison, R. A., and Pitt, W. G. (**1999**). ''Ultrasonic enhancement of antibiotic action on Escherichia coli biofilms: An *in vivo* model,'' Antimicrob. Agents Chemother. **43**, 1211–1214.

Rosenschein, U., Rozenszajn, L. A., Bernheim, J., Keren, G., Alter, A., Frimerman, A., Laniado, S., Roth, A., and Miller, H. I. (**1995**). ''Safety of coronary ultrasound angioplasty: Effects of sonication in intact canine coronary arteries,'' Cathet Cardiovasc. Diagn. **35**, 64–71.

Sontum, P. C., and Christiansen, C. J. (**1994**). ''Precision and accuracy of analysis of air-filled albumin microspheres using Coulter Multisizer Mark II,'' Pharm. Biomed. Analysis **12**, 1233–1241.

Taniyama, Y., Tachibana, K., Hiraoka, K., and Namba, T. (**2002**). ''Local delivery of plasmid DNA into rat carotid artery using ultrasound,'' Circulation **105**, 1233–1239.

Ward, M., Wu, J., and Chiu, J.-F. (**2000**). ''Experimental study of the effects of Optison® concentration on sonoporation *in vitro*,'' Ultrasound Med. Biol. **26**, 1169–1175.

Wu, J. (**2002**). ''Theoretical study on shear stress generated by microstreaming surrounding contrast agent attached to living cells,'' Ultrasound Med. Biol. **28**, 125–129.

Wu, J., Ross, J. P., and Chiu, J.-F. (**2002**). ''Reparable sonoporation generated by microstreaming,'' J. Acoust. Soc. Am. **111**, 1460–1464.

Yeng, C. H., Anapolski, M., and Cooper, T. G. (**2002**). ''Measurement of volume changes in mouse spermatozoa using an electronic sizing analyzer and a flow cytometer: validation and application to an infertile mouse model,'' J. Androl. **23**, 522–528.

Zderic, V., Vaezy, S., Martin, R. W., and Clark, J. I. (**2002**). ''Ocular drug delivery using 20-kHz ultrasound,'' Ultrasound Med. Biol. **28**, 823–829.

Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length

Marcus Q. Rhinelander[a)]
Department of Marine Science, University of Otago, P.O. Box 56, Dunedin, New Zealand

Stephen M. Dawson[b)]
Department of Marine Science, University of Otago, P.O. Box 56, Dunedin, New Zealand

(Received 2 April 2003; accepted for publication 1 February 2004)

Multiple pulses can often be distinguished in the clicks of sperm whales (*Physeter macrocephalus*). Norris and Harvey [in Animal Orientation and Navigation, NASA SP-262 (1972), pp. 397–417] proposed that this results from reflections within the head, and thus that interpulse interval (IPI) is an indicator of head length, and by extrapolation, total length. For this idea to hold, IPIs must be stable within individuals, but differ systematically among individuals of different size. IPI stability was examined in photographically identified individuals recorded repeatedly over different dives, days, and years. IPI variation among dives in a single day and days in a single year was statistically significant, although small in magnitude (it would change total length estimates by $<3\%$). As expected, IPIs varied significantly among individuals. Most individuals showed significant increases in IPIs over several years, suggesting growth. Mean total lengths calculated from published IPI regressions were 13.1 to 16.1 m, longer than photogrammetric estimates of the same whales (12.3 to 15.3 m). These discrepancies probably arise from the paucity of large (12–16 m) whales in data used in published regressions. A new regression is offered for this size range. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1689346]

PACS numbers: 43.80.Ka [WA] Pages: 1826–1831

I. INTRODUCTION

Sperm whale sounds consist almost exclusively of short-duration, broadband clicks (Watkins, 1980) which almost certainly function in echolocation (Møhl *et al.*, 2000; Jaquet *et al.*, 2001). These clicks form the basis of several types of vocal behaviors defined by click rate (Watkins, 1980; Gordon, 1987; Weilgart and Whitehead, 1988). Only the most common type, ''regular'' or ''usual'' clicks, is analyzed here. Multiple pulses of decaying amplitude, each roughly 1 ms long and 3–7 ms apart, are often present in regular clicks (Backus and Schevill, 1966).

In a remarkable paper published in 1972, Norris and Harvey combined acoustic and morphological evidence to propose a mechanism of sound production in sperm whales. According to Norris and Harvey (1972), impulsive sounds are produced by pneumatic action in the *museu de singe*, an organ at the anterior of the snout, approximately underneath the blowhole (see Clarke, 1979). Part of the acoustic energy of this initial sound radiates forward into the water, forming the first pulse of the click, while the remainder is ducted posteriorly through the spermaceti organ until it reflects off the air-filled nasofrontal sac, on the front of the skull. This reflection passes a second time through the spermaceti organ until part of its energy escapes into the water, forming the second pulse. A portion of the energy of this second pulse is reflected posteriorly by the air-filled frontal sac, again traveling through the spermaceti, continuing the cycle until the sound energy decays. Norris and Harvey proposed that the time between the pulses within a single click (the interpulse interval, or IPI) is the time it takes sound to pass twice through the spermaceti organ, or twice the length of the spermaceti organ times the speed of sound in spermaceti. Because the length of the spermaceti organ is correlated to total length (Nishiwaki *et al.*, 1963) they proposed that IPIs could provide a measure of whale size. A slightly different mechanism for sound generation has been proposed recently (Møhl, 2001; Møhl *et al.*, 2002), but the key point that IPI is twice the travel time between the sound reflectors in the head remains unaltered.

Several studies have used this idea to estimate lengths of sperm whales. While some studies correlated these acoustic measurements with visual estimates (Norris and Harvey, 1972; Møhl *et al.*, 1981), in some cases (e.g., Alder-Fenchel, 1980 and Møhl *et al.*, 1981) lengths calculated acoustically have exceeded the maximum recorded in the literature (18.3 m; Rice, 1989). Of published studies, only Gordon (1991) has analyzed IPIs from whales of independently measured length. He analyzed IPIs of 11 individuals whose lengths had been measured from photographs taken at a known height above the sea surface (see Gordon, 1990). Lengths that he calculated using Møhl *et al.*'s (1981) equation were 3 to 6 meters longer than lengths he measured photographically. He discussed several sources of error that could be responsible for this, and generated a new regression equation to fit the IPI/whale length data he gathered.

While Gordon's (1991) empirical equation fits the size

[a)]Current address: 36 Boston Hill Rd., Andover, Connecticut 06232
[b)]Electronic mail: steve.dawson@stonebow.otago.ac.nz

 0001-4966/2004/115(4)/1826/6/$20.00

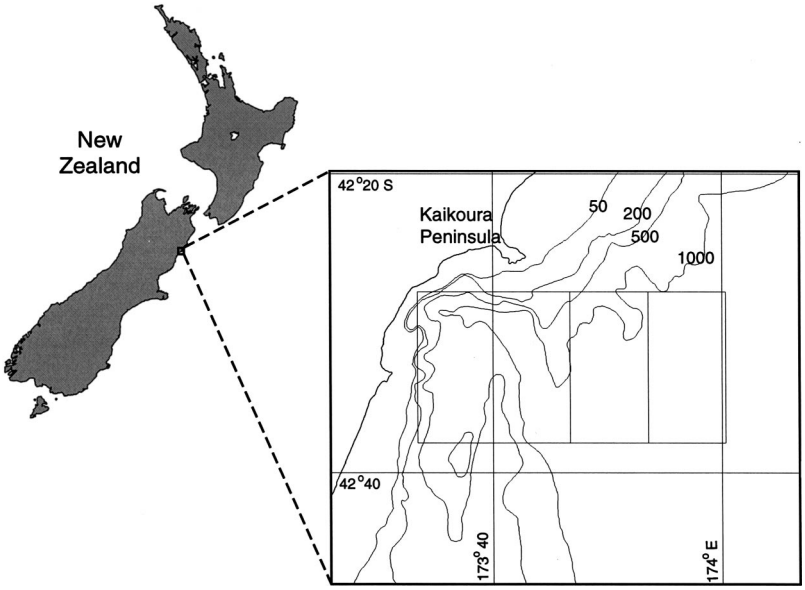


FIG. 1. Map of New Zealand showing Kaikoura, its bathymetry, and our study area (box). Bathymetry units are meters.

range of sperm whales in the tropical Indian Ocean, the limited sample size and the existence in the dataset of only one animal over 12 m long may limit its wider use. Sperm whales show extreme sexual dimorphism; the smaller females and young are mainly confined to the tropics while mature males range to high latitudes (Rice, 1989). More data are needed to derive a reliable relationship between IPI and length for all size classes. Additionally, basic questions about the stability of IPIs within individuals remain unanswered.

At Kaikoura, New Zealand, male sperm whales are present year-round, and are routinely found within a few kilometers of shore (Childerhouse *et al.*, 1995; Jaquet *et al.*, 2000). Many of these animals are larger than those in Gordon's (1991) dataset (up to 15.8 m; Dawson *et al.*, 1995). In this paper, we use recordings of photographically identified individuals, made over multiple dives, days, and years, to assess the stability of IPIs within individuals. We also compare acoustic length estimates with stereo-photogrammetric length measurements of the same individuals.

II. METHODS

Data for this project were collected as part of an ongoing research program off Kaikoura, New Zealand. Fieldwork has spanned two or three field seasons, ranging from 3 to 6 weeks each, every year from 1990–2000. A total of 394 days was spent on the water. Virtually all data were collected in a 10- by 20-nautical mile block south of Kaikoura Peninsula (Fig. 1) from a 6.6-m rigid-hulled inflatable boat powered by a 90-hp outboard motor. Data collection was limited to sea states of Beaufort 4 or below (Beaufort 2 for photogrammetry).

A custom-built directional hydrophone, useful to a range of 3–5 nautical miles, was used to locate and track sperm whales. Fluke photographs were taken at the start of the dive to allow individual identification (Arnbom, 1987) with databack-equipped Nikon F4S, F90x, and F5 cameras and 300-mm lenses. We used a stereo camera system (Dawson *et al.*, 1995) to measure the distance from the blowhole to the emargination of dorsal fin, and extrapolated that to total length using the regression calculated by Best (1990; very similar to the one provided by Gordon, 1991). When conditions permitted stereo photography, we attempted to take three pairs of stereo photographs during each encounter with each whale. We used the best pair (closest, sharpest, least tilt) from any one encounter for analysis. Consistent camera alignment was assured by measurement of a 2.67-m floating PVC bar photographed in triplicate each time the system was used. Measurements were made using a Wild Heerbrug A9 analog stereo plotter. Stereo pairs were measured only if the whale's image occupied $>25\%$ of the frame width and the camera baseline was $<10°$ off parallel.

Acoustic recordings were made either with a Sippican hydrophone connected to a Uhër 4400 report monitor analog tape recorder (1990 to 1995) or with a Sonatech 8178 hydrophone and a Sony TCDD10-PROII Digital Audio Tape (DAT) recorder (1996–present). The frequency response of the Sippican hydrophone was uneven, varying by 20 dB over its range from 200 Hz to 30 kHz. The Uhër recorder was operated at tape speeds of either 1 7/8 or 3 3/4 in. per second, giving frequency responses of ± 3 dB from 25–13 000 Hz and 20–16 000 Hz, respectively. The newer recording chain has a far superior frequency response. The Sonatech 8178 is relatively flat (± 3 dB) from 100 Hz to 180 kHz and the DAT recorder is flat (± 1 dB) from 20 Hz to 22 kHz. The Sippican hydrophone was deployed at 20 m, and the Sonatech hydrophone at 50 or 70 m.

Recordings were considered for analysis only if we gained a high-quality ID photograph before the dive and the recorded vocalizations could be unambiguously attributed to the individual of interest. A signal-to-noise ratio of >10 dB was also required. The portion of the dive suitable for analysis was limited by several factors. The hydrophone was streamed behind the boat while maneuvering for the ID photo. Once the whale fluked, the boat would drift and the hydrophone sink. At shallow hydrophone depths, surface reflections often overlapped with pulses within clicks. To avoid this, we did not analyze the first 2 min of a whale's dive. To minimize changes in IPI caused by increasing pressure (in-

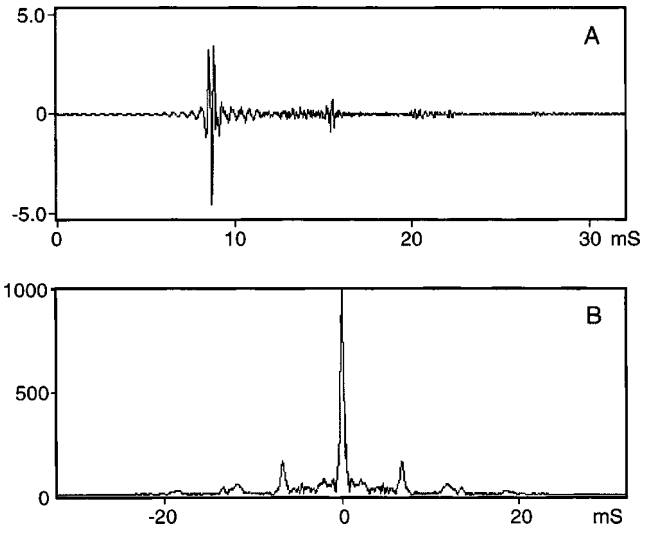


FIG. 2. Sperm whale click after bandpass filtration from 2–13 kHz (A) and its normalized cross-correlation function (B).

creasing speed of sound in spermaceti; Goold *et al.*, 1996), changes in orientation or decreasing signal-to-noise ratio, we did not analyze clicks recorded more than 8 min after fluke-up.

For digitizing, recordings were low-pass filtered with a Kemo VBF-8 Butterworth filter set in series, giving attenuation of 96 dB/octave. A corner frequency of 13 kHz was chosen to reduce the effect of differences between two recording systems and to eliminate aliasing caused by the transition band of the DAT recorder's low-pass filter. Data were digitized in 1-min segments at 44.1 kHz (16 bits) using CANARY 1.2.4 software running on a Macintosh PowerBook G3. Goold and Jones (1995), and our own trials, showed that the 3-6-kHz band revealed IPI structure most clearly, but dramatically reduced the energy in the signal. To compromise, we used a wider passband (2–13 kHz) for digital filtering prior to IPI measurement.

Waveforms of clicks chosen for analysis were cross correlated with themselves using CANARY's built-in correlator. Separate peaks in the correlation function occur at times when the waveform of the signal shows repeated patterns (Fig. 2), such as echoes (Clements, 1998). Cross correlations of sperm whale clicks show peaks at integer multiples of the interpulse interval (Goold, 1996). Thus, the time between the peaks in a cross correlation of a sperm whale click corresponds to the interpulse interval of the whale.

All correlations in this project were the complex envelope of normalized cross-correlation functions. The complex envelope is approximately equal to the amplitude envelope of the absolute value of the correlation function (Charif *et al.*, 1995). Its use reduces the short-wavelength oscillations related to high frequencies in the input signal, which can obscure peaks in correlation values. Correlations were normalized to reduce effects of amplitude variation among clicks. Only correlation functions that showed a single distinct peak, separated from the central peak (at time 0 and with a correlation coefficient of 1) by a distinct space, were accepted for analysis. If accepted, the time and correlation value of the peak were measured with the cursor and stored in CANARY's DataLog and exported for statistical analysis. ANOVA and regression analyses were performed in DATADESK 5.0.1.

III. RESULTS AND DISCUSSION

A total of 888 recordings made at Kaikoura between 1990 and 2000 was available for analysis. Of these, 345 recordings were of 66 photographically identified whales. To examine IPI stability we limited the dataset to individuals from which we had multiple high-quality recordings over at least 3 years.

One hundred and fifty-six 1-min segments were digitized from the recordings. Eighty-nine segments produced correlations that passed our criteria for measurement. Within these, a total of 1997 individual clicks were clear enough for unambiguous measurement of interpulse interval. As in previous studies (e.g., Alder-Fenchel, 1980; Møhl *et al.*, 1981; Gordon, 1991), only a small proportion of the recordings available for analysis contained clicks of a quality suitable for measurement. The number of measurable clicks per 1-min segment varied from 5 to 57.

Three factors made it straightforward to attribute clicks to a particular whale. First, sperm whales at Kaikoura are almost always alone (Childerhouse *et al.*, 1995), and are typically spaced more than 1 nautical mile apart. Second, we only attempted to record the first few minutes of the dive, during which time the target whale's clicks were very obviously louder than those from neighboring whales. Third, though it was seldom necessary, we sometimes tracked the diving whale with our directional hydrophone. Taken together, these factors make it very unlikely that the clicks we measured were from a whale other than the one we identified on fluke-up.

The multiple pulse structure within sperm whale clicks is not always evident if clicks are recorded from head on, but is present at other orientations (Møhl, 2001). If Norris and Harvey's (1972) theory is correct, IPIs of any one individual should be stable over short periods of time, providing that the whale's orientation is consistent. Hence, IPIs measured in the initial stages of the dive (when, in this study, orientation is directly away from the hydrophone) should be consistent over different dives on the same day, and likewise show little change when measured over a single year. Conversely, the theory predicts that IPIs should decrease with depth (due to increasing pressure and thus increasing density and speed of sound in spermaceti; Goold *et al.*, 1996) and increase over longer periods (due to growth). They should differ among animals of different size. We tested these predictions.

A. Are IPIs stable within short (1-min) segments of a dive?

Mean interpulse intervals of 1-min segments of recordings ranged from 6.442 to 7.866 ms. Coefficients of variation (CV) of interpulse interval within segments ranged widely (0.02% to 4.98%), but variation was partly attributable to differences in recording technology. Mean CVs of segments taken from the Uhër recorder were twice those of segments taken from the DAT recorder (mean CV 1.47% and 0.74%, respectively), presumably due to instability in tape speed, to

which DAT recorders are immune. The mean CV from the DAT recordings (0.74%) indicates that the measured interpulse intervals were highly stable over short periods.

B. Are IPIs of an individual stable over different dives in a single day?

Because IPI is predicted to decrease with depth (Goold *et al.*, 1996), and depth is predicted to increase over the first several minutes of the dive, only the third and fourth minutes of dive data were analyzed. High-quality recordings of this segment of multiple dives in a single day were available for four individuals (two dives each on a single day). Two of these showed significant differences (ANOVA: LSD, $p. < 0.01$) while two did not ($p. > 0.9$). In all cases, absolute differences between the mean IPIs of the dives were very small (<1%). Significant variation in IPIs of a given individual over short periods is not expected under Norris and Harvey's (1972) hypothesis, but in this case the magnitude of the variation is more important than its statistical significance. The absolute magnitude of these differences would change acoustic length estimates by a maximum of ca. 15 cm in a 15-m whale, indicating that IPIs of individual animals are stable over different dives on a single day.

C. Are IPIs of an individual stable over different days in a single year?

Variation among different days in a single year should be small, but increase with time between measurements. This was also investigated with ANOVA, and again, only the third and the fourth minutes of the dive were considered. For one individual, IPIs differed significantly between two dives recorded in the same month, though the absolute difference was small (<1%). For the second individual, we have data from six dives on different dates in a single year. IPIs varied significantly over the different dates. Least-significant difference (LSD) *post hoc* tests showed that dives 2 or fewer months apart did not show significant differences in IPI ($p. > 0.1$, absolute differences <1%), while all of the dives longer apart than this were significantly different ($p. < 0.0001$, maximum differences <3%).

D. Do an individual's IPIs increase over several years?

Existing data on growth of sperm whales come from length measurements and tooth layer analysis of animals killed in whaling operations (e.g., Nishwaki *et al.*, 1963). Noninvasive measurement techniques offer the possibility of measuring growth in individuals. ANOVA LSD *post hoc* tests showed that IPIs varied significantly by year for five of the six individuals for which 2 or more years of data existed. The one case that did not show a significant difference (NN80) had only 2 consecutive years of data. Linear regression of IPI against year was used to quantify the form of the change. Regression slopes were significantly different from zero for four individuals, though they varied widely (0.010 to 0.280), as did amount of variation explained (r^2 range 7.1% to 81.4%). The three animals for which IPI regressions were highly significant and explain a large amount of the variation were those for which the longest span of years was available for analysis. Growth curves from Nishiwaki *et al.* (1963) show that a 14-m whale should grow approximately 1 meter in 6 years. Such a size increase equals 7% of the animal's total length and is close to the maximum change we found (6.2%).

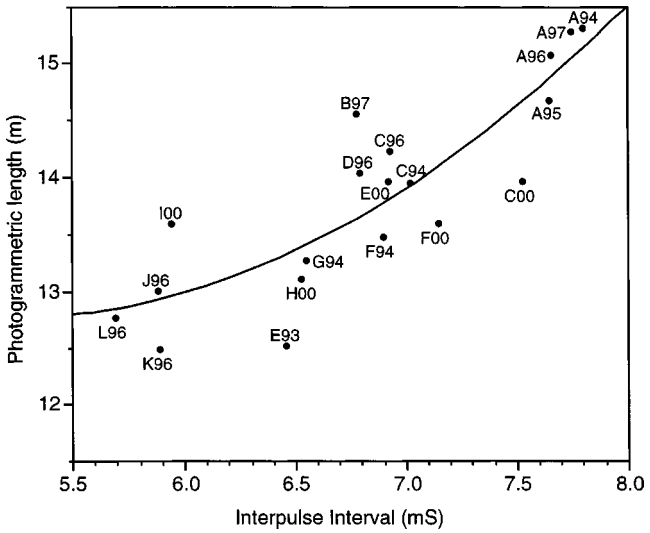


FIG. 3. Relationship between mean interpulse interval and mean photogrammetrically estimated length, for individually identified whales off Kaikoura. The curve fitted is a second-order polynomial (equation given in the text). The letter code indicates the individual, and the number the year that the estimates were made. Individuals are: A (HR80), B (HL120), C (LSL20), D (NN80), E (LSR60), F (LNL100), G (MTR80), H (HR210), I (HL200), J (HL180), K (HR110), L (MTL40).

E. Do IPIs differ among individuals?

Variation of IPIs among individuals was investigated with ANOVA. In a population of different animals, each animal would be expected to differ in length from some, but not all, others. The most recent year for which an estimate existed was used for comparison. As expected, ID was a significant factor in the variation of IPIs among the seven animals for which we had suitable data. LSD *post hoc* tests revealed that of the 36 possible combinations of individuals, only two were not significantly different (NN80-HL120 and NN160-LNL100).

F. How closely do IPIs correlate with independent measures of length?

A fundamental test of Norris and Harvey's (1972) hypothesis is whether IPIs are closely correlated with independent length measurements of the same whales. This analysis included only those individuals for which we had both acoustic and stereo-photogrammetric length estimates within the same year.

Photogrammetric estimates of total lengths ranged from 12.51 to 15.28 m (for individuals measured several times in a year, we used the mean of its measurements for that year). Stereo-photogrammetric data showed greater variability (data range/mean) over different years than the acoustic data for two of the four individuals, but the reverse was true in the other two cases (Fig. 3). The inherent flexibility of the ani-

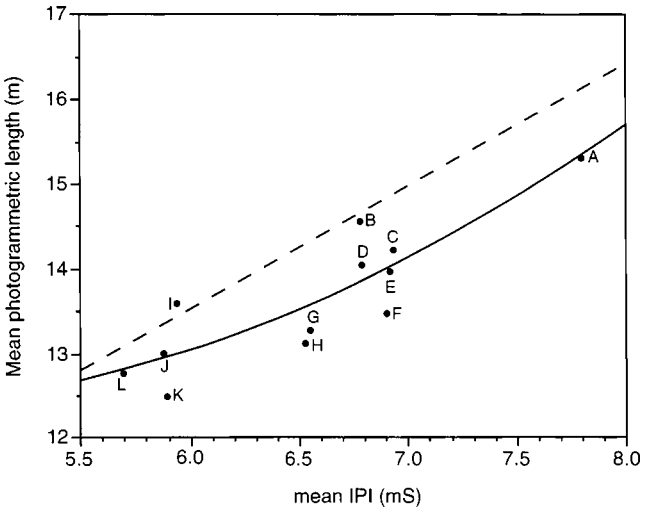


FIG. 4. Relationship between mean interpulse interval and mean photogrammetrically estimated length, for individually identified whales off Kaikoura, showing for each individual only the year for which there are the most photogrammetric data. The curve fitted (solid line) is a second-order polynomial (equation given in the text). Letter codes for individuals are as in Fig. 3. The dashed curve is that of Gordon (1990).

TABLE I. Weighted means of whale lengths calculated from IPIs (via Gordon's 1990 equation) and from stereo photogrammetry.

Whale ID	Year	Acoustic length (m)	Stereo photogrammetric length (m)	Difference
HL120	1997	14.6	14.5	0.1
HL180	1996	13.3	13.0	0.3
HL200	2000	13.4	13.6	−0.2
HR80	1994	16.1	15.3	0.8
HR80	1995	16.0	14.7	1.3
HR80	1996	15.9	15.1	0.8
HR80	1997	16.0	15.3	0.7
HR110	1996	13.4	12.5	0.9
HR210	2000	14.3	13.1	1.2
LNL100	1994	14.8	13.5	1.3
LNL100	2000	15.2	13.6	1.6
LSL20	1994	15.0	14.0	1.0
LSL20	1996	14.9	14.2	0.7
LSL20	2000	15.8	14.0	1.8
LSR60	1993	14.2	12.5	1.7
LSR60	2000	14.9	14.0	0.9
MTL40	1996	13.1	12.8	0.3
MTR80	1994	14.3	13.8	0.5
NN80	1996	14.7	14.0	0.7

mals (which is readily observable from the stereo photographs) would have contributed to this, as would the difficulty of measuring exactly the same points each time. Photogrammetry, despite some imprecision in this application, is the only method that has been shown to make accurate, remote, measurements of objects of known length (Gordon, 1991; Dawson *et al.*, 1995).

Photogrammetric length estimates, and measurements of IPI (from the same year) are available for 12 individuals. For four individuals we have both measurements in 2–4 years, and have plotted the yearly mean of both (Fig. 3). A second-order polynomial ($y=0.371x^2-3.906x+23.056$), used because the proportion of total length occupied by the spermaceti sac increases with increasing length (Fujino, 1956), shows a significant relationship ($r^2=0.743$; $p<0.0001$; Fig. 3). We concede that this calculation contains an element of pseudoreplication, because measurements of photogrammetric length vs IPI of the same individual in different years are unlikely to be statistically independent. Some sources of noise in both the photogrammetric and IPI data have been mentioned before. The four individuals measured over several years were also measured several times in any one year. For Fig. 4, we chose the year with the most photogrammetric data, and used the average of those estimates, so that each whale appears only once. Also, this averaging process should result in a more accurate photogrammetric estimate. The second-order polynomial ($y=0.251x^2-2.189x+17.120$) shows a highly significant relationship ($r^2=0.76$, $p<0.001$; Fig. 4).

Two of the published regressions relating IPI to length (Norris and Harvey, 1972; Møhl *et al.*, 1981) rely on estimates of the speed of sound in spermaceti. Estimates of this value depend on temperature and pressure (Flewellen and Morris, 1978; Goold *et al.*, 1996) and were incorrectly measured in the regression of Norris and Harvey (1972). Mohl *et al.*'s equation, when combined with some of the reported IPI values, produces lengths in excess of the largest recorded for sperm whales. Gordon's (1991) regression equation relating IPI to photogrammetric length is the only published equation to generate length estimates consistently within the known size range of sperm whales. However, our data indicate that it overestimates length in the size range of the whales off Kaikoura. In all but one case, Gordon's equation produced length estimates that were longer than our photogrammetric estimates (Table I; mean=6.2% or 0.86 m; range=−0.2 to +1.8 m). That Gordon's regression contained only one individual longer than 12 m probably explains this discrepancy, and limits its usefulness for measuring males. Likewise, our equation might not apply well to animals smaller than 12 m.

Two age–length keys exist with which to compare the length estimates from this study. Nishiwaki *et al.* (1963) used measurements of 227 individuals from the North Pacific, while Gaskin and Cawthorn (1973) measured 88 whales from the Cook Strait region. While the two keys give very different age estimates for whales over 15 m, there is general agreement on ages corresponding to the lengths found here. Minimum ages derived from our photogrammetric estimates were between 15 (Gaskin and Cawthorn, 1973) and 17 years (Nishiwaki *et al.*, 1963). Physical maturity in males occurs at a mean length of between 15.5 and 15.9 m (Gaskin and Cawthorn, 1973). By this standard, none of the whales we measured at Kaikoura is physically mature. Length at sexual maturity is less clear: estimates range from 9.5 m (Nishiwaki *et al.*, 1963) to 13.8 m (Gaskin, 1970). The data presented here, and those of Dawson *et al.* (1995), suggest that most sperm whales at Kaikoura are sexually mature males, between puberty and physical maturity.

IV. CONCLUSIONS

Our results support the key predictions of Norris and Harvey's (1972) hypothesis that interpulse intervals within

sperm whale clicks can provide a reliable measure of body size in sperm whales. IPIs of individuals, measured near the beginning of dives, are stable over different dives and days. Our data suggest that increases in IPIs over several years are due to growth and that change in IPIs of individuals over several years can be a more sensitive measure of growth than photogrammetric measurements, because of smaller variance. Finally, IPIs differ significantly among whales of different size, and IPIs are significantly related to photogrammetric measures of body length.

ACKNOWLEDGMENTS

This project was part of a research program at Kaikoura that has been funded by the New Zealand Lottery Board, World Wide Fund for Nature (NZ), Telecom NZ Ltd, the University of Otago, Department of Conservation, and the New Zealand Whale and Dolphin Trust. Reckitt and Colman (NZ) Ltd, Ocean Electronics, and Hutchwilco each assisted by sponsoring items of equipment. We are very grateful to all those who helped us collect the data at sea, especially Liz Slooten, Chris Richter, Nathalie Jaquet, Lesley Douglas, and Olaf Jaeke. Otago University's Department of Surveying provided access to the stereo plotter, and Albert Chong advised on its use.

Alder-Fenchel, H. S. (**1980**). ''Acoustically derived estimates of the size distribution for a sample of sperm whales (*Physeter catodon*) in the Western North Atlantic,'' Can. J. Fish. Aquat. Sci. **37**, 2358–2361.

Arnborm, T. (**1987**). ''Individual identification of sperm whales,'' Rep. Int. Whal. Commn **37**, 201–204.

Backus, R. H., and Schevill, W. E. (**1966**). ''Physeter clicks,'' in *Whales, Dolphins and Porpoises*, edited by K. S. Norris (University of California Press, Berkeley), pp. 510–528.

Best, P. B. (**1990**). ''A note on proportional body measurements in sperm whales,'' Paper SC/42/Sp9 presented to the IWC Scientific Committee, June 1990 (unpublished).

Charif, R. A., Mitchell, S., and Clark, C. W. (**1995**). CANARY 1.2 User's Manual (Cornell Laboratory of Ornithology, Ithaca, NY).

Childerhouse, S. J., Dawson, S. M., and Slooten, E. (**1995**). ''Abundance and seasonal residence of sperm whales at Kaikoura, New Zealand,'' Can. J. Zool. **73**, 723–731.

Clarke, M. R. (**1979**). ''The head of the sperm whale,'' Sci. Am. **240**, 128–141.

Clements, M. (**1998**). ''Digital Signal Acquisition and Representation,'' in *Animal Acoustic Communication: Sound Analysis and Research Methods*, edited by S. L. Hopp, M. J. Owren, and C. S. Evens (Springer, Berlin), pp. 27–57.

Dawson, S. M., Chessum, C. J., Hunt, P. J., and Slooten, E. (**1995**). ''An inexpensive, stereophotographic technique to measure sperm whales from small boats,'' Rep. Int Whal. Commn. **45**, 431–436.

Flewellen, C. G., and Morris, R. J. (**1978**). ''Sound velocity measurements on samples from the spermaceti organ of the sperm whale (*Physeter catodon*),'' Deep-Sea Res. **25**, 269–277.

Fujino, K. (**1956**). ''On the body proportions of the sperm whales (*Physeter catodon*),'' Sci. Rep. Whal. Res. Inst., Tokyo **11**, 47–83.

Gaskin, D. E. (**1970**). ''Composition of schools of sperm whales *Physeter catodon Linn*,'' East of New Zealand. NZ J. Mar. Fresh. Res. **4**, 456–471.

Gaskin, D. E., and Cawthorn, M. W. (**1973**). ''Sperm whales (*Physeter catodon* L.) in the Cook Strait Region of New Zealand: Some data on age, growth, and mortality,'' Norw. J. Zool. **21**, 45–50.

Goold, J. C. (**1996**). ''Signal processing techniques for acoustic measurement of sperm whale body lengths,'' J. Acoust. Soc. Am. **100**, 3431–3441.

Goold, J. C., Bennell, J. D., and Jones, S. E. (**1996**). ''Sound velocity measurements in spermaceti oil under the combined influences of temperature and pressure,'' Deep-Sea Res. **43**, 961–969.

Goold, J. C., and Jones, S. E. (**1995**). ''Time and frequency domain characteristics of sperm whale clicks,'' J. Acoust. Soc. Am. **98**, 1279–1291.

Gordon, J. C. (**1987**). ''The behaviour and ecology of sperm whales off Sri Lanka,'' Ph.D. thesis, University of Cambridge, U.K.

Gordon, J. C. D. (**1990**). ''A simple photographic technique for measuring the length of whales from boats at sea,'' Rep. Int. Whal. Commn **40**, 581–588.

Gordon, J. C. (**1991**). ''Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations,'' J. Zool. **224**, 301–314.

Jaquet, N., Dawson, S., and Slooten, E. (**2000**). ''Seasonal distribution and diving behaviour of male sperm whales off Kaikoura: Foraging implications,'' Can. J. Zool. **78**, 407–419.

Jaquet, N., Dawson, S. M., and Douglas, L. (**2001**). ''Vocal behavior of male sperm whales: Why do they click?'' J. Acoust. Soc. Am. **109**, 2254–2259.

Møhl, B. (**2001**). ''Sound transmission in the nose of the sperm whale Physeter catodon. A post mortem study,'' J. Comp. Physiol., A **187**, 335–340.

Møhl, B., Larsen, E., and Amundin, M. (**1981**). ''Sperm whale size determination: Outlines of an acoustic approach,'' FAO Fisheries Series **5**, 327–332.

Møhl, B., Madsen, P. T., Wahlberg, M., Au, W. W. L., Nachtigall, P. E., and Ridgeway, S. H. (**2002**). ''Sound transmission in the spermaceti complex of a recently expired sperm whale calf,'' ARLO **4**, 19–24.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648.

Nishiwaki, M., Ohsumi, S., and Maeda, Y. (**1963**). ''Change of form in the sperm whale accompanied with growth,'' Sci. Rep. Whal. Res. Inst., Tokyo **17**, 1–13.

Norris, K. S., and Harvey, G. W. (**1972**). ''A theory for the function of the spermaceti organ of the sperm whale (*Physeter macrocephalus*),'' in Animal Orientation and Navigation, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville (NASA, Washington D.C.), Special Publication, pp. 397–419.

Rice, D. W. (**1989**). ''Sperm whales (*Physeter macrocephalus*),'' in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 177–223.

Watkins, W. A. (**1980**). ''Acoustics and the behavior of sperm whales,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 283–290.

Weilgart, L. S., and Whitehead, H. (**1988**). ''Distinctive vocalizations from mature male sperm whales (*Physeter macrocephalus*),'' Can. J. Zool. **66**, 1931–1937.

Low-frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean

Sharon L. Nieukirk
Cooperative Institute for Marine Resources Studies, Oregon State University, Hatfield Marine Science Center, 2030 S. Marine Science Drive, Newport, Oregon 97365

Kathleen M. Stafford
National Marine Mammal Laboratory, 7600 Sand Point Way NE, Seattle, Washington 98115

David K. Mellinger and Robert P. Dziak
Cooperative Institute for Marine Resources Studies, Oregon State University, Hatfield Marine Science Center, 2030 S. Marine Science Drive, Newport, Oregon 97365

Christopher G. Fox
NOAA Pacific Marine Environmental Laboratory, Newport, Oregon 97365 and Dept. of Commerce/NOAA/NGDC, 325 Broadway, E/GC, Boulder, Colorado 80305-3328

(Received 12 September 2003; revised 15 January 2004; accepted 19 January 2004)

Beginning in February 1999, an array of six autonomous hydrophones was moored near the Mid-Atlantic Ridge (35 °N–15 °N, 50 °W–33 °W). Two years of data were reviewed for whale vocalizations by visually examining spectrograms. Four distinct sounds were detected that are believed to be of biological origin: (1) a two-part low-frequency moan at roughly 18 Hz lasting 25 s which has previously been attributed to blue whales (*Balaenoptera musculus*); (2) series of short pulses approximately 18 s apart centered at 22 Hz, which are likely produced by fin whales (*B. physalus*); (3) series of short, pulsive sounds at 30 Hz and above and approximately 1 s apart that resemble sounds attributed to minke whales (*B. acutorostrata*); and (4) downswept, pulsive sounds above 30 Hz that are likely from baleen whales. Vocalizations were detected most often in the winter, and blue- and fin whale sounds were detected most often on the northern hydrophones. Sounds from seismic airguns were recorded frequently, particularly during summer, from locations over 3000 km from this array. Whales were detected by these hydrophones despite its location in a very remote part of the Atlantic Ocean that has traditionally been difficult to survey. [DOI: 10.1121/1.1675816]

PACS numbers: 43.80.Ka, 43.30.Sf [WA] Pages: 1832–1843

I. INTRODUCTION

Passive acoustic experiments have become an important tool in surveying remote areas of the sea that are difficult to investigate by more traditional techniques. Acoustic surveys of cetacean habitat are a powerful means of identifying the species present (Clark *et al.*, 1996; Clark and Charif, 1998; Stafford *et al.*, 1999; Watkins *et al.*, 2000), locating and tracking individuals (Clark *et al.*, 1996; Clark and Fristrup, 1997; McDonald *et al.*, 2001), identifying sounds associated with different regions (Stafford *et al.*, 1999; 2001), and determining patterns of seasonal distribution and relative abundance (Thompson *et al.*, 1992; Clark *et al.*, 1996; Moore *et al.*, 1998, Stafford *et al.*, 1999, 2001; Mellinger *et al.*, 2004). The extent to which this information can be obtained from acoustic data depends largely on study design, including locations and dates of recordings, instrument type (autonomous moored instrument, vessel-deployed hydrophones, etc.), sampling rate of recordings, and types of nonacoustic data collected concurrently.

In 1999, a consortium of U.S. investigators deployed an array of autonomous hydrophones (Fox *et al.*, 2001) to monitor seismic activity along the Mid-Atlantic Ridge south of the Azores (Smith *et al.*, 2002; Fig. 1). Although this experiment was designed to monitor low-frequency earthquakes, the instruments were also capable of recording the low-frequency calls of several species of balaenopterid whales (Payne and McVay, 1971; Winn and Perkins, 1976; Watkins, 1981; Edds, 1982; Clark, 1994). These hydrophones were located within potential migratory corridors for at least two species of baleen whales (Charif *et al.*, 2001), in an area far offshore that is not often covered by marine mammal surveys (Mellinger and Barlow, 2003, p. 23). The aim of the acoustic analyses described here was to document seasonal occurrence in the central North Atlantic of several cetacean species' vocalizations. This information will complement existing North Atlantic acoustic data on balaenopterids (Clark, 1995; Clark and Charif, 1998; Charif and Clark, 2000; Mellinger *et al.*, 2000; Charif *et al.*, 2001; Mellinger and Clark, 2003) and augment visual survey and historical whaling data on the seasonal movements and distributions of whales in the North Atlantic.

II. BACKGROUND: TYPES OF SOUNDS

The autonomous hydrophones used in this study were deployed for an experiment designed to monitor seismic activity along the Mid-Atlantic Ridge, and were configured to record sound frequencies only between 1 and 50 Hz. This includes the frequency range of the most common sounds of

blue and fin whales, as well as a portion of the frequency range of minke and humpback whale vocalizations. Sounds associated with Bryde's whales in the Atlantic are documented only from the Caribbean Sea and are known to vary geographically (Oleson *et al.*, 2003), so it was unclear what types of vocalizations might be present in the recordings. In addition, it was uncertain how to distinguish Bryde's whale sounds from other ocean sounds in the frequency band available. For these reasons, Bryde's whale sounds were excluded from the analysis. Sounds of sperm whales were above the frequency range of the instruments, and sounds of sei whales are poorly known, so these species were likewise excluded. In addition to natural sounds, manmade noise such as ship and seismic profiling sounds could be recorded in this frequency band. Because of increased interest in such manmade sounds (NRC, 2000, 2003), the distinctive signals of seismic airguns were analyzed.

The 20-Hz pulses of fin whales in the Atlantic and Pacific Oceans have been described in detail elsewhere (Thompson *et al.*, 1979; Watkins, 1981; Watkins *et al.*, 1987; Edds, 1988; Thompson *et al.*, 1992; Clark *et al.*, 2002). Typically the pulses are tones sweeping from 25–44 Hz down to 16–20 Hz over 0.5–1 s that occur with regular interpulse spacing (Watkins, 1981; Thompson *et al.*, 1992). Series of pulses occur in long, patterned, song-like sequences that change with geographic location and possibly with time (Cummings *et al.*, 1986; Watkins *et al.*, 1987; Thompson *et al.*, 1992; Clark *et al.*, 2002).

North Atlantic blue whale vocalizations were first described by Edds (1982) from the Gulf of St. Lawrence. Typically these are long, patterned sequences of sounds in the 15–20-Hz frequency band. Three types of sounds were described by Mellinger and Clark (2003). The first consists of a two-part, A–B phrase, with part A an 8-s tone, followed 0–5 s later by part B, an 11-s frequency-modulated downsweep. Usually these sounds appeared as A–B pairs, but sometimes sequences containing only part A, and occasionally only part B, were recorded. The second type of sound attributed to blue whales was a very short (2–5 s), quiet, 9-Hz tone, while the third was an arch-like sound that started at about 55 Hz, swept up to 70 Hz, then descended to 35 Hz. By far the most common call type identified by Mellinger and Clark (2003) was the A–B pair.

Series of low-frequency pulsed sounds, and sequences of these sounds (pulse trains), have been reported from minke whales (Winn and Perkins, 1976; Swift *et al.*, 1996). These pulsed sounds and pulse trains have been recorded in and near the Caribbean (Winn and Perkins, 1976; Mellinger *et al.*, 2000), the western North Atlantic (Clark, 1994), and in the St. Lawrence Estuary (Edds-Walton, 2000). Pulse trains have typically been characterized as decelerating series of pulses, although pulse rates in observations from other data sources vary, speed up, slow down, or remain constant (DKM, personal observation, 1999). The frequency range reported by Winn and Perkins (1976) was from below 100 Hz up to at least 800 Hz, although these data were filtered with a high-pass cutoff frequency of 100 Hz so the true lower-frequency component was not identified.

Song is the best known and most complex of the sounds made by humpback whales. It consists of a series of variable sounds that occur in repetitive patterns known as units, phrases, and themes (Payne and McVay, 1971). Humpback song changes throughout the winter–spring display season, yet all whales in a population seem to make these changes at approximately the same time and sing very similar songs (Payne and Payne, 1985; Guinee *et al.*, 1983). Song characteristics have been used to identify population differences (Payne and Guinee, 1983; Winn *et al.*, 1981; Helweg *et al.*, 1990, 1998). The true function of song is unknown, but it is currently thought to be a type of male display, as all identified singers to date have been male (Tyack, 1981; Clapham and Mead, 1999). Humpback whales sing primarily on the wintering grounds (Payne and McVay, 1971; Tyack, 1981), but occasionally song or parts of song are recorded on the feeding grounds and along the whales' migratory route (Clapham and Mattila, 1990; Mattila *et al.*, 1987; McSweeney *et al.*, 1989; Norris *et al.*, 1999). Songs typically range in frequency from less than 20 Hz to over 4 kHz (Thomson and Richardson, 1995) and may be composed of a series of grunts, squeaks, moans, or other sounds. Less frequently recorded are the 50- to 10-kHz sounds made within social groups on the wintering grounds (cf. Tyack, 1983; Silber, 1986; Tyack and Whitehead, 1983) and the sounds associated with feeding (Thompson *et al.*, 1986; Cerchio and Dahlheim, 2001).

Sounds associated with seismic exploration, particularly sounds from airgun arrays, have garnered increasing interest recently as there are concerns regarding the potential impact of airgun noise on marine mammals (NRC, 2003). The loud impulses produced by airguns are created as air, pressurized within cylinders, is released suddenly into the water (Parkes and Hatton, 1986; Dragoset, 2000). The expansion of this air mass and the following contraction and re-expansion create loud explosive sounds of very short duration and broad frequency which are used to probe rock layers beneath the seafloor. The sound-pressure source level (SPSL) of a single airgun ranges from 216–232 dB re: 1 μPa at 1 m, while arrays of up to 70 airguns can produce SPSLs of up to 259 dB re: 1 μPa at 1 m (Richardson *et al.*, 1995). Typically the sounds associated with both commerical and research airguns occur repetitively every 10–20 s over a time span of days to weeks, with occasional interruptions for such actions as turning the ship that tows the airgun array.

III. METHODS

In February 1999, six autonomous hydrophones were moored along the Mid-Atlantic Ridge between 15°–35 °N and 33°–50 °W (Fig. 1). Hydrophones are referred to by geographic location as NW, NE, CW, CE, SW, SE (Table I). These hydrophones monitored sound continuously, using a low-pass filter cutoff centered at 50 Hz to reduce signal aliasing, and recorded the filtered signals to disk at a sampling rate of 110 Hz. Each mooring package consisted of an anchor, an acoustic release, an autonomous hydrophone logging system, and flotation. The logging system was composed of an International Transducer Corporation 1032 hydrophone, a preamplifier and filter designed to prewhiten ocean ambient noise spectra from 1–50 Hz, a digital re-

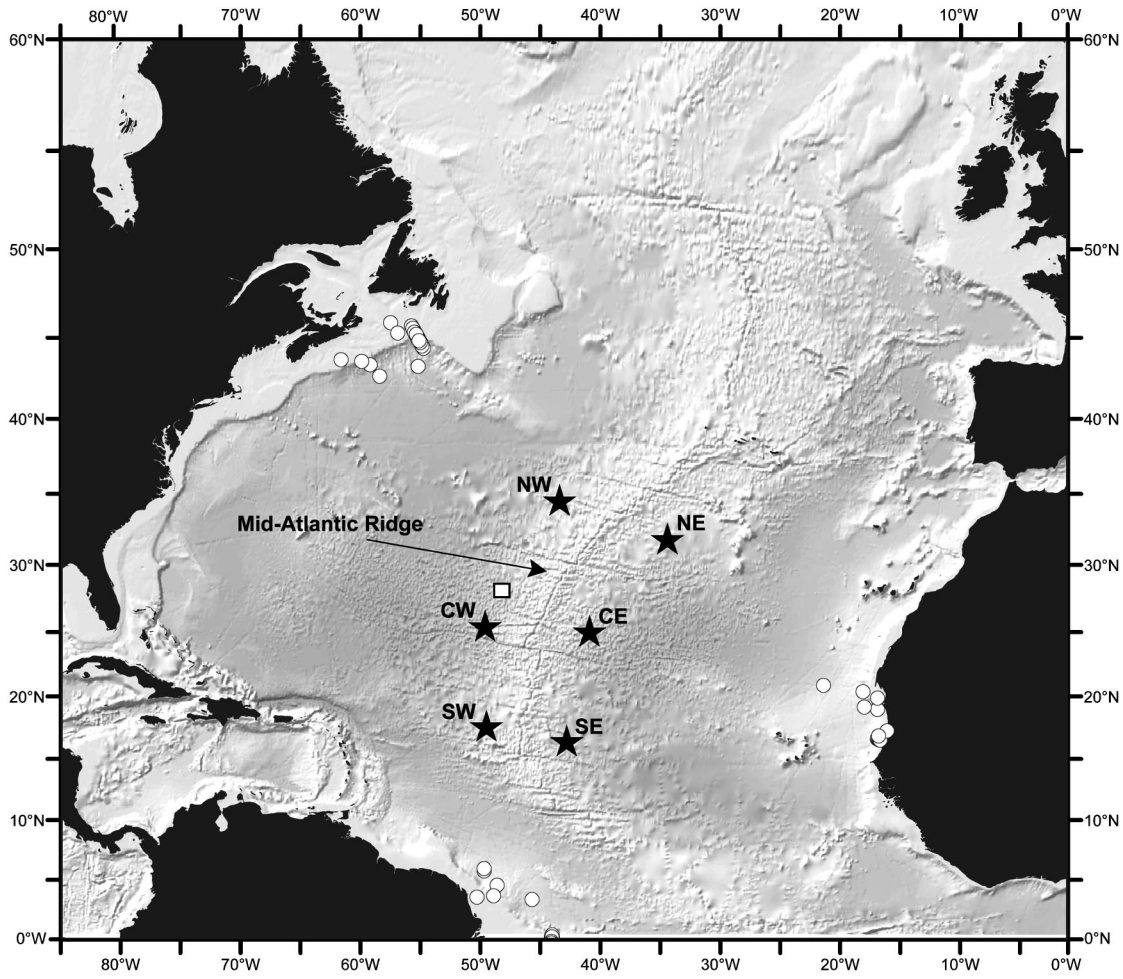


FIG. 1. Locations of six autonomous hydrophones (filled stars) moored along the Mid-Atlantic Ridge, and approximate locations (circles) of seismic airgun activity located via the array. The square represents the location of research airgun activity during the summer of 1999.

corder, and a pressure-resistant titanium case (Fox *et al.* 2001). The instruments were moored above the seafloor such that the hydrophones were suspended near the deep sound-channel axis at depths of about 800 m (Table I). The hydrophone spacing, approximately 700–800 km, was such that vocalizations from an individual whale would rarely be recorded on more than one hydrophone simultaneously, and locating vocalizing animals was not attempted. Each instrument was designed to record for just over 12 months. The data were archived on each instrument's hard drive until recovery, at which time the disks were replaced and the instrument redeployed for another 12 months.

TABLE I. Approximate locations, bottom depth, and mooring depth of the six autonomous hydrophones moored near the Mid-Atlantic Ridge. Instrument depth is the number of meters from the surface and varies with year due to differences in mooring line length.

Hydrophone	Approximate mooring location	Approximate bottom depth (m)	1999–2000 instrument depth (m)	2000–2001 instrument depth (m)
SE	16 °N 43 °W	4565	865	767
SW	16 °N 49 °W	4715	815	746
CE	26 °N 40 °W	5105	905	866
CW	26 °N 50 °W	5182	982	874
NE	32 °N 35 °W	3927	927	925
NW	35 °N 43 °W	4179	679	686

The recovered data were monitored for whale vocalizations by visual examination of spectrograms. Continuous time–frequency spectrograms (frame and FFT size 256 samples (2.29 s), overlap 50%, Hanning window, effective filter bandwidth 1.8 Hz) of the acoustic signals recorded from the six hydrophones were displayed via a program developed by National Oceanographic and Atmospheric Administration/Pacfic Marine Environmental Laboratory (NOAA-PMEL) written in IDL® (Interactive Data Language, Research Systems, Inc., Boulder, Co). Data for each day from Feb. 1999 to Mar. 2001 were visually examined for biological sounds, seismic activity, and manmade noise. Sounds were identified as blue, fin, or minke whale vocalizations based on their similarity to published sounds, as specified above. One other sound was identified as biological in origin because it was above the frequency band of seismic sounds, which are typically less than 20 Hz; was frequency modulated; and occurred in an irregular temporal pattern. Presence or absence of all of these sound types was recorded for each hour of data examined, and the resulting data were aggregated monthly to show the percentage of hours in each

month in which each sound type was present at least once. Recognizing that call identification can be somewhat subjective, 20 percent of the data were randomly selected and call identifications were verified by a second analyst experienced in identifying whale vocalizations. Sounds of seismic airguns were identified by their broadband impulse character, their high degree of regular repetition, and their regular occurrence for hours without stopping.

Sounds with high signal-to-noise ratio and few or no interfering sounds were selected for measurement of time and frequency characteristics. The detected biological sounds occurred in series, usually with a very regular repetition interval, and were inferred to be from one whale or one group of whales. The series would sometimes be interrupted by silent intervals that have been associated with surfacing to breathe (Cummings and Thompson, 1971; Watkins *et al.*, 1987). Occasionally two or more overlapping series would be observed, but sounds from the two series could normally be distinguished by differences in loudness, in which case only the louder series would be measured. Sound measurements were made in a *session*, an uninterrupted period of time that included part or all of one series. The sound characteristics measured included duration, initial and final frequencies, and frequency range. When appropriate, *intercall interval* (the time from the end of one call to the beginning of the following), *intergroup interval* (time between groups of pulses), and *long interval* (time between call series substantially longer than the usual intercall or intergroup interval) were measured. Statistics are reported as mean and standard deviation. Measured values were then compared to published values of species known to occur in the North Atlantic.

Due to their extremely high source levels, sounds from airgun sources were regularly recorded on three or more hydrophones, and often all six hydrophones, simultaneously. Such multiple arrivals allowed the use of arrival time delays between instruments to estimate the location of the airgun source. A modified least-squares method and software developed by NOAA-PMEL for earthquake localization (Fox *et al.*, 2001) were used to estimate the locations of seismic survey vessels.

IV. RESULTS

From Feb. 1999 to Mar. 2001, over 17 000 h of data were examined for signals of interest. Analysis identified over 6000 earthquakes (Smith *et al.*, 2002), ship noise, seismic vessel survey (airgun) sounds, sounds of fin, blue, and minke whales, and sounds that were biological in origin but could not be definitively ascribed to a particular baleen whale. The most common biological sound recorded was a series of short, downswept pulses in the 15–30-Hz range that were identifiable as fin whale calls (Fig. 2). Calls with a high signal-to-noise ratio were measured from approximately 1 h of data from each of three hydrophones (NW, NE, and CW). On average, calls lasted 0.9±0.1 s (N whales=3, N calls =467). The intercall interval between the ''classic'' calls (Clark *et al.* 2002) averaged 17.5±0.4 s (N whales=3, N calls=380). However, a longer interval (38.1±1.9 s, N whales=3, N calls=80) that sometimes included a pulse of the type Clark *et al.* (2002) refer to as a ''backbeat'' was also observed. As this backbeat was not always visible, most likely due to the lower amplitude of this type of call, we did not make detailed measurements of the time between backbeat and classic pulses. The longer intervals between pulse series, referred to as ''rests'' by Watkins *et al.* (1987), averaged 135.3±21.9 s (N whales=3, N calls=9).

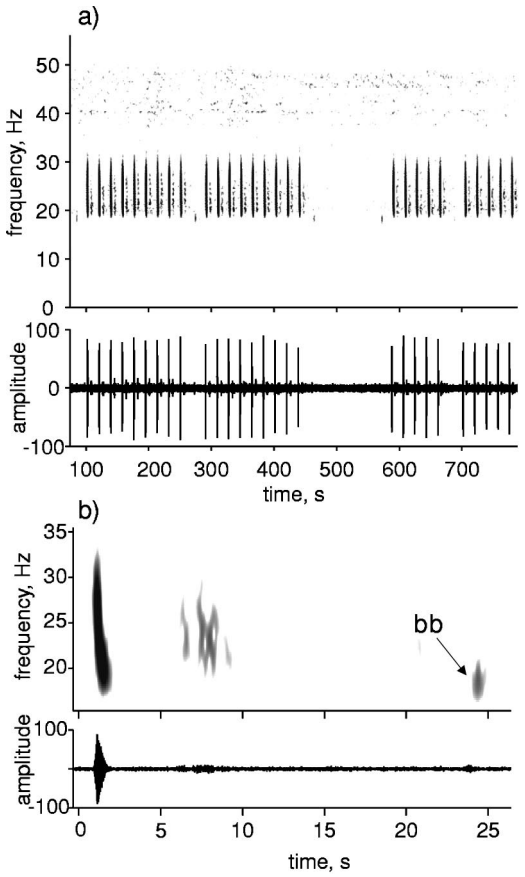


FIG. 2. (a) Spectrogram and time series of a series of fin whale calls recorded on the CW hydrophone, 15 January 2000 [spectrogram parameters: frame and FFT length 4.7 s (512 samples), overlap 0.75, Hamming window, for a filter bandwidth of 0.9 Hz]. (b) Detail of fin whale classic pulse and backbeat (bb; arrow) [spectrogram parameters: frame length 0.6 s (64 samples), FFT length 512 samples, overlap 0.9375, Hamming window, for a filter bandwidth of 7 Hz].

Fin whale calls were detected seasonally on all six hydrophones [Fig. 3(a)]. The majority of pulses was observed on the northern hydrophones from October to April, with peak detections in the winter months. In some months these calls were detected in over 85% of the hours examined. The same seasonal pattern was seen on the central hydrophones, although the proportion of hours with calls was much lower at these latitudes. At the southern hydrophones, relatively few hours with fin whale calls were observed. In the 2 years of the study, the seasonal occurrence of fin whale sounds was similar with the exception of the NE hydrophone, where fin whale calls occurred more often and later into the year in 2000–2001 than in 1999–2000.

The North Atlantic blue whale A–B phrase was the second most common whale sound recorded by this array [Fig. 4(a)]. These sounds occur in long, patterned series, interrupted by gaps that may represent breathing intervals (Cum-

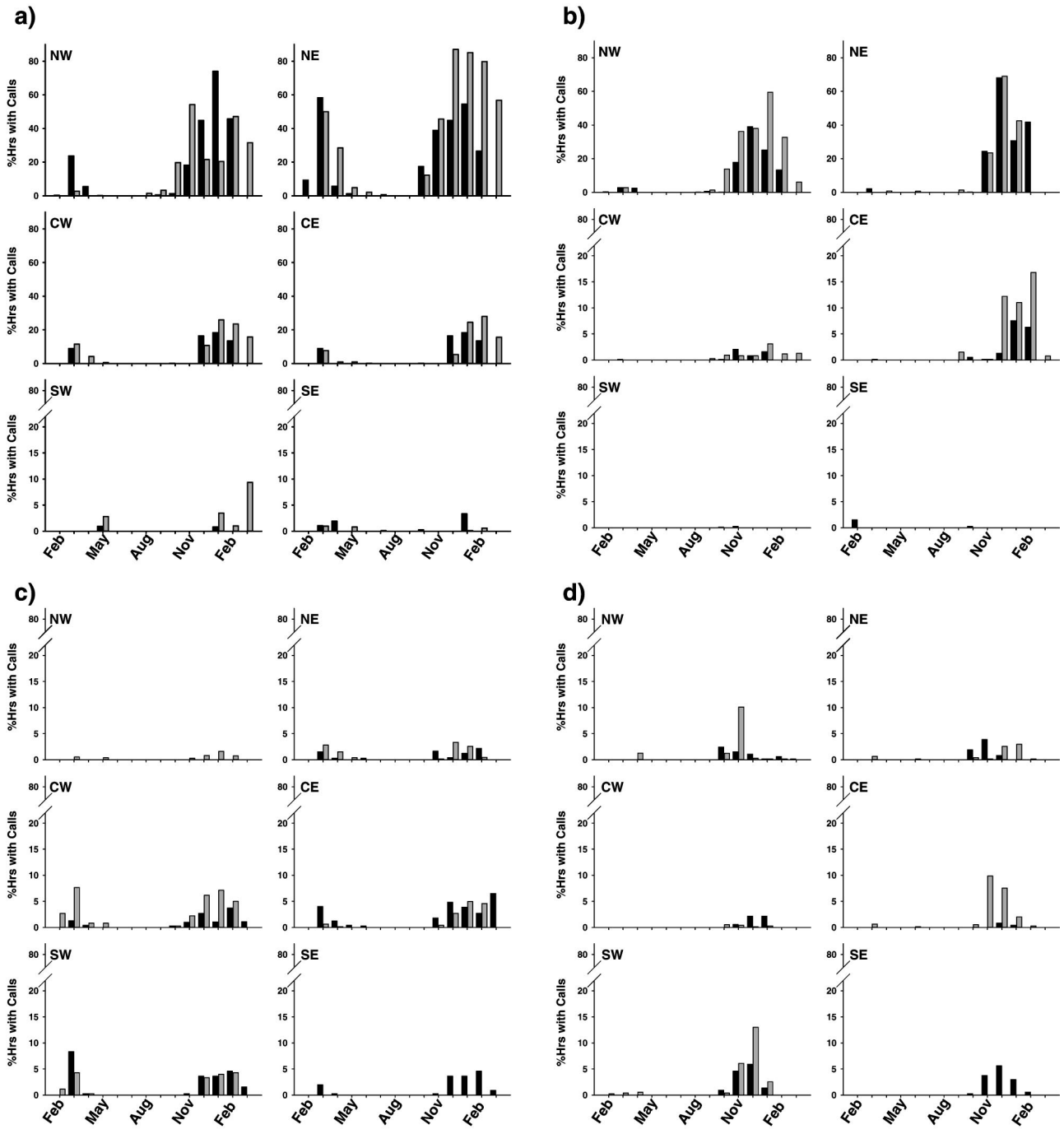


FIG. 3. Histograms of seasonal patterns of calls detected by the six autonomous hydrophones. Dark bars are data from Feb. 1999–Feb. 2000; light bars from Feb. 2000–Mar. 2001. Bars represent proportion of hours per month in which calls were detected. Seasonal patterns of sounds from (a) fin whales; (b) blue whales; (c) minke whales; (d) downswept vocalizations from unidentified baleen whales.

mings and Thompson, 1971). Here, we use the terminology of Mellinger and Clark (2003) to describe these sounds. Time–frequency characteristics were measured for a total of 794 blue whale phrases from ten different sessions when blue whale sounds were present, representing ten time periods and three locations in the North Atlantic. Of these phrases, 556 were A-only phrases and 227 were A–B phrases; 11 A–B phrase sequences also included arch sounds. In A-only phrases, part A swept from 18.4±0.2 Hz to 17.6±0.4 Hz and lasted 11.8±1.1 s ($N=10$). The time between consecutive A-only phrases was 60.0±3.8 s ($N=8$). In A–B phrases, part A lasted an average of 11.1±1.6 s ($N=8$), and part B swept from 18.3±0.2 Hz to 16.0+0.4 Hz over 9.0±0.9 s ($N=8$). Time between the two phrase parts was 4.8±0.8 s ($N=8$). The ratio of the numbers of parts A and B within a series varied from 1 to 155 over the time periods during which calls were measured ($\overline{\times}=26.3$ ±52, $N=8$). The number of phrases in each sequence was also quite variable, ranging from 2 to 11 ($\overline{\times}=6.5\pm2.6$, $N=10$). Arch sounds were noted in only one time period, and because the upper-frequency limit of the recordings was 50 Hz, only the lower-frequency portions of these sounds were recorded [Fig. 4(b)] Eleven arch sounds were measured.

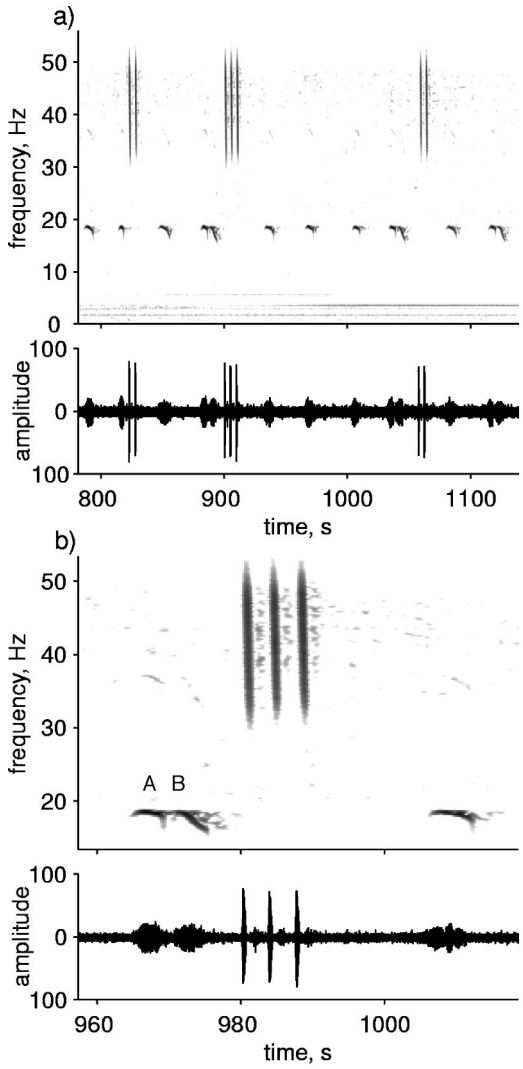


FIG. 4. (a) Spectrogram and time series of a series of blue whale AB calls and arch sounds recorded on the NE hydrophone, 15 November 1999 [spectrogram parameters same as in Fig. 2(a)]. (b) Detail of blue whale call [spectrogram parameters same as in Fig. 2(a) except overlap is 0.875].

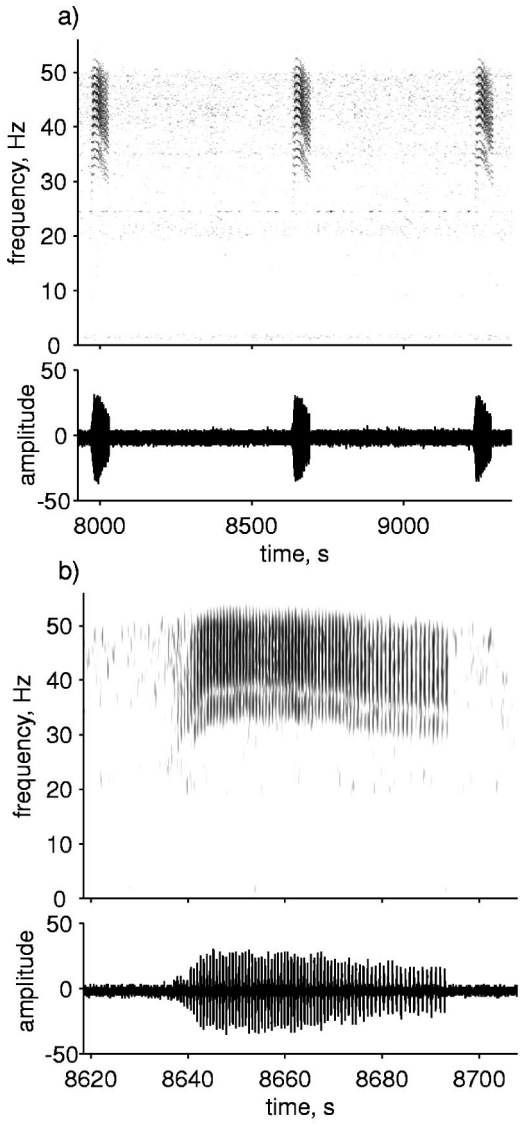


FIG. 5. Spectrograms and time series minke whale calls recorded on the SE hydrophone, 27 January 2000. (a) Series of minke whale calls. Because of the cutoff frequency of our hydrophones (50 Hz), these are considered partial calls, but are still similar enough to minke whale calls recorded elsewhere to be identifiable [spectrogram parameters: frame length 9.3 s (1024 samples), FFT length 2048 samples, overlap 0.5, Hamming window, for a filter bandwidth of 0.4 Hz]. (b). Detail of minke whale call. Note the decelerating pulse rate, a feature of nearly all of the calls examined [spectrogram parameters same as Fig. 2(b) except overlap is 0.75].

These occurred in groups of 2–3 during only two of the 20 sequences measured for this time period. Arch sounds swept from the top of our recording frequency range (ca. 50 Hz) down to 32.0 ± 1.0 Hz. On average, these sounds lasted 2.5 ± 0.4 s ($N=11$).

Atlantic blue whale calls were detected primarily on the northern hydrophones during the winter months [Fig. 3(b)]. Trends were similar between the 2 years. Call occurrence peaked in December and January on both the NW and NE hydrophones. As with the fin whale data, the seasonal pattern was similar on the CE and CW hydrophones but at greatly reduced rates. Call occurrence on the southern hydrophones was quite low in both years.

Series of pulsive calls (Fig. 5) were detected in both years [Fig. 3(c)] in the hydrophone recordings. These appeared to be the lower-frequency part of pulse trains that have been recorded from minke whales (Winn and Perkins, 1976; Swift *et al.*, 1996), having the same approximate pulse repetition rate and pulse train length (Mellinger *et al.*, 2000). Measurements of frequency and interpulse interval were made at the beginning, middle, and end of each pulse train because pulse rate and frequency band varied within the train. On average, the interpulse interval calculated from individual session averages ($N=26$) was 0.7 ± 0.1 s at the beginning of a pulse train, 0.7 ± 0.1 s at the middle of a pulse train, and 0.9 ± 0.04 s at the end. The ending interpulse interval was significantly longer (t-test, $p<0.05$), indicating that these pulse trains slowed down toward the end. The pulse trains, averaged over the sequences measured ($N=24$), contained 68 ± 6 pulses and lasted 48.0 ± 4.8 s, and successive trains occurred at intervals of 528 ± 94 s ($N=25$). The mean lower frequency was 30.6 ± 4.4 Hz ($N=24$), with a minimum frequency over all measured pulse trains of 21.0 Hz.

Pulse trains occurred most often during the months of December through March [Fig. 3(c)] and were most common

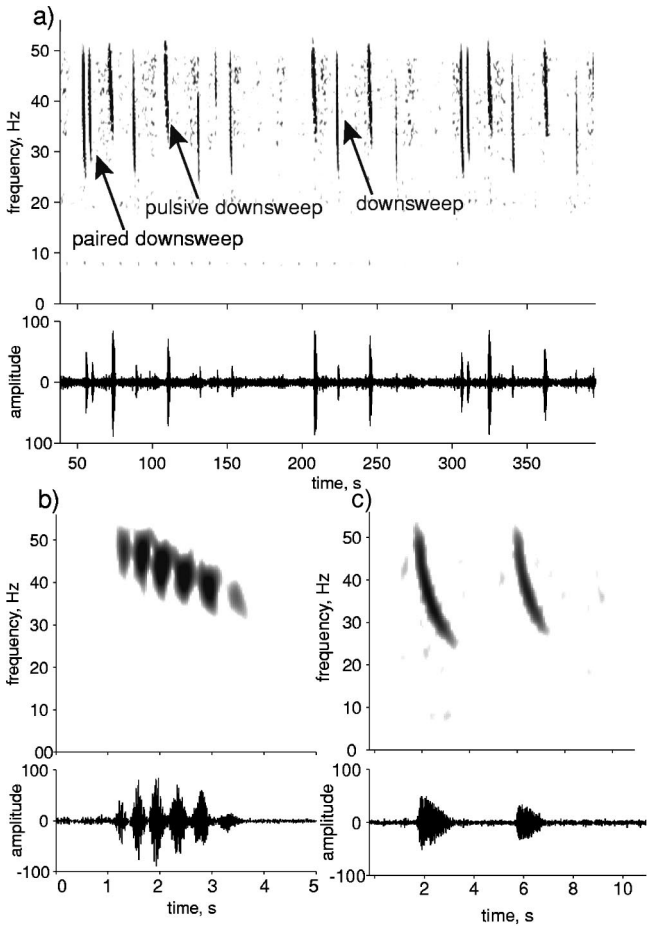


FIG. 6. (a) Spectrogram and time series of calls that are likely from a baleen whale recorded on the CW hydrophone, 26 November 1999 [spectrogram parameters: frame and FFT length 2.3 s (256 samples), overlap 0.5, Hamming window, for a filter bandwidth of 1.7 Hz]. This call typically consists of 1 or 2 downsweeps followed by a pulsive downsweep. (b) Detail of pulsive downsweep [spectrogram parameters: frame length 0.3 s (32 samples), FFT length 512 samples, overlap 0.875, Hamming window, for a filter bandwidth of 14 Hz]. (c) Detail of downsweep [spectrogram parameters: frame length 0.6 s (64 samples), FFT length 256 samples, overlap 0.875, Hamming window, for a filter bandwidth of 7 Hz].

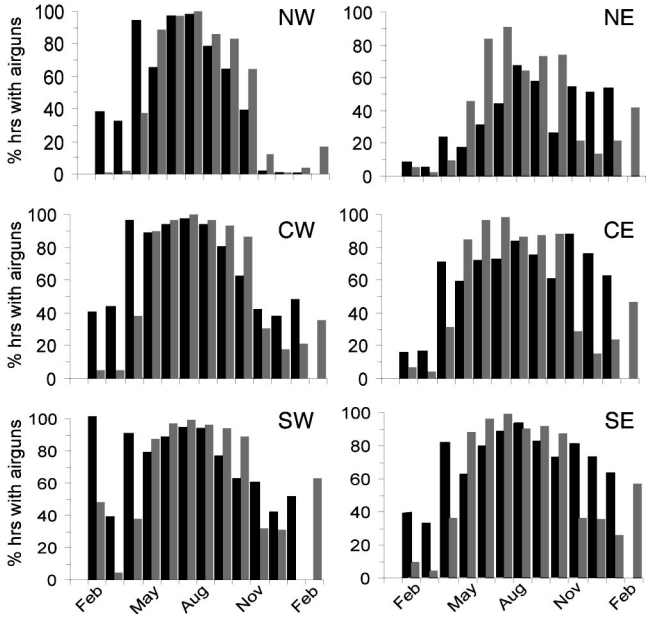


FIG. 7. Seasonal patterns of airgun pulses detected by the six autonomous hydrophones. Dark bars are data from Feb. 1999–Feb. 2000; light bars from Feb. 2000–Mar. 2001. Bars represent proportion of hours per month in which calls were detected.

on the central and southern hydrophones. This same seasonal pattern was seen on the NE hydrophone but at much lower rates of occurrence. Almost no pulse trains were detected on the NW hydrophone.

Less frequently, sounds were observed that swept from the maximum frequency of the autonomous hydrophone instrument, 50 Hz, down to approximately 30 Hz (Fig. 6). These sounds, here called *downsweep vocalizations*, sometimes occurred in pairs, and due to their general similarity to known baleen whale sounds, it is likely that these are baleen whale vocalizations. Cursory examination of the data revealed that there were typically two types of downsweep vocalizations: a simple, 3.0 ± 0.1-s downsweep from about 49.7 ± 0.9 Hz down to 29.0 ± 0.5 Hz ($N=6$), and a second, slightly longer (4.11 ± 0.1 s) sound more pulsive in nature. This second sound was recorded less often than the simple downsweep, was less likely to occur in pairs, and typically swept from about 50.2 ± 0.2 Hz down to 34.4 ± 0.4 Hz ($N=5$). There was no obvious pattern to the order of simple downsweeps and pulsive downsweeps. Simple downsweeps often occurred in series that did not include pulsive downsweeps. The exact nature of these sounds was difficult to determine, as it appeared there was additional content above the cutoff frequency of the recording system's filter. Both types of downsweeps were recorded primarily in winter and resemble a variety of sounds, including those recorded in the presence of blue whales in the St. Lawrence Estuary (C. Berchok, personal communication 2003), sounds described from the Norwegian Sea that are likely from fin whales (C. Clark, personal communication 2003), and the lower frequencies of some parts of humpback whale songs. In addition, these two sounds were often made in sequences exceeding 15 min that included no time gaps. If these were parts of a humpback whale song, the animal was apparently singing the same phrase or theme repeatedly. These sounds were recorded the least often of all those discussed in this manuscript, and were recorded most often on the central hydrophones in the late fall and winter months [Fig. 3(d)].

Sounds associated with seismic airguns were recorded routinely on all hydrophones, and trends were similar in the two years (Fig. 7). Typically airguns were heard every 10–20 s (Fig. 8). Although airgun sounds tended to dominate recordings during the summer months, loud whale vocalizations could still be detected during intense airgun activity (Fig. 8). Occasionally the array recorded airguns from more than one location, masking cetacean sounds and on four occasions making the spectrogram data impossible to use. The high received level of these impulses on multiple hydrophones made it possible to estimate the locations of the ships conducting the airgun surveys. During the summer months, airguns operated off Nova Scotia, Canada, probably in support of exploration in the Sable Island region (Fig. 1). From spring through fall seismic vessels, presumably commercial, were located working off the coast of western Africa and

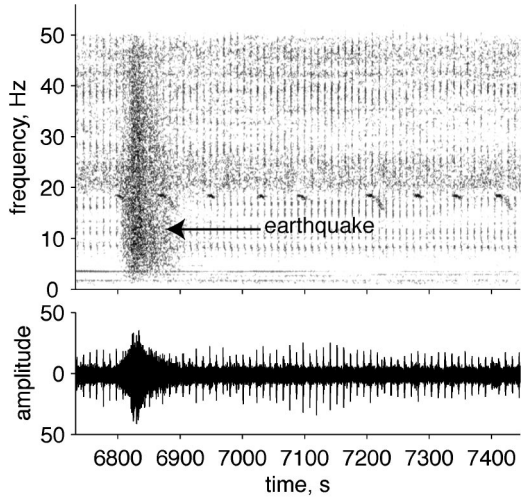


FIG. 8. (a) Spectrogram and time series of a series of airgun pulses, an earthquake (arrow), and a series of blue whale A–B calls recorded on the NE hydrophone, 11 February 2000 [spectrogram parameters same as in Fig. 2(a)].

northeast of Brazil. Seismic vessels operating in other areas of active exploration, such as the North Sea and the Gulf of Mexico, were not observed by this array due to bathymetric blockage.

V. DISCUSSION

Vocalizations of at least three, and possibly four, species of baleen whales were recorded on this array, despite its remote, offshore location. For all species and all hydrophones, most vocalizations were recorded in the late fall, winter, and early spring. This distinct seasonal pattern in detections may be due to changes in the vocal activity or migratory behavior of these whales. The wintertime peak in vocalizations apparent in our data coincides with what is thought to be the breeding season, and what is known to be the calving season, for baleen whales (Gaskin, 1982; Stewart and Leatherwood, 1985). It is male humpback whales that sing complex songs (Tyack, 1981; Darling and Bérubé, 2001), and there is limited evidence that it is male blue whales (McDonald *et al.*, 2001) and fin whales (Croll *et al.*, 2002) that produce loud, repetitive series of vocalizations. These long, repeating series of sounds made by male baleen whales during the winter months may be an advertisement signal (Tyack, 1981; Watkins *et al.*, 1987), similar to that produced by males of many terrestrial species during the breeding season (cf. Clutton-Brock and Albon, 1979; Searcy and Yasukawa, 1996; Gerhardt and Huber, 2002). Vocal advertisement could potentially occur year-round, as baleen whales are known to produce sounds in all months of the year (Clark and Gagnon, 2002), but probably peaks during the wintertime breeding season. Thus, increased vocal activity during the breeding season may be one explanation for the seasonal pattern of detections in these data.

In addition to this potential seasonal change in vocal behavior, the high numbers of detections during the winter months may also be due to an increase in the number of vocalizing individuals in the areas monitored by our hydrophones. Most baleen whales are distributed in productive, high-latitude regions during the summer, and migrate to less-productive, lower-latitude areas, sometimes far offshore, for calving and breeding in the winter months (Gaskin, 1982). Our array is in an area that could be a migratory destination or route for baleen whales. In other areas of the North Atlantic, the number of vocalizations detected has been strongly correlated ($r>0.85$) with the estimated number of individuals heard (Clark and Charif, 1998), and it therefore seems likely that the higher numbers of detections here are indicative of more vocalizing whales in the region. Because blue and fin whales vocalize in all months of the year (Clark and Charif, 1998; Charif and Clark, 2000; Stafford *et al.*, 2001; Clark and Gagnon, 2002) and very few of these vocalizations are detected on this mid-Atlantic array during the summer months, the number of vocalizing animals in the vicinity of this array during the summer months must be small. Airgun activity also peaks during the summer, but it is unlikely airguns are completely obscuring whale sounds, as calls are detected during some months of frequent airgun occurrence in the fall, and the repetition rate of airguns is such that most whale sounds can be detected between pulses (Fig. 8) (cf. Clark and Charif, 1998). Thus, the seasonal signal in our data may be due to changes in both the vocal activity and the distribution of the whales. These explanations are not mutually exclusive, nor does the explanation for one species necessarily apply to other species. These hypotheses cannot be tested with the data presented here but suggest directions for future research.

A. Fin and blue whales

The most common vocalizations recorded on our array were sounds from fin and blue whales. Fin whale detections far outnumbered those from blue whales in this study, despite the fact that we could probably detect the loud, very-low-frequency blue whale calls (SL=188 dB re: 1 μPa at 1 m; Cummings and Thompson, 1971) at greater distances from our array than the somewhat quieter, higher-frequency fin whale sounds (SL=183 dB re: 1 μPa at 1 m; Cummings and Thompson, 1994). The fin whale was also the species heard most often on other North Atlantic arrays (Clark, 1995; Clark and Charif, 1998). This is likely a reflection of the relative number of animals; the fin whale population estimate for the North Atlantic is approximately 50 000 animals, while that of the blue whale is 1000–2000 animals (Sigurjonsson, 1995).

Both fin and blue whale sounds were recorded primarily on the two northern hydrophones (~32–35 °N), less on the central hydrophones (~26 °N), and least on the southern hydrophones (~17 °N). This pattern agrees with what we know of their distribution from whaling, visual survey, and other acoustic survey data. Very little is known of fin and blue whale migratory movements, and it is currently unclear where calving, mating, and wintering occur (Jonsgard, 1966; Waring *et al.*, 2002). For both species, the summer feeding range is limited to the north by the ice edge, and may extend as far south as the British Isles in the eastern Atlantic and the Carolinas in the western Atlantic (Jonsgard, 1966; Rorvik and Jonsgard, 1981). In winter, this distribution likely shifts further south and offshore (Kellogg, 1929; Rorvik and Jonsgård, 1981; Sigurjonsson, 1995). Sighting and stranding data

from fall and winter months record these species as far south as the Gulf of Mexico, the Mediterranean Sea, and the coast of northwest Africa (~21 °N) (Kiripichnikov, 1950; Jonsgård, 1966; Sergeant, 1977; Rorvik and Jonsgård, 1981; Sigurjonsson, 1995). Given that our hydrophone array is far offshore and south of estimated fin and blue whale habitat, it is not surprising that the majority of detections was on the northernmost hydrophones. The near absence of these sounds on the southernmost hydrophones may show that blue and fin whales do not vocalize at this latitude or that they do not occur this far south.

B. Minke whales

Although the sampling rate of our hydrophones prevented us from detecting the full 25–800-Hz bandwidth of minke whale pulse trains (Winn and Perkins, 1976; Mellinger *et al.*, 2000), we did record at least some of the lower-frequency parts of these sounds. Current population estimates, based on data from visual surveys and whaling records, suggest that there are over 100 000 minke whales in the North Atlantic (Sigurjonsson, 1995). Despite their abundant numbers, minke whale sounds were recorded relatively rarely by our array; these sounds were typically detected in less than 10% of the hours sampled. Minke whale sounds are of lower amplitude (~165 dB re: 1 μPa at 1 m; Schevill and Watkins, 1972) than blue and fin whale vocalizations and therefore may be detected by our hydrophones only when whales are in close proximity to the array.

Unlike the fin and blue whale sounds, most of our minke detections were on the central and to a lesser extent the southern hydrophones. Few detections were made on the northern hydrophones, and the majority of these was on the NE hydrophone. Minke whale distribution is generally limited to the north by ice, and to the south distribution extends to the tropics, including Bermuda, Puerto Rico, the West Indies (Winn and Perkins, 1976; Mitchell, 1991; Mellinger *et al.*, 2000) and possibly to more offshore, deep-ocean waters in winter (Horwood, 1990; Mitchell, 1991). Very little is known of the distribution of minke whales in the southeastern North Atlantic. There are a few scattered sightings of minke whales at 20 °N 20 °W and at 11 °N 22 °W (Folkow and Blix, 1991) and in the Azores during May and June, and to a lesser extent July and August (Bento, 2002). It is unclear why we did not record more minke whale vocalizations on the northern hydrophones (~31 °N) given this species's distribution; perhaps vocalizing minke whales do not move within range of our array very often at this latitude, or perhaps they are not vocalizing while at this latitude. Many minke pulse trains may also have been missed because they were above the 50-Hz upper frequency limit of the filtering/recording system. More data will be needed to explain the paucity of recordings on these northern hydrophones.

As with the fin and blue whale data, there was a marked seasonal pattern in detections of minke whale vocalizations. Pulse trains were detected from November to April, peaking in December–February; very few sounds from minke whales were recorded during the summer months. This pattern agrees with findings of other acoustic studies in the North Atlantic. In the West Indies and Bermuda, sounds attributed to minke whales were recorded from October to April (Gagnon and Clark, 1993; Clark, 1994; Nishimura and Conlon, 1994; Clark, 1996) and in the eastern North Atlantic (28 °N 20 °W) in December (Folkow and Blix, 1991). Northeast of our array, minke whale sounds were detected from September to November in SOSUS data (Clark *et al.*, 2003). If minke whales breed in open ocean areas (Sigurjonsson, 1995), they could be moving into range of our array, and increasing their vocal activity during the winter months.

C. Downsweep vocalizations

In our data, detections of the simple and pulsive downsweep vocalizations that are potentially baleen whale sounds were quite limited. Most detections were during the winter months, but a few detections were as late as April. These sounds could be from humpback whales; we would expect to record humpback song or song fragments on our hydrophones during the winter, as this array is positioned within the migratory route of, and adjacent to, humpback wintering grounds (Charif *et al.*, 2001; Clapham and Mead, 1999; Stevick *et al.*, 1998; Palsboll *et al.*, 1997). Most sounds produced by singing humpbacks are above the 50-Hz cutoff frequency for our instruments, so identifying these sounds definitively as humpback whale vocalizations is difficult. Clark and Charif (1998) point out that humpback whale sounds are of lower intensity than other baleen whale sounds, and so detection range will be limited. These sounds could also be from blue or fin whales, as they are similar to sounds tentatively attributed these species (respectively, C. Berchok and C. Clark, personal communication 2003). In the future we plan to increase the sampling rate of the hydrophones to help determine the true nature of these sounds.

D. Airguns

Since this hydrophone array was deployed, the periodic impulses produced by seismic exploration vessels operating around the Atlantic basin were the dominant signal detected. Concern over the potential effects of anthropogenic noise on marine life has been such that the National Research Council of the (U.S.) National Academy of Science has commissioned three studies on this topic to date (NRC 1994, 2000, 2003). Although seismic airgun arrays are designed to direct the majority of emitted energy downward through the seafloor, their sound emission horizontally is also significant (NRC, 2003). Airgun survey vessels were often located 3000 km or more from our array (Fig. 1), yet airgun pulses were still clearly recorded on each hydrophone. The broadband frequency range and repeated firing of these guns make them a major contributor to the low-frequency sound field in the North Atlantic.

Airgun activity in shallow water has been shown to significantly damage the ears of fish (McCauley *et al.*, 2000) and has been implicated in the stranding of beaked whales (Malakoff 2002; NRC 2003). Its effect on the baleen whales studied here is unknown; possible effects include masking of conspecific sounds, increased stress levels, changing vocalizations, and ear damage (Richardson *et al.*, 1995). Most of

the seismic vessels we located were operating in marine mammal habitat, including that of the critically endangered northern right whale.

Airgun pulses were recorded year-round but were most common from late spring through fall. This pattern is the opposite of the peak occurrences for all baleen whale calls. It is possible that the seasonal patterns seen in baleen whale calls are due to airgun interference: that is, the calls are produced in the summer months but obscured by airguns. However, because calls are detected during some months of frequent airgun occurrence in the fall, because the repetition rate of airguns is such that most whale sounds can be detected between pulses (Fig. 8), and because the data were visually inspected, we don't believe that many calls were missed due to interference (cf. Clark and Charif, 1998).

VI. CONCLUSIONS

Sounds from fin, blue, minke, and possibly another species of baleen whale were recorded on six autonomous hydrophones moored near the Mid-Atlantic Ridge. The low sampling rate of our recordings and the unknown acoustic repertoire of other species made the identity of some vocalizations uncertain. Recording data at a higher sampling rate could help determine the origin of this call type and provide data on the full frequency range of minke whale calls and the recently documented short, narrow-band sounds produced by Bryde's whales (*B. edeni*) (Oleson *et al.*, 2003). In addition, given the recent interest in the contribution of airgun noise to the marine environment and its potential adverse effects on marine mammals (cf. NRC 2000, 2003), we recommend continued monitoring of this area airgun sounds and a more formal measurement of this source of noise.

The utility of remote acoustic monitoring has been well established for determining the occurrence of calling whales in regions and during times that are not feasible for traditional visual survey methods (Clark and Charif, 1998; Stafford *et al.*, 1999, 2001; Watkins *et al.*, 2000; Mellinger *et al.*, 2003, 2004). The constraints of interpreting such acoustic data include the difficulty of associating the number of sounds recorded with the number of animals present, the detection range of the sounds, and seasonal, behavioral, and demographic variation in calling behavior data (Clark and Charif, 1998; Mellinger and Barlow, 2003). If remotely acquired acoustic data are to provide the information necessary for the management and recovery of large whales, future efforts must address these obstacles. Despite the limitations, the acoustic data from this array have given us insight into the vocal behavior and occurrence of highly mobile, migratory animals in a very remote part of the Atlantic Ocean that has traditionally been difficult to survey.

ACKNOWLEDGMENTS

Funding for the deployment, maintenance, and data analysis for the hydrophone array along the Mid-Atlantic Ridge was provided by the National Science Foundation via grants awarded to R. P. Dziak (Grants OCE-9811575, OCE-0137164, and OCE-0201692). This work was also supported by ONR Contracts N00014-00-F-0395 and N00014-03-1-0099 awarded to D. K. Mellinger. Development of the hydrophones was provided by in-kind funding from NOAA-PMEL. We are indebted to Haru Matsumoto for design and construction of the hydrophones used in this experiment; Andy Lau for the development of the PMEL software used for analysis of the data; Matt Fowler, Joe Haxel, Sarah Follett, and Rita Bento for their help with various stages of the analysis; and Sara Heimlich and two anonymous reviewers for manuscript revisions. This is PMEL contribution #2603.

Bento, R. (**2002**). ''Monitorização acústica da sazonalidade de misticetos na Crista Média Atlântica'' (Acoustic monitoring of mysticete sounds along the Mid-Atlantic Ridge). Masters thesis, Universidade dos Açores, Ponta Delgada.

Cerchio, S., and Dahlheim, M. (**2001**). ''Variation in feeding vocalizations of humpback whales *Megaptera novaeangliae* from southeast Alaska,'' Bioacoustics **11**, 277–295.

Charif, R. A., and Clark, C. W. (**2000**). ''Acoustic monitoring of large whales off north and west Britain and Ireland: A two-year study, October 1996–September 1998,'' Joint Nature Conservation Committee Technical Report **313**.

Charif, R. A., Clapham, P. J., and Clark, C. W. (**2001**). ''Acoustic detections of singing humpback whales in deep waters off the British Isles,'' Marine Mammal Sci. **17**, 751–768.

Clapham, P. J., and Mattila, D. K. (**1990**). ''Humpback whale songs as indicators of migration routes,'' Marine Mammal Sci. **6**, 155–160.

Clapham, P., and Mead, J. G. (**1999**). ''*Megaptera novaeangliae*,'' Mammalian Species **604**, 1–9.

Clark, C. W. (**1994**). ''Application of U.S. Navy underwater hydrophone arrays for scientific research on whales,'' Rep. Int. Whal. Commn. **44**, 1–12.

Clark, C. W. (**1995**). ''Matters arising out of the discussion of blue whales,'' Rep. Int. Whal. Commn. **45**, 210–212.

Clark, C. W. (**1996**). ''Whales '93: Hot tracks and greatest hits; blue whales '93; fin whales '93; minke whales '93, North Atlantic. A seven-CD set,'' Cornell Laboratory of Ornithology, Bioacoustics Research Program, 159 Sapsucker Woods Road, Ithaca, NY 14850.

Clark, C. W., and Charif, R. A. (**1998**). ''Acoustic monitoring of large whales to the west of Britain and Ireland using bottom-mounted hydrophone arrays, October 1996–September 1997,'' Report No. 281, Joint Nature Conservation Committee, Aberdeen, Scotland.

Clark, C. W., and Fristrup, K. M. (**1997**). ''Whales '95: A combined visual and acoustic survey of blue and fin whales off southern California,'' Rep. Int. Whal. Commn. **47**, 583–600.

Clark, C. W., Charif, R. A., Mitchell, S., and Colby, J. (**1996**). ''Distribution and behavior of the bowhead whale, *Balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring migration off Pt. Barrow, Alaska,'' Rep. Int. Whal. Commn. **46**, 541–554.

Clark, C. W., Borsani, J. F., and Notarbartolo-Di-Sciara, G. (**2002**). ''Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea,'' Marine Mammal Sci. **18**, 286–295.

Clark, C. W., Charif, R. A., Loveday, P., and Tasker, M. (**2003**). ''Seasonal occurrences of blue, fin, humpback, and minke whales in the eastern North Atlantic based on passive acoustic monitoring,'' Seventeenth Annual Conf., European Cetacean Society.

Clark, C. W., and Gagnon, G. C. (**2002**). ''Low-frequency vocal behaviors of baleen whales in the North Atlantic: Insights from IUSS detections, locations, and tracking from 1992 to 1996,'' J. Underwater Acoust. 52.

Clutton-Brock, T. H., and Albon, S. D. (**1979**). ''The roaring of red deer and the evolution of honest advertisement,'' Behaviour **69**, 145–169.

Croll, D. A., Clark, C., Acevedo, A., Tershey, B., Flores, S., Gedamke, J., and Urban, J. (**2002**). ''Only male fin whales sing loud songs,'' Nature (London) **417**, 809.

Cummings, W. C., Thompson, P. O., and Ha, S. J. (**1986**). ''Sounds from Bryde, *Balaenoptera edeni*, and finback, *B. physalus*, whales in the Gulf of California,'' Fish. Bull. **84**, 359–370.

Cummings, W. C., and Thompson, P. O. (**1971**). ''Underwater sounds from the blue whale, *Balaenoptera musculus*,'' J. Acoust. Soc. Am. **50**, 1193–1198.

Cummings, W. C., and Thompson, P. O. (**1994**). ''Characteristics ans seasons of blue and finback whale sounds along the U.S. west coast as recorded at SOSUS,'' J. Acoust. Soc. Am. **95**, 2853.

Darling, J. D., and Bérubé, M. (**2001**). ''Interactions of singing humpback whales with other males,'' Marine Mammal Sci. **17**, 570–584.
Dragoset, W. (**2000**). ''Introduction to airguns and airgun arrays,'' Leading Edge **19**, 892–897.
Edds, P. L. (**1982**). ''Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River,'' J. Mammal. **63**, 345–347.
Edds, P. L. (**1988**). ''Characteristics of finback, *Balaenoptera physalus*, vocalizations in the St. Lawrence Estuary,'' Bioacoustics **1**, 131–149.
Edds-Walton, P. L. (**2000**). ''Vocalizations of minke whales (*Balaenoptera acutorostrata*) in the St. Lawrence Estuary,'' Bioacoustics **11**, 31–50.
Folkow, L. P., and Blix, A. S. (**1991**). ''Norwegian whale sighting and acoustic surveys in the Atlantic Ocean during the winter of 1989/90,'' Rep. Intl. Whal. Comm. **41**, 531–538.
Fox, C. G., Matsumoto, H., and Lau, T.-K. A. (**2001**). ''Monitoring Pacific Ocean seismicity from an autonomous hydrophone array,'' J. Geophys. Res. **106**, 4183–4206.
Gagnon, G. J., and Clark, C. W. (**1993**). ''The use of U.S. Navy IUSS passive sonar to monitor the movements of blue whales,'' Proc. 10th Biennial Conf. Biol. Mar. Mamm., p. 50.
Gaskin, D. E. (**1982**). *The Ecology of Whales and Dolphins* (Heinemann, London).
Gerhardt, H. C., and Huber, F. (**2002**). *Acoustic Communication in Insects and Anurans* (University of Chicago Press, Chicago).
Guinee, L. N., Chu, K., and Dorsey, E. M. (**1983**). ''Changes over time in the song of known individual humpback whales (*Megaptera novaeangliae*),'' in *Communication and Behavior of Whales*, edited by R. Payne (Westview, Boulder), pp. 59–80.
Helweg, D. A., Cato, D. H., Jenkins, P. F., Garrigue, C., and McCauley, R. D. (**1998**). ''Geographic variation in South Pacific humpback whale songs,'' Behaviour **135**, 1–27.
Helweg, D. A., Herman, L. M., Yamamoto, S., and Forestell, P. H. (**1990**). ''Comparison of songs of humpback whales (*Megaptera novaeangliae*) recorded in Japan, Hawaii and Mexico during the winter of 1989,'' Sci. Rep. Cetacean Res. **1**, 1–20.
Horwood, J. (**1990**). *Biology and Exploitation of the Minke Whale* (CRC Press, Boca Raton).
Jonsgård, A. (**1966**). ''The distribution of Balaenopteridae in the North Atlantic Ocean,'' in *Whales, Dolphins and Porpoises*, edited by K. S. Norris (University of California Press, Berkeley), pp. 114–123.
Kellogg, R. (**1929**). ''What is known of the migrations of some of the whalebone whales,'' Smithsonian Inst. Ann. Rep. **1928**, 467–494.
Kiripichnikov, A. A. (**1950**). ''Observations on the distributions of large cetaceans,'' Priroda **10**, 63–64.
Malakoff, D. (**2002**). ''Suit ties whale deaths to research cruise,'' Science **298**, 722–723.
Mattila, D. K., Guinee, L. N., and Mayo, C. A. (**1987**). ''Humpback whale songs on a North Atlantic feeding ground,'' J. Mammal. **68**, 880–883.
McCauley, R. D., Jenner, C., Bannister, J. L., Cato, D. H., and Duncan, A. J. (**2000**). ''Blue whale calling in the Rottnest trench, Western Australia, and low frequency sea noise,'' Austral. Acoust. Soc. Conf., Joondalup, Australia, 15–17 November, 2000.
McDonald, M., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). ''The acoustic calls of blue whales off California with gender data,'' J. Acoust. Soc. Am. **109**, 1728–1735.
McSweeney, D. J., Chu, K. C., Dolphin, W. F., and Guinee, L. N. (**1989**). ''North Pacific humpback whale songs: A comparison of southeast Alaskan feeding ground songs with Hawaiian wintering ground songs,'' Marine Mammal Sci. **5**, 139–148.
Mellinger, D. K., and Barlow, J. (**2003**). ''Future directions for acoustic marine mammal surveys: Stock assessment and habitat use,'' Report of a workshop held in La Jolla, CA, 20–22 November, 2002, Contribution No. 2557, NOAA/PMEL, Seattle.
Mellinger, D. K., and Clark, C. W. (**2003**). ''Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic,'' J. Acoust. Soc. Am. **114**, 1108–1119.
Mellinger, D. K., Carson, C. D., and Clark, C. W. (**2000**). ''Characteristics of minke whale (*Balaenoptera acutorostrata*) pulse trains recorded near Puerto Rico,'' Marine Mammal Sci. **16**, 739–756.
Mellinger, D. K., Stafford, K. M., and Fox, C. G. (**2004**). ''Seasonal occurrence of sperm whale (*Physeter macrocephalus*) sounds in the Gulf of Alaska, 1999–2001,'' Marine Mammal Sci. **20**, 48–62.
Mitchell, E. D. (**1991**). ''Winter records of the minke whale (*Balaenoptera acutorostrata* Lacépède 1804) in the southern North Atlantic,'' Rep. Int. Whal. Commn. **41**, 455–457.
Moore, S. E., Stafford, K. M., Dahlheim, M. E., Fox, C. G., Braham, H. W., Polovina, J. J., and Bain, D. E. (**1998**). ''Seasonal variation in reception of fin whale calls at five geographic areas in the North Pacific,'' Marine Mammal Sci. **14**, 617–627.
National Research Council (NRC). (**1994**). *Low-Frequency Sound and Marine Mammals: Current Knowledge and Research Needs* (National Academies, Washington, D.C.).
National Research Council (NRC). (**2000**). *Marine Mammals and Low-Frequency Sound* (National Academies, Washington, D.C.).
National Research Council (NRC). (**2003**). *Ocean Noise and Marine Mammals* (National Academies, Washington, D.C.).
Nishimura, C. E., and Conlon, D. M. (**1994**). ''IUSS dual use: Monitoring whales and earthquakes using SOSUS,'' Mar. Technol. Soc. J. **27**, 13–21.
Norris, T. F., McDonald, M. A., and Barlow, J. (**1999**). ''Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration,'' J. Acoust. Soc. Am. **106**, 506–514.
Oleson, E. M., Barlow, J., Gordon, J., Rankin, S., and Hildebrand, J. A. (**2003**). ''Low-frequency calls of Bryde's whales,'' Marine Mammal Sci. **19**, 406–419.
Palsboll, P. J., Allen, J., Bérubé, M., Clapham, P. J., Feddersen, T. P., Hammond, P. S., Hudson, R. R., Jorgensen, H., Katona, S., Larsen, A. H., Larsen, F., Lien, J., Mattila, D. K., Sigurjonsson, J., Sears, R., Smith, T., Sponer, R., Stevick, P., and Øien, N. (**1997**). ''Genetic tagging of humpback whales,'' Nature (London) **388**, 767–769.
Parkes, G., and Hatton, L. (**1986**). *The Marine Seismic Source* (Reidel, Dordrecht).
Payne, R., and McVay, S. (**1971**). ''Songs of humpback whales,'' Science **173**, 585–597.
Payne, K., and Payne, R. (**1985**). ''Large scale changes over 19 years in the songs of humpback whales in Bermuda,'' Z. Tierpsychol. **68**, 89–114.
Payne, R. S., and Guinee, L. N. (**1983**). ''Humpback whale (*Megaptera novaeangliae*) songs as an indicator of stocks,'' in *Communication and Behavior of Whales*, edited by R. S. Payne (Westview, Boulder), pp. 333–358.
Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).
Rorvik, C. J., and Jonsgard, A. (**1981**). ''Review of Balaenopterids in the North Atlantic Ocean,'' in Mammals in the Seas, Vol III: General Papers and Large Cetaceans, pp. 269–286, FAO Fisheries Series No. 5.
Schevill, W. E., and Watkins, W. A. (**1972**). ''Intense low-frequency sounds from an Antarctic minke whale, *Balaenoptera acutorostrata*,'' Breviora **388**, 1–8.
Stewart, B. S., and Leatherwood, S. (**1985**). ''Minke whale: *Balaenoptera acutorostrata* (Lacépède, 1804),'' in *Handbook of Marine Mammals, Volume 3: The Sirenians and Baleen Whales*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 91–136.
Searcy, W. A., and Yasukawa, K. (**1996**). ''Song and female choice,'' in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca), pp. 454–473.
Sergeant, D. E. (**1977**). ''Stock of fin whales *Balaenoptera physalus* L. in the North Atlantic Ocean,'' Rep. Int. Whal. Commn. **27**, 460–473.
Sigurjonsson, J. (**1995**). ''On the life history and autecology of North Atlantic rorquals,'' Whales, Seals, Fish, and Man: Proc. Int. Sympos. Biol. Mar. Mamm. in the North East Atlantic, Tromsø, Norway, 29 November–1 December, 1994, pp. 425–441.
Silber, G. K. (**1986**). ''The relationship of social vocalizations to surface behavior and aggression in the Hawaiian humpback whale (*Megaptera novaeangliae*),'' Can. J. Zool. **64**, 2075–2080.
Smith, D. K., Tolstoy, M., Fox, C. G., Bohnenstiehl, D. R., Matsumoto, H., and Fowler, M. J. (**2002**). ''Hydroacoustic monitoring of seismicity at the slow-spreading Mid-Atlantic Ridge,'' Geophys. Res. Lett. **29**, 1–4.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999**). ''Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific,'' J. Acoust. Soc. Am. **106**, 3687–3698.
Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**2001**). ''Geographic and seasonal variation of blue whale calls in the North Pacific,'' J. Cetacean Res. Manage. **3**, 65–76.
Stevick, P. T., Øien, N., and Matilla, D. K. (**1998**). ''Migration of a humpback whale (*Megaptera novaeangliae*) between Norway and the West Indies,'' Marine Mammal Sci. **14**, 162–166.
Swift, R., Menhenett, C., and Gordon, J. C. D. (**1996**). ''Feasibility study: Using sonobuoys to study cetacean acoustic behavior in the coastal waters

of the Inner Hebrides, Scotland,'' European Research on Cetaceans, p. 76.
Thomson, D. H., and Richardson, W. J. (**1995**). ''Marine mammal sounds,'' in *Marine Mammals and Noise*, edited by W. J. Richardson, C. R. Greene, Jr., C. I. Malme, and D. H. Thomson (Academic, San Diego), pp. 159–204.
Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). ''20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico,'' J. Acoust. Soc. Am. **92**, 3051–3057.
Thompson, P. O., Cummings, W. C., and Ha, S. J. (**1986**). ''Sounds, source levels, and associated behavior of humpback whales, Southeast Alaska,'' J. Acoust. Soc. Am. **80**, 735–740.
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. (**1981**). ''Interactions between singing Hawaiian humpback whales and conspecifics nearby,'' Behav. Ecol. Sociobiol. **8**, 105–116.
Tyack, P. (**1983**). ''Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds,'' Behav. Ecol. Sociobiol. **13**, 49–55.
Tyack, P., and Whitehead, H. (**1983**). ''Male competition in large groups of wintering humpback whales,'' Behaviour **83**, 132–154.
Waring, G. T., Quintal, J. M., and Fairfield, C. P., editors (**2002**). U.S. Atlantic and Gulf of Mexico marine mammal stock assessments—2002. NOAA Tech. Memo. NMFS-NE-169.
Watkins, W. A. (**1981**). ''Activities and underwater sounds of fin whales,'' Sci. Rep., Whales Res. Inst. **33**, 83–117.
Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). ''The 20-Hz signals of finback whales (*Balaenoptera physalus*),'' J. Acoust. Soc. Am. **82**, 1901–1912.
Watkins, W. A., Daher, M. A., Reppucci, G. M., George, J. E., Martin, D. L., DiMarzio, N. A., and Gannon, D. P. (**2000**). ''Seasonality and distribution of whale calls in the North Pacific,'' Oceanography **13**, 62–67.
Winn, H. E., and Perkins, P. J. (**1976**). ''Distribution and sounds of the minke whale, with a review of mysticete sounds,'' Cetology **19**, 1–12.
Winn, H. E., Thompson, T. J., Cummings, W. C., Hain, J., Hudnall, J., Hays, H., and Steiner, W. W. (**1981**). ''Song of the humpback whale—population comparisons,'' Behav. Ecol. Sociobiol. **8**, 41–46.

Erratum: "Forward scattering of pulses from a rough sea surface by Fourier synthesis of parabolic equation solutions" [J. Acoust. Soc. Am. 114, 1266–1280 (2003)]

David A. Miles, Robin N. Hewitt, Marcus K. Donnelly,[a] and Timothy Clarke
QinetiQ, Winfrith Technology Centre, Dorset DT2 8XJ, U.K.

(Received 12 January 2004; accepted for publication 18 January 2004)

[DOI: 10.1121/1.1675815]

PACS numbers: 43.20.Fn, 43.30.Hw, 43.10.Vx [SLB]

In the introduction, the reference to the work where conformal mapping was implemented within the computer model FEPE is incorrect. The reference given [Ref. 7, M. D. Collins, ''Application and time-domain solution of higher order parabolic equations in underwater acoustics,'' J. Acoust. Soc. Am. **86**, 1097–1102 (1989)] is only concerned with the development of the FEPE computer model. The appropriate reference for the development of the FEPE-CM (FEPE/EFEPE coupled to the conformal mapping algorithm) is missing. The implementation of a conformal mapping algorithm to include the effects of a deterministic rough surface (based on the work by Dozier [Ref. 6, J. Acoust. Soc. Am. **75**, 1415–1432 (1984)]) within FEPE/EFEPE was carried out and validated in the following paper:

Norton, G. V., Novarini, J. C., and Keiffer, R. S., ''Coupling scattering from the sea surface to a one-way marching propagation model via conformal mapping: Validation,'' J. Acoust. Soc. Am. **97**, 2173–2180 (1995).

[a]Now at: Systems Engineering Assessment (SEA), Beckington Castle, PO Box 800, Frome, BA11 6TB, England.

 0001-4966/2004/115(4)/1844/1/$20.00