CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 117, No. 1 January 2005

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 1, JANUARY 2005

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 1, JANUARY 2005

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 1, JANUARY 2005

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

ERRATA

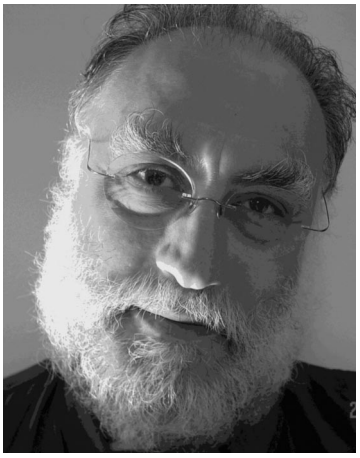
ACOUSTICAL NEWS—USA

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

Patricia Keating—For contributions to the integration of the phonetic and linguistics aspects of speech production.



Manell Zakharia—For contributions to underwater acoustic imaging.

ASA Awards Presented at 2004 International Science and Engineering Fair (ISEF)

Special awards were presented by representatives of ASA to five high school students at the 2004 ISEF held in Portland, Oregon in May of this year under the corporate sponsorship of Intel Corporation. The fair has been held since 1950 to stimulate interest in scientific and engineering careers among high school students around the world and the interest in the fair has been steadily growing over the years. This year, the fair included 1429 entrants from the United States and 40 other countries who exhibited projects that represented one or more categories of 14 technical areas of science and engineering.

Kerrie Standlee, Chair of ASA Judges, presents ASA awards to Intel ISEF Science Fair participants (1 to r): Jonathan Bell, Joline Fan, Alison Slowey, Benjamin Schmidt and Alexander Dillard.

Of the 1429 entrants at the fair, the ASA judging team, after a preliminary review of all project titles, concluded that 53 of the projects potentially fell within the technical areas of the Acoustical Society of America (ASA) and warranted further consideration for the special awards presented by ASA. The three judges forming the ASA judging team split the 53 projects into three groups and each judge took a group of projects to visit and decide if the project was in fact enough of an acoustical by related project to retain it in the group of projects warranting further consideration. With this approach, the field of eligible projects was reduced to 32. Each of the 32 projects was then examined in more detail by at least two of the three judges and then notes compared to develop a final list of eight finalists that would be interviewed and ranked for consideration for the awards bestowed by the ASA.

Normally, ASA awards a first place prize of $500 and three non-cash Honorable Mention awards with all award recipients receiving a free one-year subscription to JASA on CD ROM. However, because there were two students who had such outstanding projects in different areas of acoustics it was decided that two first place awards of $250 would be awarded instead of the one first place award of $500. With this decision, a first place award of $250 was given to Joline Marie Fan from Upper Arlington High School in Upper Arlington, Ohio for her Engineering category project entitled, *Acoustic Excitation for Enhancing Interfacial Dynamics of Mesobubbles*. During her interview with the judging team, the 16-year-old Ms. Fan demonstrated that she had an unquestionable understanding of the way in which acoustic waves interact with mesobubbles in multi-bubble systems to regulate the dynamic behavior of the bubbles. One of the judges with a Ph.D. from MIT told Ms. Fan that she had learned as much about her topic of interest at age 16 as he did when he was 30 years old. All judges agreed and told Ms. Fan that, based on her work presented at the ISEF, she should find it very easy to

 0001-4966/2005/117(1)/1/6/$22.50

receive a scholarship in the field of acoustics at almost any university with an acoustics program and we encouraged her to continue her interest in acoustics.

A second, first place award of $250 was given to 15-year-old Benjamin Alexander Schmidt from Center Wellington District High School of Fergus, Ontario, Canada for his Engineering category project entitled, *SMART: Static Microphone Arrays for Robot Tracking*. Mr. Schmidt's project caught the judges' eyes because he used simple materials in developing a microphone array for his project, he developed a mathematical model of his system, and he used a computer to control his array and test his mathematical model. Mr. Schmidt's presentation during his interview by the judges demonstrated that he, like Ms. Fan, had a very strong understanding of the acoustic principals associated with the project. Mr. Schmidt was encouraged to continue his interest in acoustics and consider looking for a university in which he could further his interest.

Honorable Mention awards were given to 16-year-old Alison Karee Slowey of Saint Michael's Academy in Bryan, Texas for her Earth and Space Sciences project entitled, *Sound Wave Propagation in Brine Pools*; to 15-year-old Jonathan Brooks Bell of R. Nelson Snider High School in Fort Wayne, Indiana for his Engineering project entitled, *Dampening the Modes of Vibration of Multistory Scale Models Due to Seismic Waves*; and to 17-year-old Alexander William Dillard of Bradshaw High School in Florence, Alabama for his Physics project entitled, *The Effects of Acoustic Pressure Disturbances on Refraction of Visible Light in Gases*. Ms. Slowey's project was a study involving the use of acoustic energy to determine the salinity of water in the Gulf of Mexico. Mr. Bell's project was an investigation into the use of passive dampers to minimize the resonant modes of vibration in scale model multi-story buildings. Mr. Dillard's project was a study into the use of acoustic pressure waves to influence the refraction properties of light in gases. All three projects were considered excellent examples of the quality of acoustic related work brought to the fair this year and gave the judges the feeling that high school students are beginning to see acoustics as a field worth pursuing.

The ASA judging team included Dr. Alfred L. Nuttal from the Oregon Health Sciences University of Portland, Oregon, Dr. Charles Oppenheimer with Daly-Standlee & Associates, Inc., a consulting acoustical engineering firm in Beaverton, Oregon, and Mr. Kerrie G. Standlee, P.E., the Principal of Daly-Standlee & Associates, Inc. Mr. Standlee acted as the Chair of the judging team and presented the ASA awards at the award ceremony. The judging team found participating in the science and engineering fair to be a wonderful and memorable experience and they found it heartening to see the enthusiasm and interest in the field of acoustics shown by those students with projects involving acoustics. The judges would like to encourage those who have an opportunity in the future to participate in an ISEF program should it come to your community.

KERRIE G. STANDLEE, P.E.

Regional Chapter News

Greater Boston Chapter

On September 23, 2004, the Greater Boston Chapter of the Acoustical Society hosted a tour of the 500 Atlantic Avenue construction project. This is a hotel/condominium building situated atop the subsurface Vent Building 3 (VB 3) of the Boston Central Artery/Third Harbor Tunnel. The Central Artery Tunnel construction is complete and VB 3 is in full operation. 500 Atlantic Avenue is one of the first projects to take advantage of the new cityscape being opened for park and building development along the route of the former elevated Central Artery.

In the photo are ASA/GBC members who attended the tour. Gregory C. Tocci (to the far left) conducted the 500 Atlantic Avenue tour of construction above ground. Erick Thalheimer, Senior Noise Control Engineer for the ''Big-Dig'' (second from the left), lead the tour of VB 3 below ground.

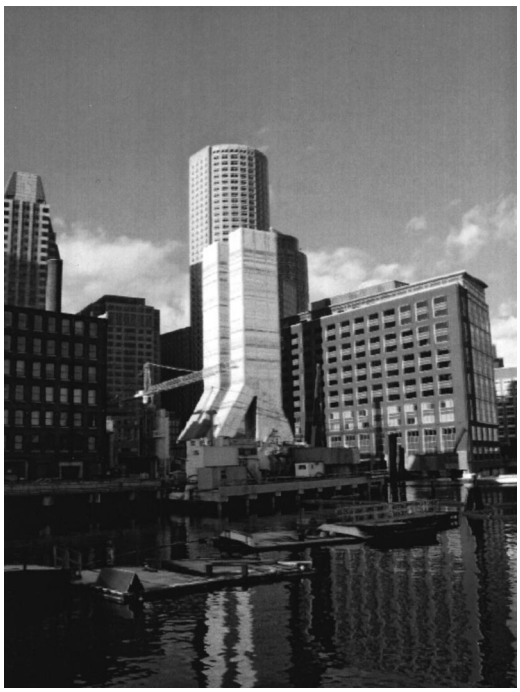
Vent Building 3 is a heavy poured concrete structure with its roof situated about 3 feet below grade. The vent building contains 23 100^{+}-in.-diam double-width centrifugal fans that, for a number of reasons, are not vibration isolated. The fans exhaust into a cluster of reinforced concrete shafts that rise up to about 250 ft above grade as shown in the accompanying figure. The land on which the vent building is situated is owned by Intell Management. The tall stacks are to be embedded in a luxury hotel/residential condominium complex that Intell is presently constructing. The new building is to be built atop the vent building and is also to be

ASA/GBC members who attended the tour.

partially supported by the stacks. The brief presentation by Greg Tocci after the tour, at 3 Cheers Restaurant, summarized the history of work on the project, including vibration measurements and SEA modeling, and the strategy used to determine whether elastomeric vibration isolation elements should be incorporated into structural connections with the vent building.

The acoustical analysis for this project has been a joint effort of Cavanaugh Tocci Associates, Inc. and R.H. Lyon Corp as acoustical consultants to the Owner (Intell) and Wilson Ihrig & Associates, Inc. as acoustical consultant to the Hotel Tenant (Six Continents Hotels).

Because of the complexity of the building and the analysis, the presentation focused on a few specific areas. These are the structural connections between the stacks and floors of the hotel and residential floors, sound levels at 125 Hz, where the NC rating (by the tangency method) of fan noise in building spaces has been generally found to be dominant, and, finally, whether vibration isolated connections of the hotel floors with the vent building stacks are required. The work in this project intentionally excluded the most logical consideration, and that is the vibration isolation of fans in the vent building. The CA/T could not allow any changes to fan installations



Vent Building 3.

that the INTELL team might propose since, at the outset of the project, the installation of tunnel ventilation fans was the chief limitation in the timely opening of the northbound tunnel, which could not afford to be delayed.
GREGORY C. TOCCI

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.

2005

16–20 May	149th Meeting joint with the Canadian Acoustical Association, Vancouver, Canada [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].
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].
18–22 July	17th International Symposium on Nonlinear Acoustics, State College, PA [Anthony Atchley, The Pennsylvania State University, 217 Applied Research Lab Building, University Park, PA 16802; Tel.: 814-865-6364 E-mail: ISNA17@outreach.psu.edu; WWW: http://www.outreach.psu.edu/c&i/isna17]
17–21 Oct.	150th Meeting joint with Noise-Con, Minneapolis, MN, [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].

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

Revision List

New Associates

Adams, Alyssa M., Lewitz and Assoc., Inc., 1505 Bridgeway Ave., #12B, Sausalito, CA 94965
Ariel, Zeev, DOVL AR LTD, 63 Drezner St., Tel Aviv, 69497, Israel
Bant, Sarah P., 39 Windsor Dr., Broughton, FL CH4 0PN, United Kingdom
Bastasch, Mark J., 3227 Northeast 48th Ave., Portland, OR 97213
Biesheuvel, Arie, Univ. of Twente, Department of Mechanical Engineering, P.O. Box 217, Enschede 7500 AE, The Netherlands
Birli, Joseph, Audiopack Technologies, 4933 Neo Parkway, Garfield Height, OH 44128
Chang, Seo, The Univ. of Seoul, Environmental Engineering, Jeonnongdong 90, Seoul DO 130-743 Republic of Korea
Chen, Xuemei, Naval Research Laboratory, Code 7102, 4555 Overlook Ave., SW, Washington, DC 20375
Combet, Christophe, Parc de la Fontaine de Duvence, 13 rue Levacher Cintrat, Marcoussis 91462, France
Duryea, Brad T., DMG Systems, P.O. Box 941634, Houston, TX 77094
Elmkjaer, Torsten H. Leth, Terma A/S, Airborne Systems, Hovmarken 4, Lystrup, DK-8520, Denmark
Fenneman, Douglas J., U.S. Army Night Vision & Elec. Sens. Direct., Science and Technology Division, 10221 Burbeck Rd., Fort Belvoir, VA 22060
Flynn, Ronan, Athlone Institute of Technology, Electronic Engineering, Dublin Rd., Athlone, Co. Westmeath, Ireland
Frere, Kenneth A., 29202 Campbell Ave., Moreno Valley, CA 92555
Gorne, Thomas A., Hochschule Fur Musik Detmold, Erich-Thienhaus-Institut (Ton Meister Dept.), Neustadt 22, Detmold D-32756, Germany
Harvey, David R., Harvey Marshall Berling Assoc., LLC, 173 West 81st St., Suite 2, Lower Level, New York, NY 10024
Hastings, Aaron L., Purdue Univ., Mechanical Engineering, 585 Purdue Mall, West Lafayette, IN 47907-2040
Imai, Satomi, 8259 Christmas Court, Charlotte, NC 28216
Jiang, Jintao, 3265 Sepulveda Boulevard, Apt. 306, Los Angeles, CA 90034
Jin, Craig T., School of Electrical and Information Eng., Computing and Audio Research Lab., Bldg. J03, Maze Cresent, Sydney NSW 2006 Australia
Johnston, Stephen E., Johnston Test Cell Group, Senior VP, 1300 SouthEast Grand Boulevard, Oklahoma City, OK 73129
Jouppi, Norman P., Hewlett Packard, MS 1181, 1501 Page Mill Rd., Palo Alto, CA 94304
Kang, Gye Nam, Space Two Corp., Theather Engineering Div., Hyundai I-Space Jamsil II, Suite 1015, Bangil 2 dong, Songpa-gu, Seoul 138-827, Korea
Larson, Robert E., Exponent, Test & Engineering Center, 23445 West Pinnacle Peak Rd., Phoenix, AZ 85027
Lavie, Antoine, Universite D'Artois, Faculte des Sciences Appliquees, Technoparc Futura, Bethune, 62400 France
Lewis, Bradley W., BAi, LLC, 4006 Speedway, Austin, TX 78751
Liu, Wensen, American Technology Corporation, Advanced Development Dept., 13114 Evening Creek Dr., South, San Diego, CA 92128
Marque, Jeffrey J., 2831 Fernwood St., San Mateo, CA 94403
Marshall, Robert L., Ecophon Certain Teed, Inc., Marketing, 145 Keystone Dr., Montgomeryville, PA 18936
Maslak, Samuel H., 961 High Rd., Woodside, CA 94062
Mathieu, Jerome, 37, Impasse des Millepertuis, Montpellier 34070, France
Mathuria, Parag H., 827 Arrowhead Dr., Apt. A, Sidney, OH 45365

Neubauer, Reinhard O., Ing.-Buero Dr. Neubauer VDI, Bldg. Physics & Acoustic Consultancy, Theresienstrasse 28, Ingolstadt, Bavaria 85049, Germany
Patton, Scott L., 53 Quannacut Rd., Westerly, RI 02891
Poissant, Sarah F., Communication Disorders, Univ. of Massachusetts, 125 Arnold House, Amherst, MA 01003
Powell, Kevin M., Charles M. Salter Assoc., 130 Sutter St., San Francisco, CA 94104
Prezelj, Jurij, Univ. of Ljubljana, Faculty of Mech. Eng., Lab. for Tech. Acoustics, Askerceva 6, Ljubljana SL 1000 Slovenija
Rivard, Thomas H., San Francisco Health Department, Environmental Health, 1390 Market St., San Francisco, CA 94102
Rothermich, Wayne, 6574 Covington Way, Goleta, CA 93117-1513
Singh, Moninderjit, P.O. Box 92487, Dubai, U.A.E. Aldes
Sitter, Max C., Max's ''Acoustical Sound Consultants'', P.O. Box 3454, Fullerton, CA 92834-3454
Smiljanic, Rajka, Linguistics, Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208
Smith, Gordon P., Western Kentucky Univ., Physics and Astronomy, 1 Big Red Way, Bowling Green, KY 42101
Smith, Rachel H., Dept. of Linguistics, Univ. of Cambridge, Sidgwick Ave., Cambridge CB3 9DA, United Kingdom
Snellen, Mirjam, Delft Technical Univ., Dept. of Earth Observ. & Space Syst. (DEOS), Kluyverweg 1, Delft 2629 HS, The Netherlands
ter Haar, Gail R., 24 Spencer Hill, Wimbledon, London SW19 4NY, United Kingdom
Tian, Xinhao, 330R, Michener PK, Northwest, Edmonton, Alberta T6H 4M5, Canada
Tidd, Richard A., Benthos Inc., Geophysical, 49 Edgerton Dr., North Falmouth, MA 02556-2826
Tsiappoutas, Kyriakos, 215 Masters Dr., Bloomington, IL 61704
Vincent, Harold T., Mikel, Inc., 151 Martine St., Fall River, MA 02723
Wang, Yadong, Dept. of Biology and Linguistics, Univ. of Maryland, 1401 Marie Mount Hall, College Park, MD 20742
Watanabe, Yuko, School of Information Environment, Tokyo Denki Univ., 2-1200, Muzai Gakuen-dai, Inzae, Chiba, 270-1382 Japan
Winter, Joseph F., Dept. Foreign Affairs & Trade, ATS, R.G. Casey Bldg., John McEwen Cr., Barton, ACT 0221, Australia
Xu, Li, Ohio Univ., School of Hearing, Speech & Language Sciences, Grover Center-W229, Athens, OH 45701
Yang, Huiyu, Feilo-USA, 4900 Harwood Court, Durham, NC 27713
York, Donald A., CRS Engineering, Inc., A/E Design, 109 Jefferson St. North, Suite 11, Huntsville, AL 35801
Zimmerman, William R., 11528 Woodmar Ln. Northeast, Albuquerque, NM 87111
Zsiga, Elizabeth C., Georgetown Univ., Linguistics, 37th and O St., North-West, Washington, DC 20057

New Students

Adams, Jason B., USC, Engineering, 3601 Watt Way-GFS 301, Los Angeles, CA 90089-1693
Ames, Heather M., 8 Vinal St., #10, Brighton, MA 02135
Argo IV, Theodore, 11915 Stonehollow Rd., #1333B, Austin, TX 78758
Ballard, Jeffrey A., Florida Atlantic Univ., Ocean Engineering, 101 North Beach Rd., Dania Beach, FL 33004
Bardhan, Neil P., Univ. of Rochester, Brain & Cognitive Sciences, 457 Meliora Hall, River Campus 270268, Rochester, NY 14627
Barnard, Andrew R., Penn State Univ., Acoustics, P.O. Box 30, State College, PA 16804
Barnhill, Colin L., 604 West 37th St., Baltimore, MD 21211
Barriault, Patrick, 2630 Prospect, App. #305, Sherbrooke, QU J1J 4G2, Canada
Borchert, Elizabeth M.O., Univ. of Minnesota, Psychology, 75 East River Rd., Minneapolis, MN 55455
Boucher, Matthew A., 508 East Graham Place, Bloomington, IN 47401
Brady, John, 1400 Martin St., #1027, State College, PA 16803
Brasovan, Philip J., 620 West Belmont Ave., Chicago, IL 60657
Bush, Adam R., 1606 Jefferson St., Madison, WI 53711
Callaway, Jason E., Seattle Arts Institute, Audio, 2323 Elliot Ave., Seattle, WA 98121
Carlson, Christopher, P.O. Box 5911 JMU, Harrisonburg, VA 22807
Chandrayadula, Tarun K., George Mason Univ., ECE, 4400 University Dr., Fairfax, VA 22030
Charmes, Emmanuel, 612 NW 13th St., Apt. 21, Boca Raton, FL 33486
Chatterley, James J., 7 Wymount Terrace, Provo, UT 84604
Chinchilla-Rodriguez, Sherol S., House Ear Inst., DAIP, 2100 West 3rd St., Los Angeles, CA 90057
Clark, Linde J., Penn State Univ., Graduate Program in Acoustics, P.O. Box 30, State College, PA 16804
Cook, Ian M., 42 Gertrude Dr., Middletown, NY 10940
Corcoran, David E., 16790 Highland Valley Rd., Ramona, CA 92065
Cutliffe, Steven P., 2312 Bluebird Ln., Rolling Meadows, IL 60008
Debever, Claire, Scripps Institution of Oceanography, Marine Physical Laboratory, 8820
Shellback Way, NTV Bldg., Room 446, La Jolla, CA 92093-0238
Defreville, Boris, 35 rue Poulet, Paris 75018, France
Ernst, Jason P., Univ. of Hartford, Engineering, 200 Bloomfield Ave., Box #2880, West Hartford, CT 06117
Fung, Mary, 1814 West 7th St., Brooklyn, NY 11223
Garner, Cassie A., 3721 North 2nd St., Lincoln, NE 68521
Goor, Mina, 1212 Southwest 16th Ave., #E, Gainesville, FL 32601
Green, Matt J., 164 North 400 West, #14, Provo, UT 84601
Gros, Emmanuelle, 463 Galt Ouest, Sherbrooke, QU J1H 1Y5, Canada
Haas, Michael, UW Platteville, 002 Morrow Hall, 1300 Greenwood Ave., Platteville, WI 53818
Hanson, Adrian L., 831 Peregrine Circle, Oregon, WI 53575
Heavens, David, 3 Greenhill Rd., Leicester LE2 3DN, England
Heidari, Amir Homayoun, North Carolina State Univ., Civil Engineering, 208 Mann Hall, Stinson Dr., Raleigh, NC 27604-7908
Hennes, Christopher C., 150 Northbrook Ln. #215, State College, PA 16803
Herrera, Luis J., Cra. 97 #86 B 15 Ap. 106, Bogota, Quirigua, Colombia
Herrmann, Kristen G., 2175 South Ocean Blvd., #303, Delray Beach, FL 33483
Hong, Robert S., 2325 Willowbrooke Ln., Iowa City, IA 52246
Horton, Gregory P., 7 State Dr., Harrisburg, PA 17112
Hurd, James R., 19353 Surrey Ln., Northville, MI 48167
Isom, Adam, Box 1737, 2000 North Parkway, Memphis, TN 38112
Ito, Yasuko, 2333 Kapiolani Blvd., Apt. #3015, Honolulu, HI 96826
Jaworski, Justin W., Duke Univ., Mechanical Engineering Department, Box 90300, Durham, NC 27708
Jong-Yeon, Kim, Catholic Univ. of Daegu, School of Med., Preventive Medicine & Public Health, 3056-6 Daemyung-4 Dong, Nam-Gu, Daegu 705-718 Korea
Juntarapaso, Yada, Graduate Program in Acoustics, School of Acoustics, P.O. Box 30, State College, PA 16804
Kaczynski, Adam, Michigan Technological Univ., 1801 Townsend, Dr., Room 232, Houghton, MI 49931
Kao, Han, 205 Cunningham Hall, White Course, University Park, PA 16802-6840
Kinney, Andrew W., 51 Starr Rd., Montoursville, PA 17754
Kondaurova, Maria V., 1104 North Salisbury St., #45, West Lafayette, IN 47906
Landis, Cody J., 3151 Calimyrna Rd., Acampo, CA 95220
LeBlanc, Jared B., 514 Kevin Dr., Lafayette, LA 70507
Lee, Connie, 3910 North Fremont St., Unit J, Chicago, IL 60613
Lie, Ki, 270 Littleton St., Apt. 435, West Lafayette, IN 47906
Liu, Jerry C., Univ. of Southern California, Linguistics, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693
Liu, YaoJan, 1950 East Stoll Rd., Lansing, MI 48906-1074
Lloyd, James D., 128 Green Forest Estates, St. Peters, MO 63376
Lopez, Miguel I., 3320 West 60th Place, Chicago, IL 60629
Lyons, Joy E., 217 I Alley, State College, PA 16801
Machac, Tamara, Trinity College, 300 Summit St., Box 700921, Hartford, CT 06106
Mantha, Sravan P., Ray W. Herrick Laboratories, Purdue Univ., Mechanical Engineering, 140 South Intramural Dr., West Lafayette, IN 47907
Marui, Atsushi, 3710 Rue Peel, Room 1, Montreal, QU H3A 1W9, Canada
Mattock, Karen J., MARCS Auditory Laboratories, Univ. of Western Sydney, Locked Bag 1797 Penrith South DC, Sydney, NSW 1797, Australia
McCoy, Sandra L., 34 Walnut St., #3, Waltham, MA 02453
McGuire, Erica, 117 Bishop Hall, East Spring St., Oxford, OH 45056
Miah, Khalid H., P.O. Box 7155, Austin, TX 78713
Miao, Hongyu, 60 Crittenden Blvd., Apt. #400, Rochester, NY 14620-4013

Mollman, Emily M., 7434 Frederiksen Court, Ames, IA 50010
Monet-Descombey, Julien A., Universite de Sherbrooke, Genie Mecanique, 2500 Boulevard de l'Universite, Sherbrooke, QU J1K 2R1, Canada
Naka, Yusuke, Boston Univ., 677 Beacon St., Boston, MA 02215
Nutter, David B., Physics and Astronomy, Brighan Young Univ., N283 ESC, Provo, UT 84602
O'Connor, Steve P., Australian National Univ. (ANU), RSISE, The Australian National Univ., Canberra ACT 0200, Australia
Oliveira, Leonardo, Yale Univ., Linguistics, 370 Temple St., Room 210, New Haven, CT 06511
Pape, Nathan A., 200 Bloomfield Ave., Box 1608, West Hartford, CT 06117
Perimeter, Mike R., 1011 South Valentia St., #18, Denver, CO 80247
Perkins, Micah L., 4713 McCord Rd., Ponca City, OK 74604
Perrot, Camille, GAUS-Universite de Sherbrooke, Genie Mecanique, 2500, boul. de l'Universite, Sherbrooke, QC J1K 2R1, Canada
Richardson, Kevin R., 134 Birchtree Court, State College, PA 16801
Riggs, Daylen B., USC, Linguistics, 3601 Watt Way-GFS 301, Los Angeles, CA 90089-1693
Robinette, Brandon L., 3946 North Bell Ave., Chicago, IL 60618
Rodda, Judith L., Univ. of Otago, Zoology, 340 Great King St., Dunedin, Otago 09001, New Zealand
Ronsee, Lauren, 1312 Louisiana St., Lawrence, KS 66044
Root, Benjamin J., 17 Warren Terrace, West Hartford, CT 06119
Rozema, Randall T., 319 Forest Ave., Bellefonte, PA 16823
Ryherd, Steven R., 1601 South 64th St., Omaha, NE 68106
Sarangapani, Sairajan, Univ. of Rhode Island, Ocean Engineering, 217 Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882
Shafer, Benjamin, 675 North 800 East, Provo, UT 84606
Slaney, Jacob H., 1328 Stonehedge Dr., Birmingham, AL 35235
Smith, Jessalyn, 155 Mohawk School Rd., New Castle, PA 16102
Stimpert, Alison K., Mammal Research Program, Hawaii Inst. of Marine Biology, 46-007 Lilipuna Rd., Kaneohe, HI 96744
Szeto, Wai Man, Flat 15A, Block 2, Villa Athena, Ma On Shan, N.T. Hong Kong
Tanen, Robert M., Acoustical Engineering, Univ. of Hartford Branch, 2986 Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117
Taylor, Kristen A., Marine Mammal Research Program, Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, HI 96734
Thirwani, Kapil C., 2 Tintern Ave., London, Kingsbury, NW9 0RJ, United Kingdom
Thompson, Eric R., 8201 Surrey Court, Fort Wayne, IN 46815
Tono, Henrique L., 212 North Duke St., #214, Durham, NC 27701
Valmont, Elizabeth, 1047 South Westlake Ave., Los Angeles, CA 90006
Waters, Zachary J., 89 Ocean View Highway, Westerly, RI 02891
White, Robert D., 2222 Fuller Court, 602A, Ann Arbor, MI 48105
Wickstrom, Sueann M., 968 NW First Ave., Chisholm, MN 55719
Yack, Tina M., 7212 Vassar Ave., La Mesa, CA 91941
Yealy, Christopher M., 90 Foxtown Dr., Abbottstown, PA 17301
Zettler, Cynthia M., 116 Summer Terrace Ln., Atlanta, GA 30342

New Electronic Associates

Anderson, Richard H., Dynamics Technology, Inc., 1555 Wilson Blvd., Suite 703, Arlington, VA 22209-2405
Barnes, Jennifer J., Barnes Music, P.O. Box 57, Broadmeadow, NSW 2292, Australia
Barrable, Ross, Harmony Sound Gardens, PMB 4301 Suite B-3, 63 North Pagosa Blvd., Pagosa Springs, CO 81147
Blackwood, Diane J., 3990 Elwood Court, Concord, CA 94519
Brunskog, Jonas, LTH, Lund Univ., Engineering Acoustics, P.O. Box 118, Lund SE 221-00, Sweden
Buchanan, Robert A., 7 Stratford Place, St. John's, NL A1G 1T5, Canada
Conoir, Jean-Marc, LAUE, Univ. du Havre, place R. Schuman, Le Havre 76610, France
Daves, Brian W., 3707 Queen Ave., North, Minneapolis, MN 55412
Goodman, H. D., 4410 Lees Corner Rd., Chantilly, VA 20155
Graham, Joelle G., 25 Birchview Rd., Nepean, ON K2G 3G3, Canada
Hundley, Allen, DPS Technology, 5662 Calle Real, #311, Goleta, CA 93117
Kennedy, Geoffrey P., RSW Engineering, Inc., 1700 Broadway St., #404, San Francisco, CA 94109
Kristiansen, Kim, Dali A/S, Project Department, Dali Alle 1, Noerager 9610, Denmark
Livengood, Kim J., Univ. of Missouri, Fisheries and Wildlife Sciences, 302 ABNR, Columbia, MO 65211
Luppe, Francine, LAUE, Universite du Havre, place R. Schuman, Le Havre, NO 76610, France
Masahiro, Ikeda, Yamaha, 6600 Orangethorpe Ave., Buena Park, CA 90620
Miklovic, Donald W., AETC Inc., 8910 University Center Ln. #900, San Diego, CA 92122
Moleti, Arturo, Physics, Univ. of Roma Tor Vergata, Via della Ricerca Scientifica 1, Roma 00133 Italy
Nielsen, Jakob M., Dali A/S, Project Department, Dali Alle 1, Noerager 9610, Denmark
Nishiguchi, Isoharu, Kanagawa Institute of Technology, System Design Engineering, 1030, Shimoogino, Atsugi-shi, Kanagawa-ken 243-0292, Japan
Pannell, Chris N., Optronic Laboratories Inc., 4632 36th St., Orlando, FL 32811
Quick, Andrew, SonaMed Corporation, 1250 Main St., Waltham, MA 02451
Reddy, Balakrishna, Speech Processing, Elvista Digital System Pvt. Ltd., 4121, 6th Cross, 19th Main, HAL 2nd Satge, Bangalore, Karnataks, 560008 India
Reed, Janice L., GeoSonics Inc., 359 Northgate Dr., Warrendale, PA 15086
Swain, William B., 2620 Stratford Rd., Lawrence, KS 66213
Van Parijs, Sofie M., Cornell Univ., Lab. of Ornithology, 159 Sapsucker Woods Rd., Ithaca, NY 14850
Wade, Travis W., Carnegie Mellon Univ., Psychology, 5000 Forbes Ave., Pittsburgh, PA 15213
Wheeler, Robert L., 1754 Toepfer Rd., Akron, OH 44312

New Corresponding Electronic Associates

Cherniavskyi, Konstiantyn E., Artema St. 58/2, Apt. 29, Kiev 04050, Ukraine
de Magalhaes, Fernando L., Rua Aguiar, 57, Rio de Janeiro 20261-120, Brazil
Kim, Jeehyun, Sungdong-Gu, Oksu-Dong, Samsuing Apt. 109-501, Seoul, 133-100, South Korea
Qu, Tianshu, Center for Information Science, Peking Univ., Beijing 100871, China
Woo Hun, Ryang, Chonbuk National Univ., Div. Science Education, College Education, Jeonju, Jeonbuk 561-756, Republic of Korea

Reinstated

R. R. Capranica—*Fellow*
K. A. Siren—*Associate*
I. N. Pieleanu—*Student*
X. Sun—*Electronic Associate*

Students to Associates

D. Baskent, J. R. Baughman, K. J. Benoit-Bird, C. G. Clopper, W. R. Elsberry, A. L. Gerig, S. Hakim, M. K. Norris, M. E. Poese, B. L. Southall, L. A. L. Souza

Students to Electronic Associates

R. Balachandran, S. Fuchs, Y. E. Kim, B. S. Nelson

Associates to Electronic Associate

S. D. Browne, D. B. Gerhard, A. K. Herrmann, L. A. Leske, A. J. Schafer

Associate to Corresponding Electronic Associate

E. M. Yiu

Members to Electronic Associates

C.-Y. Tseng

Electronic Associate to Associate

J. Eldredge, S.-M. Kim, K. Krumbholz, R. E. Larson

Resigned

R. Folsom, D. Godfrey, D. McCandless—*Members*
J. Andrade, T. Burke, S. M. Reynolds, R. E. Stone, Jr.—*Associate*
C. Olsen—*Student*

Deceased

C. H. Allen, I. Malecki, J. B. Moreland—*Fellows*
A. R. Condom—*Member*

Fellows	905
Members	2386
Assoc.	2605
Students	897
Electronic Associates	248
	7041

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).]*

Physical Principles of Medical Ultrasonics, 2nd edition

C. R. Hill, J. C. Bamber, G. R. ter Haar (Editors)

John Wiley & Sons, Chichester, 2004.
xv +511 pp, $199.00 (hardcover), ISBN: 0471970026

This edition follows the highly regarded 1986 edition for which the stated purpose was ''to examine the subject of medical ultrasonics from the point of view of underlying physical principles.'' The intention was to present basic material ''in a manner that should help newcomers to the field to contribute effectively to its further advancement.'' Great advances have indeed been made in the field since then; diagnostic ultrasound now accounts for over 20% of all medical imaging procedures worldwide. The 1986 authors deserve their share of the credit. While most of the first edition is still relevant—and has been retained—much new material has been added. As before, the book is designed mostly for graduate students in physical sciences and engineering, but will also be helpful to motivated undergraduates in these fields, as well as to other physically oriented medical personnel.

Chapter 1 presents relevant aspects of acoustic theory, to which reference is made in later chapters. There is emphasis on ultrasonic pulses in lossless homogeneous media, including examples in which Green's functions are used and edge waves discussed. In Chapters 2 and 3 a wealth of practical information is offered for generating medical ultrasound fields, especially with arrays that can be focused and steered. Equally valuable is the attention given to measurement of field characteristics involving power, intensity and acoustic pressure, believed to be the most important for their relevance to imaging capabilities, and to the potential for biological effects. If I were a student or other person planning to do research or pursue some application using ultrasound transducers, I would want these chapters available to me as a guide and source of general information.

Chapters 4 and 5 deal broadly with attenuation, absorption, dispersion, and the speed of sound. They update the very extensive treatments in 1986, especially in regard to contrast agents, therapy and nonlinear propagation. In Chapter 6, a theoretical basis is outlined for understanding reflection and scattering of ultrasound in tissues. In the future, a more complete understanding of scattering from tissue structures is expected from the use of acoustic microscopy, and from employment of reconstruction techniques analogous to those used in x-ray computed tomography.

Chapter 7 takes up studies of the interaction between ultrasound and tissues by considering their physicochemical aspects. It is pointed out that, for many tissues, existing differences in shear properties are much greater than in bulk properties (such as density). It follows that techniques based on shear properties would be more sensitive to tissue differences.

The role of the human observer is examined in Chapter 8. It is shown that characteristics of the ''eye of the beholder'' influence both the design and effectiveness of the overall imaging procedure. Quantitative measures of imaging and perception are discussed.

In Chapters 9, 10, and 11 the authors describe a methodology for clinical investigation with the two stated aims: describing in outline the present state of the art; and exploring ideas and principles presented in earlier chapters that have promise for the future. Opportunities are seen for flow-imaging and elastography to become more quantitative. Also, it is anticipated that the introduction of stable, safe echo-enhancing contrast agents will open up a new range of possibilities for ultrasound imaging. The principles of operation of the SAM and SLAM microscopes are explained; these instruments provide information complementary to that given by optical microscopes.

Chapter 12 examines a number of physical phenomena that are believed to account for the various types of link between ultrasonic exposure and biological effect. These include heat generation, acoustic cavitation, radiation forces and torques, and acoustic streaming. Comparisons are made between theory and experiment. Therapeutic and surgical applications of ultrasound are taken up in Chapter 13, with emphasis on work that is considered to have a reasonably well-established scientific basis. Included are ultrasonically enhanced drug delivery, physiotherapy for pain control and fracture healing, focused beam surgery, ultrasound angioplasty, and treatment of MeniΠre's Disease.

In Chapter 14 consideration is given to various lines of evidence that have been brought to bear on questions of safety. Definitions of relevant biological endpoints are discussed, as well as the concepts of safety and hazard in the contexts of therapeutic and diagnostic practice. It is concluded that epidemiological surveys based on earlier procedures have found little firm evidence of deleterious effects from clinical use of diagnostic ultrasound. Questions continually arise, however, as new procedures and conditions are introduced; examples during the past decade are (i) the use of higher intensities and (ii) the employment of gas-containing contrast agents. While much good can come from these and other changes, the authors stress that there must be continued vigilance in recognizing and studying possibilities they might present for unwanted effects.

This volume, like the first edition, treats the physical aspects of medical ultrasound that are key to understanding how it works. Much valuable information is contained in its pages, and it is presented carefully and well. I know of no other book that treats this subject with comparable breadth and depth. It should be available for study and for reference by anyone who is involved with research into topics related to medical ultrasound, or with development of applications, and with providing technical help to clinical users.

WESLEY L. NYBORG
Physics Department
University of Vermont

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
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*
NEIL A. SHAW, Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Univ. of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,771,562

43.30.Sf FISH FINDING METHOD AND SYSTEM

David A. Betts *et al.*, assignors to Techsonic Industries, Incorporated
3 August 2004 (Class 367/111); filed 26 August 2002

A small bobber attached to a fishing line houses a transducer, sonar transmitter and receiver, microprocessor, rf transmitter and antenna, and battery. Echo returns indicating the presence of fish near the bobber position are converted to electrical signals and radio transmitted to a receiver at the fisherman's shore position, where the echoes are viewed on some display device. This display device may, in fact, be a small viewing screen attached to the fisherman's wrist or to the fish pole. The poor fish ain't got a chance.—WT

6,771,563

43.30.Tg APPARATUS FOR DEPLOYING A LOAD TO AN UNDERWATER TARGET POSITION WITH ENHANCED ACCURACY AND A METHOD TO CONTROL SUCH APPARATUS

François Bernard, Den Haag, the Netherlands
3 August 2004 (Class 367/131); filed 20 September 2002

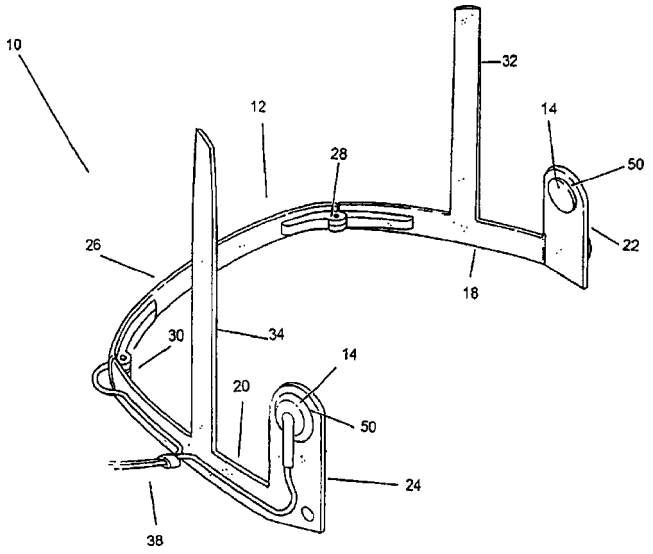
An apparatus for delivering a load to an underwater target position incorporates a sonar system that generates a number of acoustic beams to aid in target location as well as equipment to measure the sound velocity profile in the water surrounding the apparatus so that local bending of the sound rays can be computed to allow making real-time corrections of the target location.—WT

6,769,508

43.30.Tg UNDERWATER AUDIO SYSTEM

Damian Victor O'Connor, Brentwood Western Australia, Australia
3 August 2004 (Class 181/129); filed in Australia 21 May 1999

An underwater audio system consists, in part, of the device shown in the figure. The hinged support frame **12** features two arms **32** and **34** which



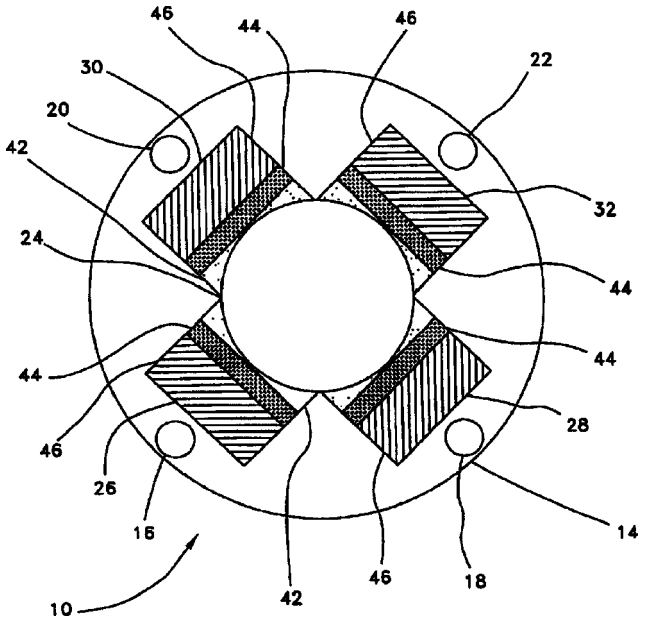
are designed to be inserted under the strap of a diver's mask. The frame is so sized that the two waterproofed speakers **14** are then held adjacent to the diver's ears. The cables **38** connect to an auxiliary portable audio device. In an alternate embodiment, there is an extension of frame **12** to the front so that a waterproofed microphone is held adjacent to the diver's mouth.—WT

6,778,470

43.30.Xm MOUNTABLE SYNTACTIC FOAM SENSOR HOUSING

Patrick J. Monahan and Robert S. Janus, assignors to The United States of America as represented by the Secretary of the Navy
17 August 2004 (Class 367/188); filed 15 September 2003

An acoustic sensor is described for monitoring the retrieval of a towed array deployed from a submarine. The sensor, shown in cross sectional view, incorporates four piezoceramic transducers **26**, **28**, **30**, and **32** surrounding a cylindrical passage tube **24**. Two of the transducers function as transmitters and the opposite two as receivers. Items **44** are the piezoceramic elements while items **46** are backing plates. The transducers are bonded to the passage



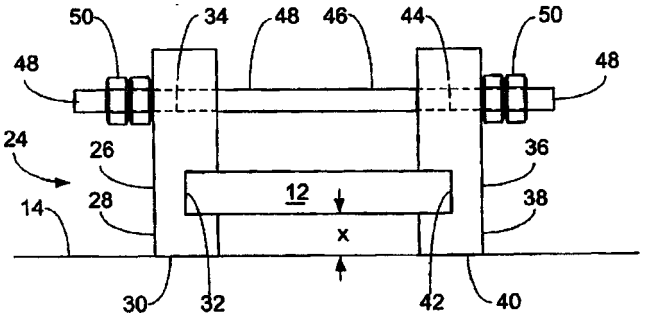
tube with a semi-rigid adhesive such as polyurethane which then forms acoustically transparent windows **42**. The remainder of the internal volume of the sensor is encapsulated with an epoxy composite whose density is adjusted by the addition of hollow microspheres to lighten the weight of the whole sensor. In operation, as a towed array is drawn in through the water-filled passage tube **24**, the pair of perpendicular beams generated by the two transmitting transducers results in different received signals at the two hydrophones depending on whether the towed array has been fully retrieved or not.—WT

6,771,564

43.30.Yj SONAR DOME AND A MOUNTING BRACKET FOR REMOVABLY CONNECTING AN ACOUSTIC SENSOR ELEMENT TO A SONAR DOME

Thomas S. Ramotowski and Patrick J. Monahan, assignors to The United States of America as represented by the Secretary of the Navy
3 August 2004 (Class 367/173); filed 15 September 2003

A pair of mounting brackets **28** and **38** for holding each of a set of piezocomposite transducer elements **12**, generally in the shape of a rectangular parallelepiped, to the inner wall **14** of a sonar dome are fashioned from flexible acoustically transparent material such as polyurethane. The two brackets are adhesively bonded to the dome at interface surfaces **30** and **40**. Because the brackets are flexible, they can be bent apart to initially insert, or



subsequently remove and replace, any one of the transducer elements **12**. Slots **32** and **42** cut into the brackets hold the transducer element at a small distance **x** away from the inner wall of the dome which, of course, is free-flooded. A restraining member such as a threaded PVC rod **48** can be inserted through bore holes **34** and **44** in the mounting brackets to lock the transducer element in position.—WT

6,778,468

43.30.Yj AUTOMATICALLY TRACKING SCANNING SONAR

Yasushi Nishimori and Ken'ich Watanabe, assignors to Furuno Electric Company Limited
17 August 2004 (Class 367/103); filed in Japan 20 May 2002

The elements of a large sonar array, generally cylindrical or spherical in shape, are excited in various combinations and with various signals, such as CW or FM, so that the frequency of the radiated signal is changed as the direction of radiation is changed. The goal is to reduce spurious false target readings because of grating lobes and high side lobes.—WT

6,760,455

43.38.Bs ELECTROSTATIC LOUDSPEAKER WITH A DISTRIBUTED FILTER

James J. Croft III and Robert C. Williamson, assignors to American Technology Corporation
6 July 2004 (Class 381/191); filed 13 July 2001

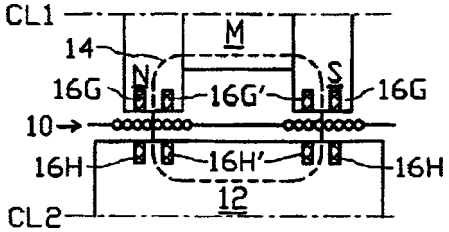
To control the directional pattern of an electrostatic loudspeaker, the fixed electrode can be divided into multiple sections connected through resistors of different values. This patent argues that it is more efficient to control the resistivity per unit area of the diaphragm itself. In this way, a distributed network is created in which the acoustic output is attenuated with increasing frequency at all points distant from the electrical connection area.—GLA

6,768,806

43.38.Dv SHORTING RINGS IN DUAL-COIL DUAL-GAP LOUDSPEAKER DRIVERS

Douglas J. Button *et al.*, assignors to Harman International Industries, Incorporated
27 July 2004 (Class 381/401); filed 18 March 1999

Dual-voice-coil dual-magnetic-gap electrodynamic transducers are a specialty of the assignee and shorting rings have been used in magnetic motor topologies to reduce voice coil inductance and therefore the harmonic distortion due to the inherent nonlinearities of the voice-coil/gap structure. The patent describes how, by strategically placing one or more annular shorting rings, the voice coil inductance can be reduced (placement near the voice coil), second and higher order even harmonic distortion can be re-



duced (placement entirely within the magnetic flux loop centerline), and third and higher order odd harmonic distortion can be reduced (placement centered in the magnetic gap). One of the embodiments of the invention shows shorting rings **16G**, **16G′**, **16H**, and **16H′** located near the voice coil that are said to mainly suppress second and higher order even harmonic distortion and to reduce voice coil inductance.—NAS

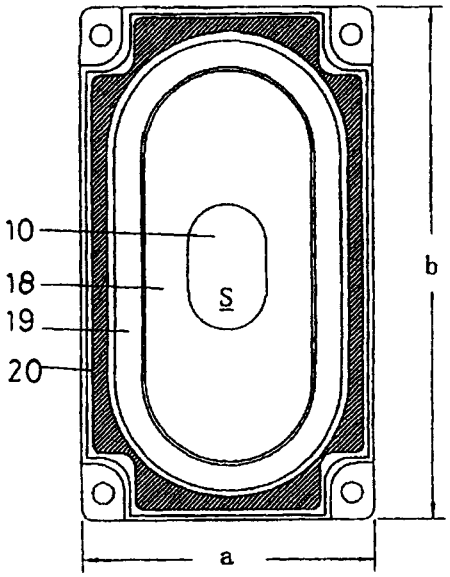
6,771,792

43.38.Dv TRACK-SHAPED LOUDSPEAKER

Hidetoshi Shirakawa, assignor to Foster Electric Company, Limited

3 August 2004 (Class 381/412); filed in Japan 28 December 1998

The aesthetics of television sets and similar entertainment equipment plays a large role in the industrial design of these devices. The patent admits that thin is in regarding the form factor industrial designers desire in the physical shape of the loudspeakers they place along the perimeter of large screen displays and other such visual devices. These elongated loudspeakers are not known for their sound output. The patent appears to describe an



oval-shaped magnetic gap topology **10**, and similarly shaped voice coil and attachment to the cone and suspension, that allows the magnetic circuit to be as large as possible, within certain constraints to prevent saturation, and to increase the magnetic flux density. The one claim concerns this oval, or racetrack-shaped, magnetic motor.—NAS

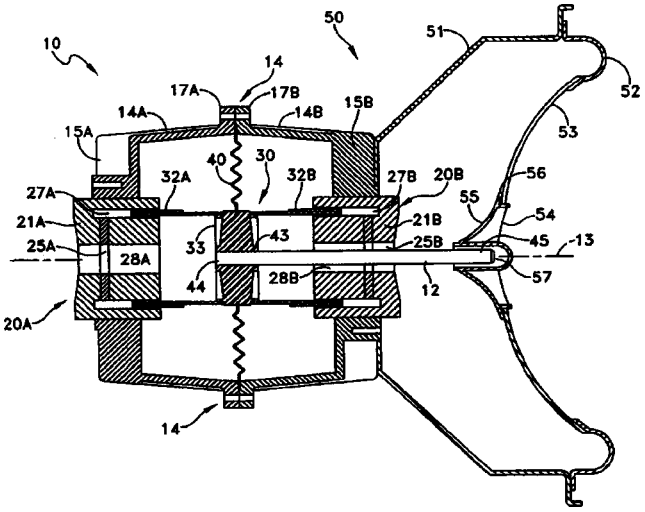
6,778,677

43.38.Dv REPAIRABLE ELECTROMAGNETIC LINEAR MOTOR FOR LOUDSPEAKERS AND THE LIKE

C. Ronald Coffin, Topsfield, Massachusetts

17 August 2004 (Class 381/418); filed 7 April 2003

The problem is that cone-type electrodynamic loudspeakers are prone to several failure modes. The suspension, the lead wires, the neck joint, or the cone can all fail, albeit usually not all at the same time. So, is there an easy way to repair these failures? In many cases, the recone technician removes all moving parts from the frame motor assembly and drops in a ''recone'' kit, gluing the suspension and spider and soldering the flex leads to the electrical terminals on the frame. Peavey has a line of loudspeakers that feature removable baskets that allow for the quick replacement of the cone, suspension, spider and voice coil. The basket can then be reconed at your leisure. The present patent describes an electromagnetic linear motor that has two voice coils, **32A** and **32B**, in a complex magnetic motor design



that drives a rod **12** that in turn is attached to the loudspeaker **50** through a special connection fitting. The motor structure can be attached or affixed to a variety of radiating mechanisms, including cones of concave and convex cross sections. In addition to the motor design, various and sundry means of attaching the cone and support structure to the motor are described. How this is an improvement over current methods of repair is not clear to this reviewer.—NAS

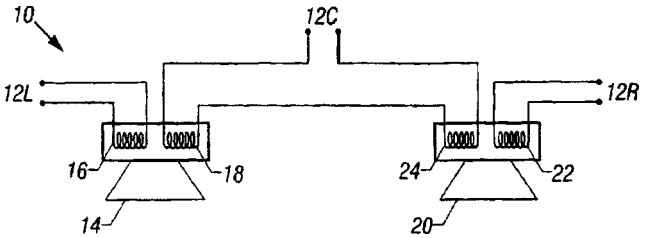
6,782,111

43.38.Dv MULTIPLE VOICECOIL AND DRIVER TRANSDUCING

Peter J. Wagner II and Charles R. Barker III, assignors to Bose Corporation

24 August 2004 (Class 381/335); filed 9 July 1998

Let's say you have a successful line of equipment that features two loudspeakers. How do you present three channels, say, left, center, and right, in those two loudspeaker channels? The patent describes how you can do this by using two loudspeakers, **14** and **20**, each with two independent voice coils, **16** and **18**, and **22** and **24**. By connecting a voice coil in the left loudspeaker with a voice coil in the right loudspeaker, in either series (as



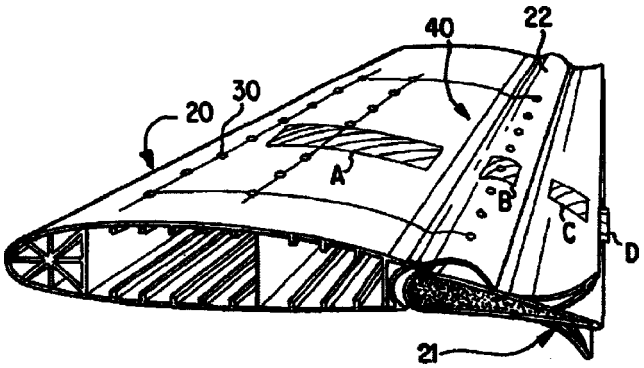
shown in the figure) or parallel connection, and then by driving this connected set of voice coils with the center channel signal and the remaining coil in each of the transducers with the left and right signals respectively, you can present the three channels. It may just be this reviewer's memory, but this scheme seems vaguely familiar.—NAS

6,752,020

43.38.Fx DEVICE FOR MEASURING PRESSURE, SOUND AND VIBRATION AND METHOD OF ANALYZING FLOW ON SURFACES OF STRUCTURAL PARTS

Gerald Sobotta and Dieter Bosch, assignors to EADS Deutschland GmbH
22 June 2004 (Class 73/702); filed in Germany 20 May 1999

This patent describes a method and apparatus for analyzing fluid flow on missiles or aircraft. If sensors are mounted on exterior surfaces, they can be subjected to hostile environments. Moreover, the presence of sensors disturbs the flow pattern that is being measured. The invention utilizes sensors imbedded in the structure or mounted on the interior of the skin. Signals in the range below 250 kHz are analyzed by artificial intelligence methods to determine the onset of turbulence, etc.—GLA



6,776,762

43.38.Fx PIEZOCOMPOSITE ULTRASOUND ARRAY AND INTEGRATED CIRCUIT ASSEMBLY WITH IMPROVED THERMAL EXPANSION AND ACOUSTICAL CROSSTALK CHARACTERISTICS

Kenneth R. Erikson *et al.*, assignors to BAE Systems Information and Electronic Systems Integration Incorporated
17 August 2004 (Class 600/459); filed 13 June 2002

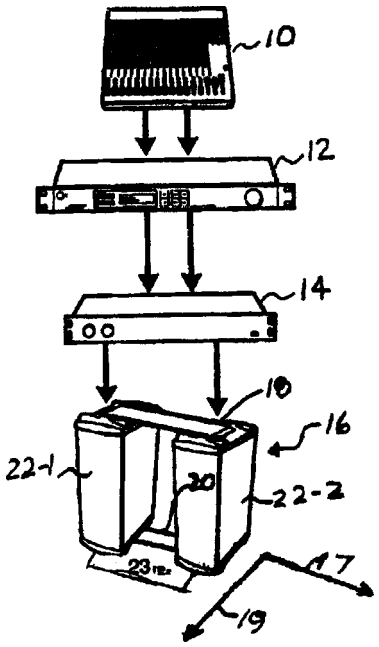
The aims of this ultrasound array are to provide a piezocomposite transducer array and integrated circuit (IC) assembly and to provide acoustical matching properties suitable to the frequency of the array and the medium to which it is being applied. The IC assembly features better acoustic isolation between transducer elements, greater mechanical reliability of the solder bond joining the array and the IC, and greater vertical attenuation of acoustic signals between the array and the IC. The device includes a piezocomposite ultrasound array with thermal expansion characteristics approximately matching silicon or other semiconductor material from which the ICs are constructed. Transducer elements are acoustically isolated laterally. Acoustical effects are isolated vertically with capacitive coupling and small area solder bumps relative to wavelength and further isolated laterally by thin supporting substrates relative to wavelength, including thinner semiconductor IC substrates.—DRR

6,766,033

43.38.Hz MODULAR BASS ARRAYING

Morten Jorgensen and Christopher B. Ickler, assignors to Bose Corporation
20 July 2004 (Class 381/387); filed 19 September 2001

It is known that a directional low-frequency loudspeaker array can be built using standard woofer boxes driven by suitably processed signals. A variety of common gradient patterns can then be created by selecting the appropriate electronic parameters. This patent describes such a system using two standard speaker modules **22-1**, **22-2**, a two-channel power amplifier **14**, and a two-channel signal processor **12**.—GLA



6,754,360

43.38.Ja SIGNAL-HORN

Attilio Granziera, assignor to Stebel SPA
22 June 2004 (Class 381/340); filed in Italy 4 February 2000

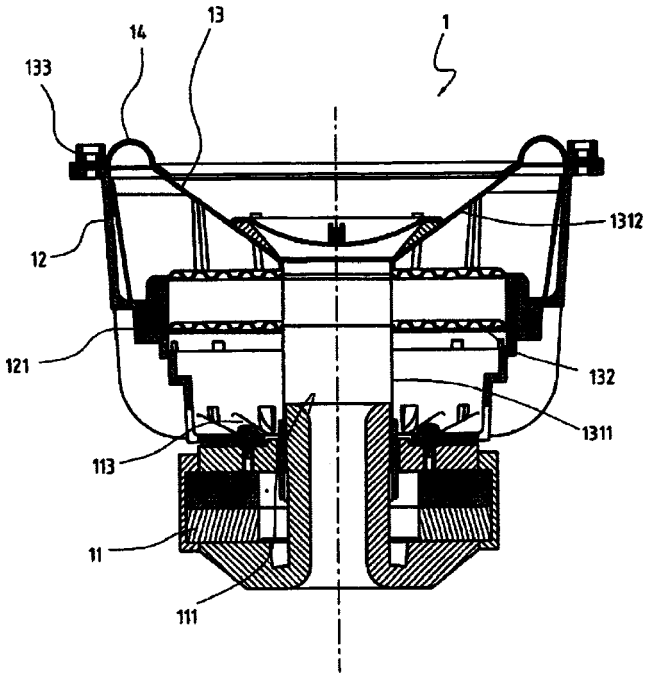
To an automobile manufacturer, the ideal horn assembly would be very small, very cheap, and very loud. This patent describes a horn design intended to deliver maximum loudness from an electromagnetic driver at a given frequency (typically between 380 and 530 Hz). A spiral sound path takes the form of an extended cylindrical throat section followed by an exponential flare. The patent includes a comparative chart of three horns that lists ''sound power'' in dB(a) but does not include the electrical power drawn by the three drivers.—GLA

6,757,402

43.38.Ja KNOCKDOWN SPEAKER

Yen-Chen Chan, Hsin Chuang City, Taipei Hsien, Taiwan, Province of China
29 June 2004 (Class 381/396); filed 14 November 2001

This somewhat elongated loudspeaker consists of three self-contained assemblies: a magnetic base **11**, a casing **12**, and a vibration unit **13**. When the music lover blows out a loudspeaker, he simply buys a new vibration



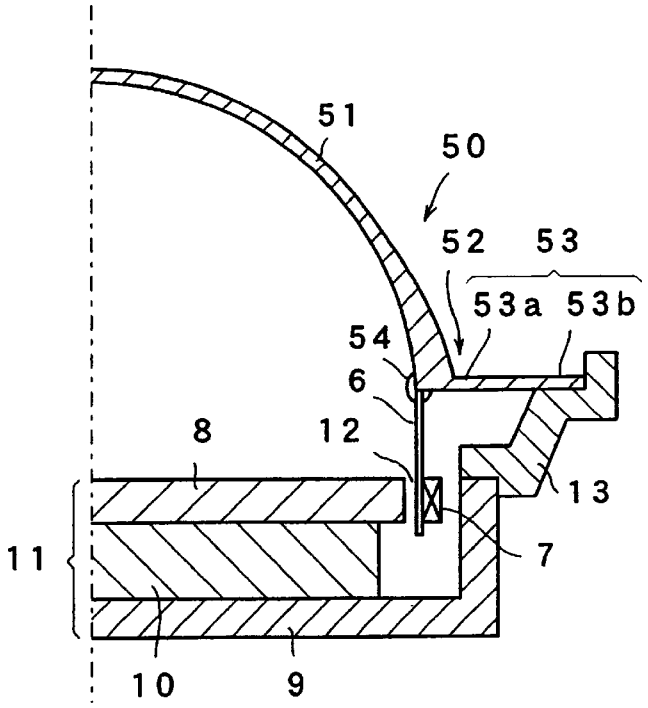
unit and reassembles the speaker. The novelty of this design is difficult to see since commercial loudspeakers with field-replaceable cone assemblies are readily available.—GLA

6,757,404

43.38.Ja LOUD SPEAKER, DIAPHRAGM AND PROCESS FOR MAKING THE DIAPHRAGM

Hiroyuki Takewa *et al.*, assignors to Matsushita Electric industrial Company, Limited
29 June 2004 (Class 381/430); filed in Japan 20 November 2000

This Matsushita patent is 42 pages long and lists 11 co-inventors. An



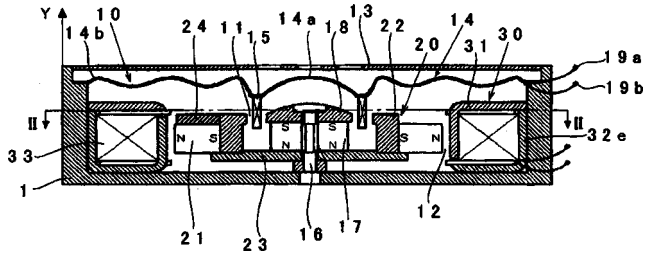
injection-molded diaphragm **50** includes a dome part **51**, a voice coil junction part **52**, and an integral surround **53**. ''Thus, the electromagnetic driving force of the voice coil can be effectively transmitted to the diaphragm so that a loudspeaker with excellent characteristics can be gained.''—GLA

6,766,034

43.38.Ja MULTIFUNCTION ACOUSTIC DEVICE

Takashi Kobayashi and Akira Nikaido, assignors to Citizen Electronics Company, Limited
20 July 2004 (Class 381/396); filed in Japan 21 September 2000

It is not unusual for a patent to be followed by one or more additional patents describing variations of or improvements to the original invention. In this case however, the speaker/vibrator covered by patent 6,711,269 [reviewed in J. Acoust. Soc. Am. **116**(3), 1320 (2004)] appears to be un-



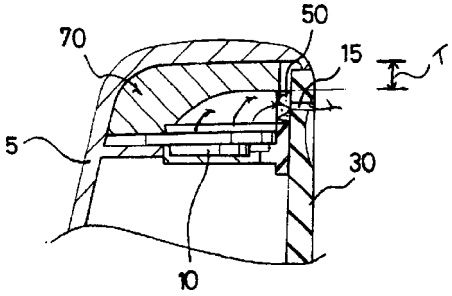
changed. The broad claims of the first patent are simply filled in with additional terms like ''side yolk,'' ''eccentric means,'' and ''having magnetic poles.'' In effect, the preferred embodiment has been patented separately.—GLA

6,763,110

43.38.Ja SOUND GUIDE FOR SPEAKER AND HANDSET FOR MOBILE COMMUNICATION USING THE SAME

Young Hwan Sung, assignor to LG Information & Communications, Limited
13 July 2004 (Class 379/433.02); filed in the Republic of Korea 18 December 1998

To reduce the size of a cellular phone handset without reducing its display area, sound from loudspeaker **10** is conducted to exit openings **15**



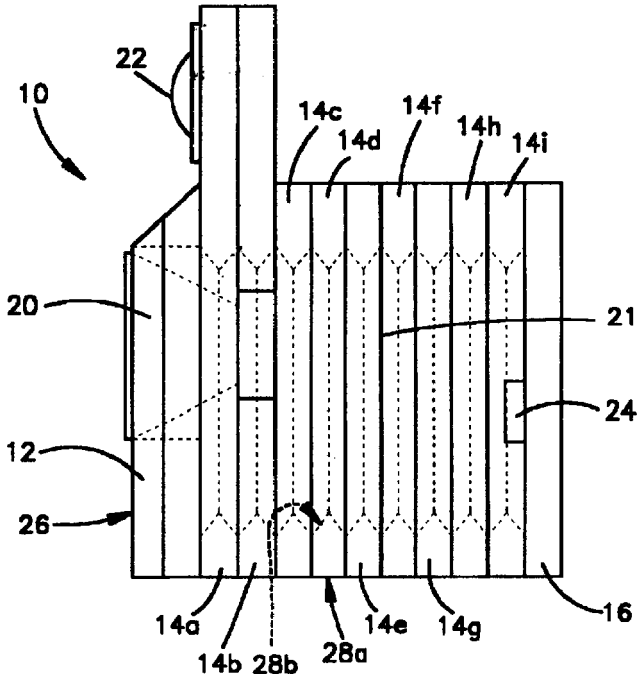
through a groove in sound guide **70**. According to the patent abstract, the speaker is mounted at a ''predetermined'' angle, and the angle at which sound is directed is also ''predetermined.'' However, the word ''prescribed'' is substituted in the 29 patent claims. To some readers, these verbal distinctions may be more interesting than the invention itself.—GLA

6,763,117

43.38.Ja SPEAKER ENCLOSURE

Barry Goldslager, Gardner, Massachusetts *et al.*
13 July 2004 (Class 381/345); filed 27 September 2001

A loudspeaker enclosure is made up of laminated sections, like a butcher block. Instead of being glued together, the sections are removable.



Thus, the user can modify the enclosure without removing the loudspeaker—a vexing problem that has heretofore been deemed unsolvable.—GLA

6,771,789

43.38.Kb ADJUSTABLE MICROPHONE APPARATUS

Hideaki Kakinuma, assignor to Vertex Standard Company, Limited
3 August 2004 (Class 381/361); filed in Japan 15 March 2000

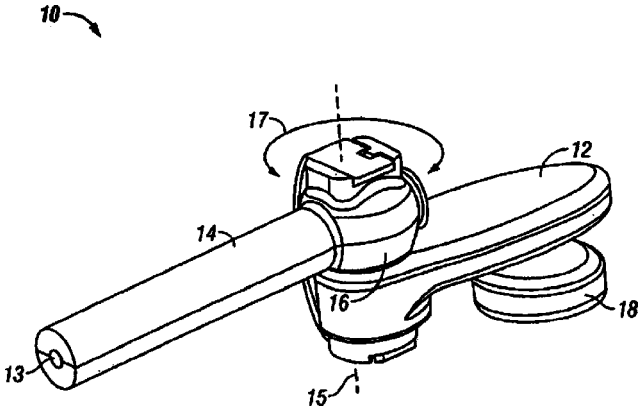
Miniature microphones that are built into various pocked-sized or hand-held communications and recording devices are subject to handling and wind noises in normal use. The patent describes several methods of adjusting an internal shield at the microphone element to isolate it from these effects.—JME

6,768,804

43.38.Kb ADJUSTABLE MICROPHONE BOOM WITH ACOUSTIC VALVE

Osman K. Isvan, assignor to Plantronics, Incorporated
27 July 2004 (Class 381/376); filed 15 November 2001

The patent discusses the design of communications headphones equipped with adjustable microphone tube boom sections in which the microphone is located in the headset itself. The hollow boom has a number of hinges and joints and is designed to afford optimized response at all usable tube lengths and adjustments thereof.—JME

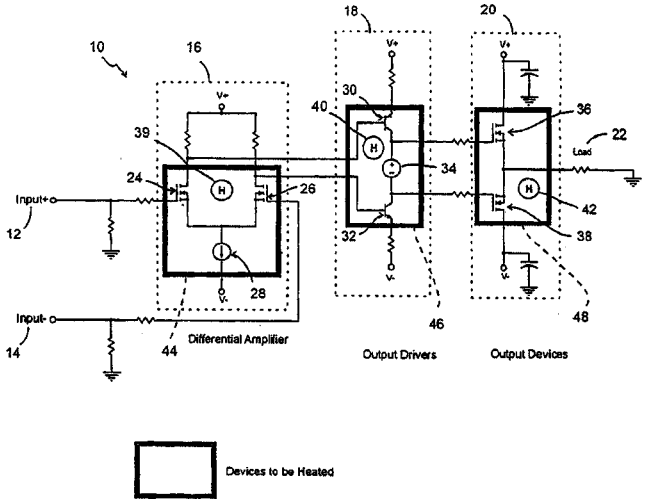


6,765,802

43.38.Lc AUDIO SOUND QUALITY ENHANCEMENT APPARATUS

Ray B. Ridley, assignor to Ridley Engineering, Incorporated
20 July 2004 (Class 361/767); filed 27 October 2000

The inventor wants to find the holy grail: transistor amplifiers that sound like tube/valve amplifiers. This popular pursuit (see the August 1998 *IEEE Spectrum* article for the pro-vacuum slant) has been covered extensively in the popular literature as well. The inventor proposes an unorthodox solution: he wants to intentionally heat up the transistor junctions in the amplifier stages (shown in bold boxes in the figure). The inventor argues that this sounds more tube like. But he doesn't present theory or measurement. What this really does is change the beta (current gain) of the transistors but the slopes don't change. So, how this helps is left to the imagination.—MK

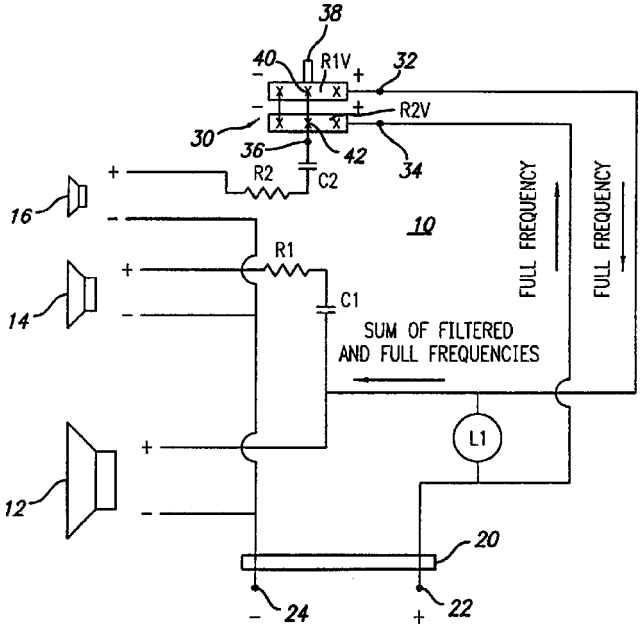


6,775,385

43.38.Lc LOUDSPEAKER FREQUENCY DISTRIBUTION AND ADJUSTING CIRCUIT

Jeffery James Coombs, assignor to James Loudspeaker, LLC
10 August 2004 (Class 381/99); filed 21 September 2000

Many loudspeaker system for the home, and even some for professional use, use passive low-, high-, and band-pass filters. These are commonly known as crossover networks. This patent describes a simple means of providing an adjustable crossover network using a series connected pair of potentiometers **R1V** and **R2V** that can "easily adjust the frequency out-



put and tonal qualities of a loudspeaker system to accommodate differing acoustical environments in which the loudspeaker system may be placed.'' The patent also discusses damping factor and sensitivity, both of which are said to be enhanced.—NAS

6,752,238

43.38.Md WATER RESISTANT AUDIBLE TOYS WITH SOUND EFFECTS

David Small and Paul S. Rago, assignors to Shoot the Moon Products 11, LLC
22 June 2004 (Class 181/149); filed 14 March 2002

Here's the problem: electric powered water guns would be more entertaining if they emitted sounds along with the water. Accordingly, the speaker assembly must be made watertight. But this creates other problems including possible pressure differentials due to shipping and water temperature variations. The inventors point out that absolute water proofing isn't needed. Indeed, they use a Mylar™ cone with an O-ring on the speaker surround.—MK

6,759,961

43.38.Md TWO-WAY COMMUNICATION BABY MONITOR WITH A SOOTHING UNIT

Karen Fitzgerald *et al.*, assignors to Mattel, Incorporated
6 July 2004 (Class 340/573.1); filed 4 October 2001

The now ubiquitous baby monitor enables parents to live somewhere other than in audio proximity to the child. If the concept is extended to a two-way radio, then the baby can now hear the parents (however distorted by poor transducers). In addition, this patent introduces parental control of a ''soothing unit'' that tries to reassure the child that sleep is really a fine thing to do. Schematics and mechanical drawings complete the picture.—MK

6,760,276

43.38.Md ACOUSTIC SIGNALING SYSTEM

Gerald S. Karr, Los Angeles, California
6 July 2004 (Class 367/197); filed 11 February 2000

This astounding patent argues that by generating tones, a simple thresholding receiver can be used to control devices at the end of the transmission channel (e.g., radio or telephone). First, this is essentially a primitive modem but without the knowledge of coding. Second, no mention is made of noisy channels of any variety. Third, the author maintains that these burst tones (sinusoids) are pleasing to the ear. Fourth, he's forgotten reverberation issues at the receiving end. He's probably never seen an acoustic coupler.—MK

6,760,696

43.38.Md FAST START VOICE RECORDING AND PLAYBACK ON A DIGITAL DEVICE

Jossef Goldberg *et al.*, assignors to Microsoft Corporation
6 July 2004 (Class 704/201); filed 1 April 1999

This patent concerns the excessive loading time for a process under Windows CE™. Suppose you want to make a recording but can't wait for the operating system to boot and/or the application to load. The proposed solution is not much more than buffering sound in memory while the process starts. Faster processors could make this issue obsolete (except for code bloat).—MK

6,782,105

43.38.Md REFLECTION SOUND GENERATOR WITH SERIES OF MAIN AND SUPPLEMENTARY FIR FILTERS

Shinichi Sahara and Yasushi Shimizu, assignors to Yamaha Corporation
24 August 2004 (Class 381/63); filed in Japan 25 November 1998

The fine art of creating natural sounding artificial reverberation at a price one can afford demands a careful balance between naturally iterative elements (such as IIR filters) and randomizing functions (such as dithered delay). The practiced ear abhors any hint of repetition in reverberant response and its ugly companion, comb filtering. What the ear seems to demand is a certain randomly based, spatially diverse, density of events, and any density beyond that critical point is more or less ignored by the ear. The patent describes several techniques in which simulated early reflections are subjected to variations in delay. Because of the subsequent relooping of the signal around certain elements, the effect of a given delay shift is actually spread out over a larger time interval. Subsequent delay paths may be shorter, but the combination can result in a much longer interval of random events. The patent is clearly written and sound throughout.—JME

6,754,358

43.38.Si METHOD AND APPARATUS FOR BONE SENSING

Peter V. Boesen, Des Moines, Iowa *et al.*
22 June 2004 (Class 381/326); filed 10 July 2001

During the past few years there has been renewed interest in bone conduction audio pickup for two-way communications systems. This patent asserts that existing bone conduction sensors have a number of limitations resulting from their general size and shape. A nonocclusive design is described that is small in size yet provides increased contact area. The invention includes means for detecting when the sensor has been displaced and for compensating for such displacement.—GLA

6,763,119

43.38.Si NECKPHONE

Chun-Bong Lee, assignor to Neckphone Company, Limited
13 July 2004 (Class 381/385); filed in the Republic of Korea 27 May 2000

Small loudspeakers and electronic circuitry can be housed in a flexible, fabric-covered housing that is draped around the user's neck like a locker room towel. The proximity of the speakers to the listener's ears can produce surprisingly good stereo reproduction. The arrangement described has been available for more than 20 years, but in this patent it is put to a different use. For 360° sound immersion, frontal sound sources are reproduced through conventional loudspeakers but rear sources are reproduced through neckphone speakers. The patent asserts that this hybrid arrangement can produce convincing surround sound in a restricted listening area, making it well suited to computer gaming.—GLA

6,781,933

43.38.Vk PSEUDO MULTI-CHANNEL STEREO PLAY-BACK

Satomi Shigaki, assignor to NEC Corporation
24 August 2004 (Class 369/47.1); filed in Japan 15 December 1997

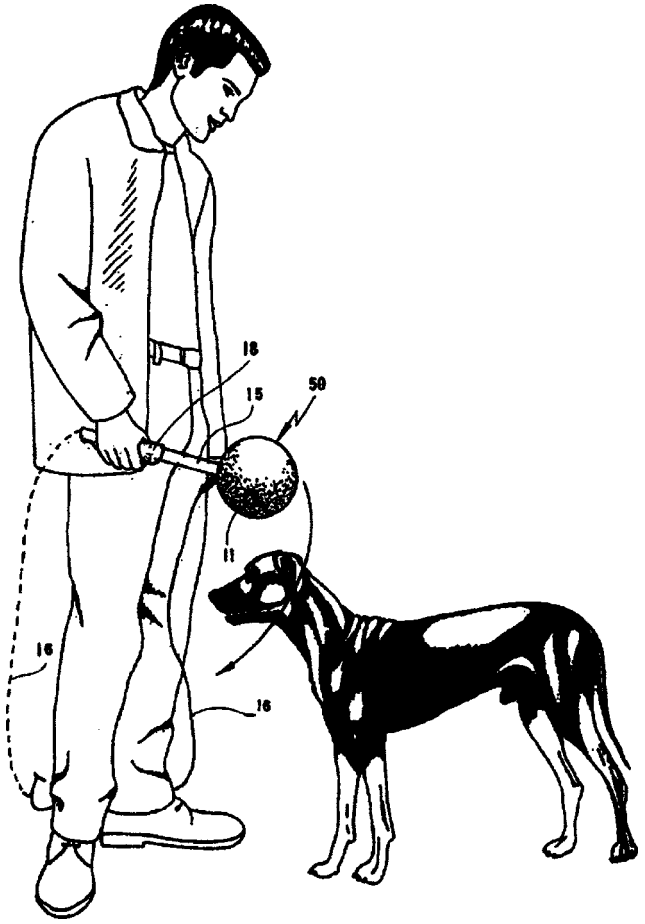
The patent describes a method of playing standard 5.1 surround program material over two loudspeakers, such as you would find in the normal laptop computer or in a conventional television set. There are by now scores of such schemes for doing this, ranging all the way from utter simplicity to very complex, finely tuned, HRTF-intensive realizations. This one appears to be more or less mid-way on the continuum.—JME

6,782,104

43.38.Vk METHOD AND DEVICE INTENDED FOR THE PICKING UP OF SOUNDS, FOR THEIR RECORDING AND THEIR PLAY-BACK, AND REPRODUCING THE NATURAL SENSATION OF A SOUND SPACE

Georges Claude Vieilledent, assignor to ARKAMYS
24 August 2004 (Class 381/26); filed in France 31 December 1998

This fanciful patent describes stereo-on-a-stick, as it were, in which a wand with a baffled pair of microphones at the end can be moved and manipulated, apparently for ''reproducing the natural sensation of a sound space.'' What is clearly being sought here is a clear distinction with conventional stereo, ''which artificially reconstitutes for the listener, by unconscious analogy with stored memories, the atmosphere of a concert hall to



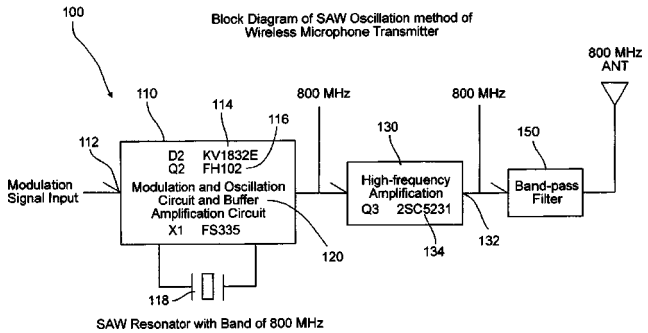
which he attributes depth and breadth.'' There is certainly more than a bit of truth here.—JME

6,778,814

43.38.Wl WIRELESS MICROPHONE APPARATUS AND TRANSMITTER DEVICE FOR A WIRELESS MICROPHONE

Yukinaga Koike, assignor to Circuit Design, Incorporated
17 August 2004 (Class 455/95); filed in Japan 28 December 1999

The patent describes a miniature wireless microphone system, complete with complementary pre- and de-emphasis and signal companding, whose transmitter is no larger than a conventional Canon XLR microphone connector. The system is intended for business purposes and as such will be operated alone and, apparently, not in an environment of multiple adjacent



wireless channels. Its rf power output is about 10 mW, operating frequency is stated as ''about 800 MHz,'' and battery life is stated as 10 h. The oscillating element is a surface acoustic wave oscillator. If the system is carefully shielded—in both directions—it should enjoy great success in many business functions.—JME

6,763,310

43.40.At MODAL ANALYSIS METHOD AND APPARATUS THEREFOR

François Lafleur *et al.*, assignors to Centre de Recherche Industrielle du Québec
13 July 2004 (Class 702/33); filed 13 May 2002

Environmental stress screening of printed circuit boards for quality control in manufacturing is accomplished by subjecting each circuit board to a predetermined sound field and determining the resonant frequencies and damping factors associated with its main vibration modes. The circuit board is mounted on a fixture which supports a number of microphones and a single accelerometer is affixed to the circuit board adjacent to one of the microphones. The outputs of all microphones and of the accelerometer are subjected to Fourier analysis and used to determine frequency-response functions, from which the modal characteristics are computed on the basis of a multiple-input single-output model.—EEU

6,752,899

43.40.Le ACOUSTIC MICROBALANCE FOR IN-SITU DEPOSITION PROCESS MONITORING AND CONTROL

Bhanwar Singh *et al.*, assignors to Advanced Micro Devices, Incorporated
22 June 2004 (Class 156/345.24); filed 16 January 2002

Metal film deposition is an important step in the manufacture of semiconductors. If the gradually increasing thickness of the metal could be accurately measured *in situ*, then the process could be terminated at exactly the right time. The patent explains that experimental work with a resonant quartz microbalance has not proved successful. Instead, an acoustic microbalance is proposed that can be placed in a deposition chamber. The microbalance can be any device or material in which the frequency or propagation speed of an acoustic wave varies with the thickness of a depositing film. ''Data from the acoustic microbalance is employed to detect a process endpoint, determine an adjustment to processing conditions..., and/or provide feedback control over current processing conditions.''—GLA

6,752,039

43.40.Tm SHOCK-ABSORBED VEHICLE STEERING WHEEL

Martin Kreuzer *et al.*, assignors to TRW Automotive Safety Systems GmbH & Company KG
22 June 2004 (Class 74/552); filed in Germany 8 March 2001

The rim of a steering wheel, according to this patent, is connected to its shaft via a resilient arrangement, whose stiffness can be changed in response to a control signal derived on the basis of a steering-column vibration signal. The resilient arrangement may include a magnetorheological or electrotheological fluid.—EEU

6,752,250

43.40.Tm SHOCK, VIBRATION AND ACOUSTIC ISOLATION SYSTEM

Edward T. Tanner, assignor to Northrop Grumman Corporation
22 June 2004 (Class 188/267); filed 27 September 2001

The isolation system of this patent employs an elastic element for vibration isolation in mechanical parallel with an actively controlled damper for shock attenuation. The damper force is optimized via a controller to which are input signals from a payload acceleration sensor and/or from measurement of the displacement of the payload relative to the base.—EEU

6,752,425

43.40.Tm SEMI-ACTIVE CONTROL OF AUTOMOTIVE STEERING SYSTEM VIBRATION WITH MAGNETO-RHEOLOGICAL DAMPING

Wei-Yi Loh *et al.*, assignors to Ford Global Technologies, LLC
22 June 2004 (Class 280/779); filed 30 October 2002

A damper, according to this patent, in essence consists of a disc that is affixed to the steering wheel shaft and located in a housing filled with a magnetorheological fluid. The disc's circumference is somewhat like that of a gear and the disc contains orifices, so that the fluid can act on the disc's rotational and actual motions. Electrical coils can activate or deactivate the magnetorheological fluid in response to a controller signal, derived by comparing a steering wheel vibration signal to a predetermined threshold value.—EEU

6,763,921

43.40.Tm REDUCED-VIBRATION TUBE ARRAY

Gerald J. Bruck and Daniel R. Bartolomeo, assignors to Siemens Westinghouse Power Corporation
20 July 2004 (Class 188/378); filed 13 September 2002

In tube assemblies, such as heat exchangers, where external bracing of tubes may interfere with flow across the tube exteriors, flow-induced vibrations of tubes are reduced by dampers located inside the tubes. Dampers according to this patent consist of strips that may be helical, for example, and that are situated in a tube essentially so as to not make contact with the interior of the tube when it does not vibrate. At some locations along the tube, the clearance between certain points on the strips and the tube interior is very small so that contact is made when the tube vibrates, resulting in damping of the tube vibrations.—EEU

6,755,094

43.40.Tm ANTI-NOISE GEAR

Winfried Rehle *et al.*, assignors to Agco GmbH & Company
29 June 2004 (Class 74/443); filed in the United Kingdom 15 August 2000

A disc of damping material is fastened coaxially to the face of a gear. This disc is of a material with a Young's modulus comparable to that of the gear's material. Mechanical fastening means, such as press fitting, retaining rings, or screws are employed, so that no adhesive is necessary.—EEU

6,763,794

43.40.Tm VIBRATION SOUND REDUCING DEVICE, AND PROCESS FOR ASSEMBLING ELASTIC MEMBRANE IN VIBRATION SOUND REDUCING DEVICE

Terukazu Torikai *et al.*, assignors to Honda Giken Kogyo Kabushiki Kaisha
20 July 2004 (Class 123/192.1); filed in Japan 14 November 1997

In a system where vibration and noise are produced by pressure pulses in a liquid, a flexible membrane is provided to release the pressure in the liquid, thereby reducing the exciting forces. The concept and devices described in this patent are applicable to automotive engines, where piston slap induces pressure pulses in the cooling water, for example.—EEU

6,763,917

43.40.Vn ELEVATOR VIBRATION REDUCTION APPARATUS INCLUDING A DEAD BAND FILTER

Kenji Utsunomiya *et al.*, assignors to Mitsubishi Denki Kabushiki Kaisha
20 July 2004 (Class 187/292); filed 10 October 2002

The vertical and/or horizontal vibrations of an elevator cab are sensed by accelerometers and reduced by means of actuators that are made to act on the basis of signals generated by a controller. The actuators are located between the cab and the elevator rails, between the cab and guide rollers that ride on the rails, or between the cab and a frame that is attached to the guide rollers. A dead-band filter is used in the control circuit, with a dead band that corresponds to the noise floor of the accelerometers. The control system is reset when the cab stops, as determined from a stop-signal generator.—EEU

6,757,620

43.40.Yq METHOD FOR EXAMINING STRUCTURES HAVING HIGH NATURAL VIBRATION FREQUENCY USING ALTERNATING MANUAL VIBRATION-EXCITING METHOD

Sung Won Yoon and Sang Keun Oh, assignors to Seoul National University of Technology
29 June 2004 (Class 702/33); filed 5 December 2002

A building can be made to vibrate measurably at its horizontal natural frequency by people atop it pushing in a synchronized manner. However, people cannot push at frequencies greater than about 2.5 Hz. In order to excite a building with a higher natural frequency, two groups of people push at half the natural frequency, alternating with each other, in unison with sounds generated by a two-beat metronome or the like. Damping is determined by the log decrement method, using measured signals obtained after the pushing is stopped.—EEU

6,776,258

43.55.Ev ACOUSTIC BLANKET SYSTEM

Duane D. Grosskrueger and Brett E. Pisor, assignors to Lockheed Martin Corporation
17 August 2004 (Class 181/294); filed 28 December 2001

This acoustic blanket system comprises a Polyimide foam interior panel encased by carbon Teflon impregnated glass fiber cover materials. The cover materials are heat-sealed at their perimeter to form a lightweight blanket that can accept a variety of fastening systems. The blanket is to be used on space vehicles to protect payloads from the acoustic energy of the rocket engine.—CJR

6,774,934

43.60.Jn SIGNAL LOCALIZATION ARRANGEMENT

Harm J. W. Belt and Cornelis P. Janse, assignors to Koninklijke Philips Electronics N.V.
10 August 2004 (Class 348/211.1); filed in the European Patent Office 11 November 1998

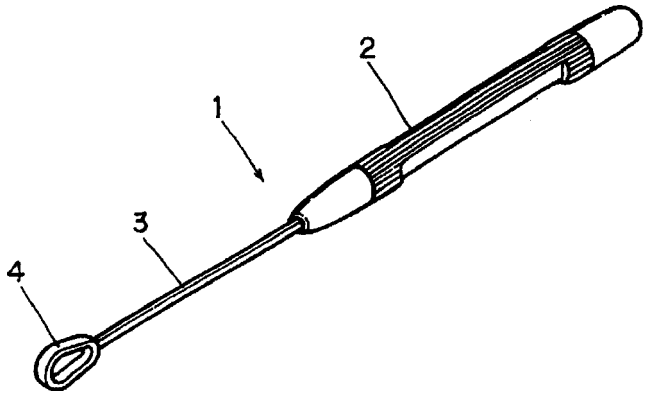
While the problems of hands-free identification and gating of audio signals were solved more than 30 years ago, it seems that similar problems in video identification of who's actually talking have not been so easily solved. The patent deals with conference systems and the desire to extend hands-free operation to include video camera management. Two or more microphones are used to sense the position of the talker, and the relative delays at the microphones are used to ''zero in'' on the position of the talker and determine the required position and aiming of the camera accordingly. The patent further discusses refinements in the technique to ensure quick and positive identification of the talker's position.—JME

6,776,786

43.64.Yp EARPICK

Yongup Kim, assignor to Soo-Hee Kim
17 August 2004 (Class 606/162); filed in the Republic of Korea 26 February 2002

An earpick is used to clean out earwax from the auditory canal. A conventional earpick has a scoop-shaped portion that substantially scrapes out the earwax. In this proposed design, the earpick includes a handle, a support bar movably coupled to the handle, and an earwax-cleaning segment coupled to the support bar. When the earpick enters and contacts the inner



wall of the auditory canal, the support bar moves within the handle. The earwax-cleaning segment has a fairly ring-shaped cross-section. The patent maintains that the new design lessens the possibility of damaging the auditory canal and eardrum.—DRR

6,768,798

43.66.Pn METHOD OF CUSTOMIZING HRTF TO IMPROVE THE AUDIO EXPERIENCE THROUGH A SERIES OF TEST SOUNDS

Morgan James Dempsey, assignor to Koninklijke Philips Electronics N.V.
27 July 2004 (Class 381/17); filed 19 November 1997

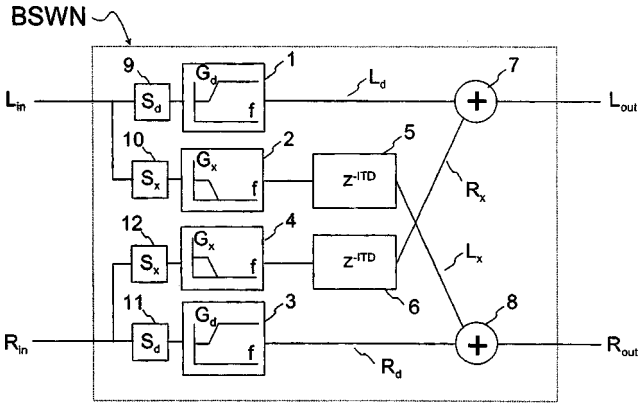
HRTFs derived from artificial head measurements may represent a ''best fit'' for the population at large but lack the precision afforded by customized measurements on individual subjects. This intelligently written patent describes a method of making a limited set of localization measurements on a subject which are then used to effectively modify the generic HRTF set to achieve a customized set for that subject.—JME

6,771,778

43.66.Pn METHOD AND SIGNAL PROCESSING DEVICE FOR CONVERTING STEREO SIGNALS FOR HEADPHONE LISTENING

Ole Kirkeby, assignor to Nokia Mobile Phones Limited
3 August 2004 (Class 381/17); filed in Finland 29 September 2000

Most methods that have been used over the years for converting stereo signals for binaural headphone listening have relied on a fairly simple model of two-way crosstalk that, to a first approximation, models what actually



happens in nature. This patent proposes a scheme in which several signal-directed gain and delay variations are introduced into the crosstalk paths, apparently to make the effect more ''real'' than nature itself.—JME

6,763,116

43.66.Ts HEARING AID AND OPERATING METHOD THEREFOR WITH CONTROL DEPENDENT ON THE NOISE CONTENT OF THE INCOMING AUDIO SIGNAL

Roland Barthel and Torsten Niederdränk, assignors to Siemens Audiologische Technik GmbH
13 July 2004 (Class 381/312); filed in Germany 24 September 2001

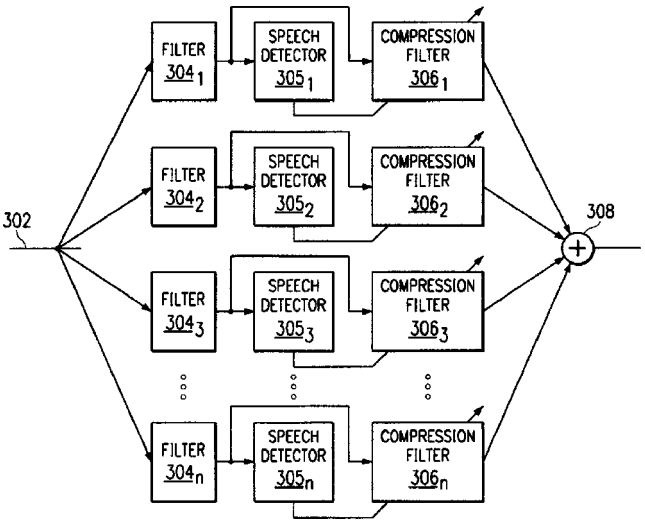
Digital mobile phones and other electrical and magnetic disturbances are sources of interference for hearing aids. The acoustic input signal to the hearing aid is analyzed for noise content and for the purpose of identifying the source of the disturbance. Based on this knowledge, the hearing aid can be automatically switched into a mode with parameter settings that minimize the detrimental effect of the noise.—DAP

6,754,355

43.66.Ts DIGITAL HEARING DEVICE, METHOD AND SYSTEM

Trudy D. Stetzler *et al.*, assignors to Texas Instruments Incorporated
22 June 2004 (Class 381/94.2); filed 7 December 2000

A digital hearing device employs speech detectors in several channels to identify the acoustic environment surrounding the wearer. The outputs of the speech detectors are used to modify the coefficients of compression filters which selectively govern the amount of attenuation or amplification provided in given frequency regions depending on whether speech or noise is present. The speech detectors may store parameters associated with certain acoustic environments, enabling the system to automatically reconfigure



the compression settings when these environments are encountered again.—DAP

6,768,801

43.66.Ts HEARING AID HAVING IMPROVED SPEECH INTELLIGIBILITY DUE TO FREQUENCY-SELECTIVE SIGNAL PROCESSING, AND METHOD FOR OPERATING SAME

Frank Wagner and Fred Zoels, assignors to Siemens Aktiengesellschaft
27 July 2004 (Class 381/312); filed in Germany 24 July 1998

A multiband filter separates the input signal into several channels. In each channel, a decision is made about whether the signals present are speech or noise. The presence of speech is sensed by detecting whether a sudden level increase has resulted in a brief peak. The type of speech signal, for example, vowels or consonants, is inferred by calculating derivatives, maxima, and minima. If speech is thought to be present in a particular channel, the signal in that channel may be boosted for the duration of the speech signal.—DAP

6,768,802

43.66.Ts BINAURAL SYNCHRONIZATION

Herbert Baechler, assignor to Phonak AG
27 July 2004 (Class 381/315); filed in the World IPO 15 October 1999

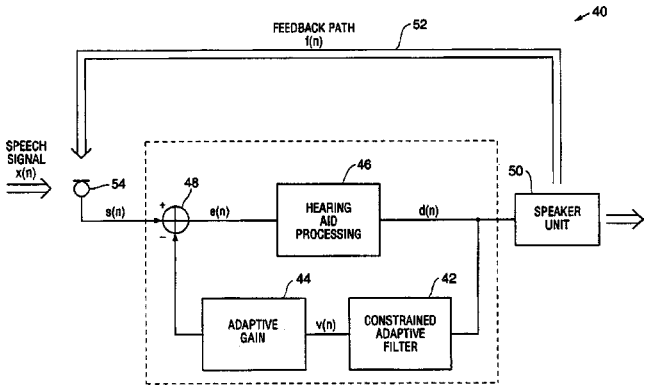
A methodology is described to synchronize the settings of the two hearing aids in a binaural fitting via wireless transmission without using a remote control. Synchronization may be initiated manually with a switch or automatically based on acoustic environment analysis. The two hearing aids can be configured to communicate as master and slave or both hearing aids can operate in the same active transmission mode.—DAP

6,754,356

43.66.Ts TWO-STAGE ADAPTIVE FEEDBACK CANCELLATION SCHEME FOR HEARING INSTRUMENTS

Fa-Long Luo *et al.*, assignors to GN ReSound AS
22 June 2004 (Class 381/318); filed 6 October 2000

Some adaptive feedback rejection algorithms used in current hearing aids are set up by training during the fitting session. These systems that also utilize constrained adaptive filtering for preventing problems with narrow-



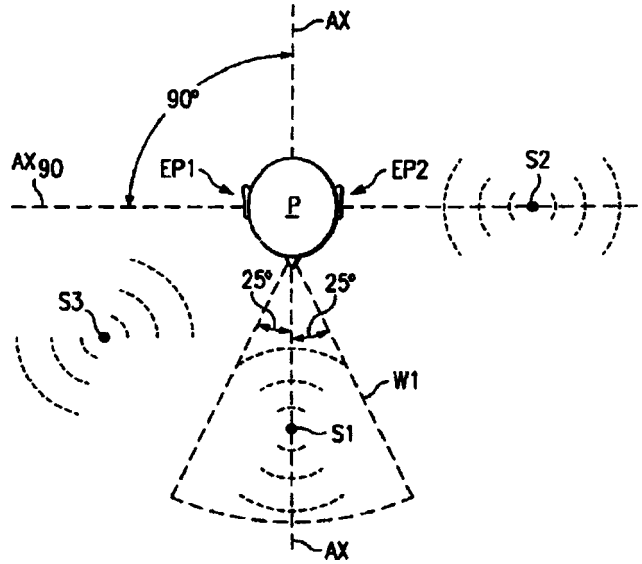
band inputs may have the disadvantage that they can't provide cancellation in the presence of large variations of the acoustic feedback path that are produced, for example, by bringing a telephone handset close to the ear. Disclosed is a two-stage feedback canceller consisting of the constrained adaptive filtering as well as an adaptive gain modification for normalizing gain.—DAP

6,778,674

43.66.Ts HEARING ASSIST DEVICE WITH DIRECTIONAL DETECTION AND SOUND MODIFICATION

Carl M. Panasik *et al.*, assignors to Texas Instruments Incorporated
17 August 2004 (Class 381/313); filed 28 December 1999

A methodology is described for suppressing undesired sounds and improving the localization ability of hearing assistive device wearers. Two spacially separated microphones each provide pickup of at least two sound sources, for example, one within and the others outside a specified desired



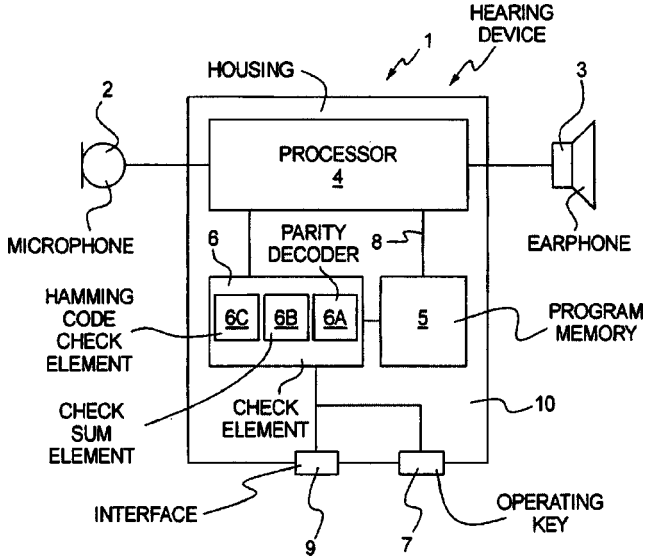
sound pickup area relative to the wearer. The microphone signals may be sent to a remote processor via wireless means. After processing to suppress the undesired signals, the resultant is provided via wireless means to a speaker in or near the wearer's ear canals.—DAP

6,782,110

43.66.Ts METHOD AND DIGITAL HEARING DEVICE FOR DETECTING AND/OR REMOVING ERRORS ARISING IN THE TRANSMISSION AND STORAGE OF DATA

Ullrich Sigwanz and Fred Zoels, assignors to Siemens Audiologische Technik GmbH
24 August 2004 (Class 381/312); filed in Germany 11 August 1997

A technique is described for performing error detection and correction for data transmissions within and to hearing aids. For example, during hearing aid programming, errors that may occur while downloading data from a PC into permanent memory in the hearing aid could be detected, triggering



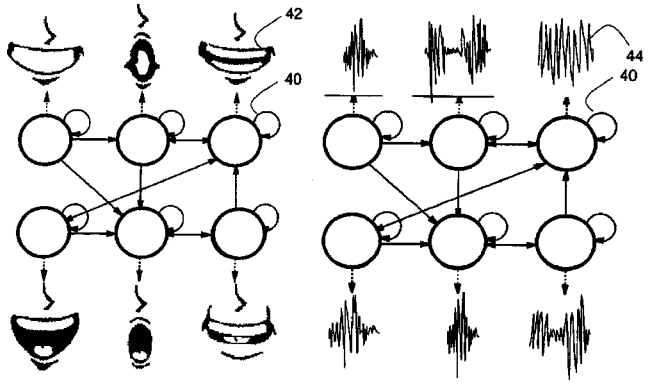
a program restart or shut down, or data retransmission. Error checking using checksums and error correction within the hearing aid saves time by eliminating time consuming data readback to the pc.—DAP

6,735,566

43.70.Aj GENERATING REALISTIC FACIAL ANIMATION FROM SPEECH

Matthew E. Brand, assignor to Mitsubishi Electric Research Laboratories, Incorporated
11 May 2004 (Class 704/256); filed 9 October 1998

This is a scheme for improving the quality of the facial displays created to accompany a previously recorded sound track. Whereas previous talking face generators have used an intermediate phonemic or visemic rep-



resentation, the plan here is to directly compute the relevant facial changes based on a hidden Markov model trained with visual features of the head

and face movements. The argument is that the relevant dynamics of facial movement are thus captured and more directly related to the audio patterns. The result is said to be applicable to either realistic image synthesis or cartoon images.—DLR

6,729,882

43.70.Jt PHONETIC INSTRUCTIONAL DATABASE COMPUTER DEVICE FOR TEACHING THE SOUND PATTERNS OF ENGLISH

Thomas F. Noble, Newbury Park, California
4 May 2004 (Class 434/167); filed 9 August 2001

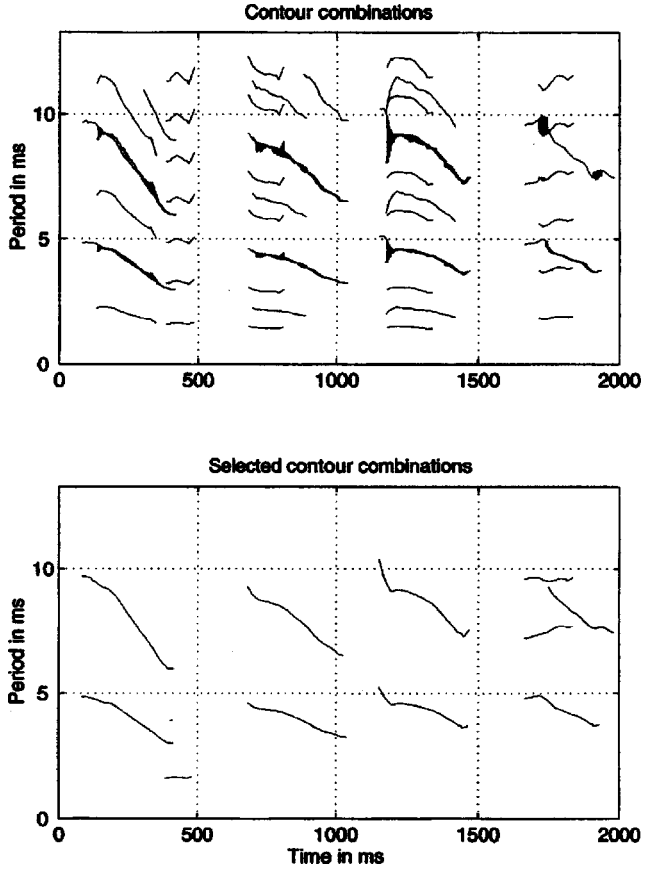
Based primarily on a syllabified English lexical database, this computer system provides a comprehensive tool for learning and teaching the structure and pronunciation of English words. Using typical database tools, a student or teacher can quickly produce pattern lists, structural charts, and audio output from selected word lists or syllable patterns. A creative user might see how these outputs could be useful in learning more about the English language.—DLR

6,745,155

43.72.Ar METHODS AND APPARATUSES FOR SIGNAL ANALYSIS

Tjeerd Catharinus Andringa *et al.*, assignors to Huq Speech Technologies B.V.
1 June 2004 (Class 702/189); filed in the Netherlands 5 November 1999

The patent describes, in considerable, but readable, detail, a set of procedures for analyzing the spectral content of a periodic or quasiperiodic



signal using a model of the basilar membrame and associated innervation as the underlying methodology. The basilar model itself consists of 100 filters ranging from 30 to 6100 Hz. The processing makes extensive use of correlations. With emphasis on the detection of periodicities, the method is reportedly able to detect formant structure as well as pitch periods. The figure shows the periodicity structure as computed for the Dutch phrase, ''nul, een, twee, drie.''—DLR

6,732,142

43.72.Ja METHOD AND APPARATUS FOR AUDIBLE PRESENTATION OF WEB PAGE CONTENT

Cary Lee Bates *et al.*, assignors to International Business Machines Corporation
4 May 2004 (Class 709/203); filed 25 January 2000

The patent covers an application for speech systhesis in which a selected portion of the computer screen is announced verbally while the user is working elsewhere, or perhaps concentrating on another portion of the screen. The announcements may be selectably triggered at fixed time intervals, upon content changes, or by other criteria. Comparing this to working while listening to music or other background, the patent argues that the scheme can improve concentration on the main task while keeping the user informed of a side issue.—DLR

6,738,457

43.72.Ja VOICE PROCESSING SYSTEM

John Brian Pickering and Graham Hugh Tuttle, assignors to International Business Machines Corporation
18 May 2004 (Class 379/88.16); filed in the United Kingdom 27 October 1999

This voice processor, intended primarily for use in a telephone voice mail system, would extract analysis parameters from one recorded voice and use those parameters to resynthsize a second recording using voice characteristics from the speaker of the first recording. The method is described using examples from linear prediction analysis, although other analysis methods would also be covered by the patent.—DLR

6,732,074

43.72.Ne DEVICE FOR SPEECH RECOGNITION WITH DICTIONARY UPDATING

Masaru Kuroda, assignor to Ricoh Company, Limited
4 May 2004 (Class 704/244); filed in Japan 28 January 1999

This system for training a speaker-dependent speech recognizer compares each pronunication of a word with the pattern stored in a speaker-independent reference dictionary. A new speaker-dependent word model is then stored. If the new word pattern is close to the existing pattern, the speaker need not pronounce additional examples of that word. The novel feature here is that a second comparison checks only the vowels. The argument is made that if just the vowels are updated, a satisfactory pattern can be produced with fewer spoken samples, easing the user's training burden.—DLR

6,738,741

43.72.Ne SEGMENTATION TECHNIQUE INCREASING THE ACTIVE VOCABULARY OF SPEECH RECOGNIZERS

Ossama Emam and Siegfried Kunzmann, assignors to International Business Machines Corporation
18 May 2004 (Class 704/251); filed in the European Patent Office 28 August 1998

The system described here essentially uses prefix and suffix analyses to extend the vocabulary of a speech recognizer. The description uses ex-

amples from German speech, which has a large inventory of affixed words. It is not clear what distinguishes this approach from the numerous prior applications of affix structure for recognition vocabularies.—DLR

6,741,963

43.72.Ne METHOD OF MANAGING A SPEECH CACHE

Daniel E. Badt *et al.*, assignors to International Business Machines Corporation
25 May 2004 (Class 704/270); filed 21 June 2000

It seems to this reviewer that the patented system is an elaborate and cumbersome addition to a speech recognizer intended to overcome an elaborate and cumbersome mode of operation. A cache of recognized phrases would retain information about the recognized items, in this case, destinations to be used in an automobile navigation system. Since the original recognizer design requires a detailed set of attributes to be laboriously added to each recognized item, a cache would allow those details to be retained, once entered. Most of the patent deals with cache mechanics, importance of entries and their frequency of use, and arrangements for purging an overflowing cache. All of this is managed by more speech input.—DLR

6,754,626

43.72.Ne CREATING A HIERARCHICAL TREE OF LANGUAGE MODELS FOR A DIALOG SYSTEM BASED ON PROMPT AND DIALOG CONTEXT

Mark E. Epstein, assignor to International Business Machines Corporation
22 June 2004 (Class 704/235); filed 1 March 2001

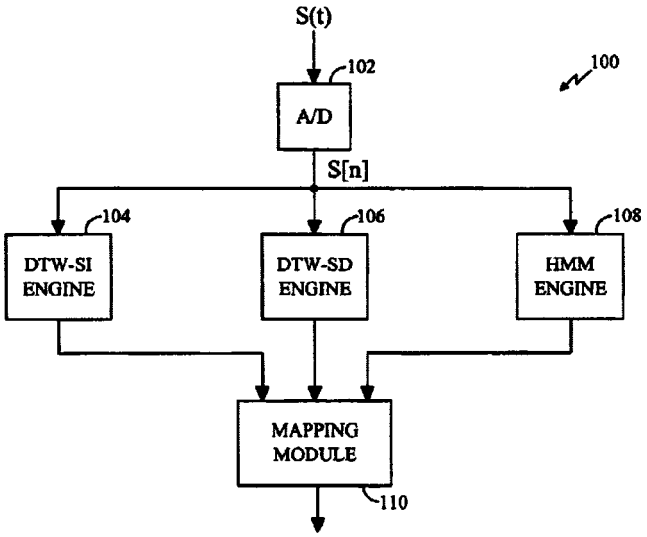
In speech-to-text systems, new data reflecting a user's changed dictation style may not significantly alter the behavior of the speech recognition system if a large amount of data already exists in a language model. Disclosed is a method of creating a hierarchy of contextual models that are organized into a treelike structure for use in converting speech to text.—DAP

6,754,629

43.72.Ne SYSTEM AND METHOD FOR AUTOMATIC VOICE RECOGNITION USING MAPPING

Yingyong Qi *et al.*, assignors to Qualcomm Incorporated
22 June 2004 (Class 704/246); filed 8 September 2000

A method is described for combining the results of different speech recognition engines and resolving differences between them. Types of engines combined include both speaker-independent and speaker-dependent dynamic time warping and hidden Markov model systems.—DAP

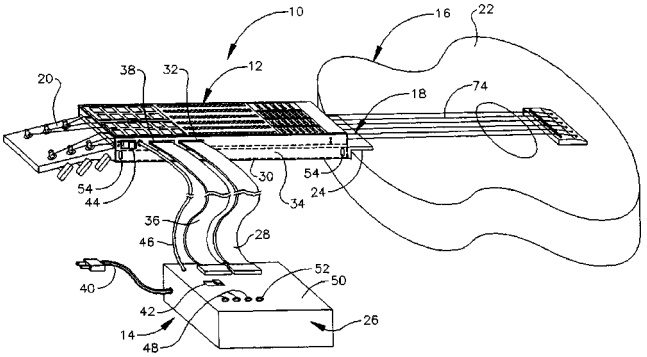


6,753,466

43.75.Gh ELECTRONIC PROGRAMMABLE SYSTEM FOR PLAYING STRINGED INSTRUMENTS AND METHOD OF USING SAME

Day Sun Lee, Long Beach, California
22 June 2004 (Class 84/669); filed 16 October 2000

Essentially, the inventor wants to turn fretted multistringed instruments into autoharps. The patent lays out how to construct a multisolenoid actuator **12** that is placed over the fretboard. The controller **26** has a chordal interface. The string activation (plucking, strumming) is left to the user (just like an autoharp).—MK



6,753,467

43.75.Hi SIMPLE ELECTRONIC MUSICAL INSTRUMENT, PLAYER'S CONSOLE AND SIGNAL PROCESSING SYSTEM INCORPORATED THEREIN

So Tanaka and Minoru Harada, assignors to Yamaha Corporation
22 June 2004 (Class 84/723); filed in Japan 27 September 2001

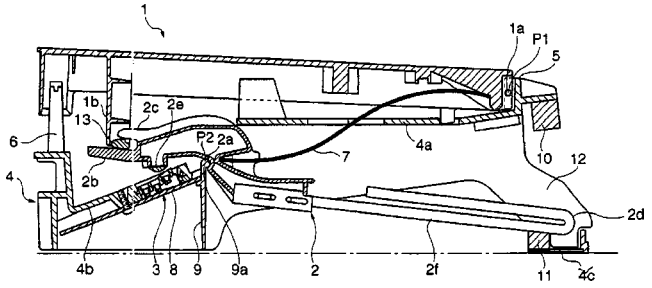
The title is a clear example of obfuscation: This is an electronic drum. Simple electronic drums use switches or thresholding to generate drum events. The patent describes several different head designs (including the use of switches for rimshots) and different front end processing circuits. Electronic drum makers will find the figures interesting but the text awkward and turgid.—MK

6,765,142

43.75.Mn ELECTRONIC KEYBOARD MUSICAL INSTRUMENT

Shinya Sakurada *et al.*, assignors to Yamaha Corporation
20 July 2004 (Class 84/719); filed in Japan 15 January 2002

Once again, the issue is the effective re-creation of the acoustic piano action in an electronic or synthetic piano (or synthesizer). As shown, a key **1** is depressed, eventually causing the switch depression **2e** to act against the



switches **3**. Note the mass for the hammer **2f** and the restoration spring **7**. The switches are an assembly of three separate switches and an algorithm for using their output is given in the patent figures.—MK

6,756,534

43.75.St MUSIC PUZZLE PLATFORM

Alexander Gimpelson *et al.*, assignors to Quaint Interactive, Incorporated
29 June 2004 (Class 84/610); filed 8 August 2002

The inventors propose a musical jigjaw puzzle: tracks or instruments are presented to the user for final assembly into a whole musical piece. To add complexity, slightly altered tracks (''distractions'') are also included.—MK

6,754,351

43.75.Wx MUSIC APPARATUS WITH DYNAMIC CHANGE OF EFFECTS

Ryoji Tanji, assignor to Yamaha Corporation
22 June 2004 (Class 381/61); filed in Japan 22 May 1997

It is well known that a DSP microprocessor can run out of processing cycles if the demands are too high. In this patent, filed in 1997, Yamaha claims to have invented a way to turn off the various signal processing sections in the firmware by manual control. Given that DSPs had been around for 30+ years at the time of filing, their claim of originality is indeed dubious.—MK

6,757,573

43.75.Wx METHOD AND SYSTEM FOR AUTHORING A SOUNDSCAPE FOR A MEDIA APPLICATION

Eric Ledoux *et al.*, assignors to Microsoft Corporation
29 June 2004 (Class 700/94); filed 2 November 1999

How can a naive user create a changing sound environment? The Microsoft team proposes a Windows™ based tool that uses a graph with triggers that control graph traversal, much like a Petri net. Unfortunately, the inventors seem ignorant of the existing graphical sound score editors like Max (Puckette, 1991). Additionally, MPEG-4 is given short shrift. Windows hackers will appreciate the Visual Basic script in Appendix A.—MK

6,762,357

43.75.Wx RESONANCE APPARATUS, RESONANCE METHOD AND COMPUTER PROGRAM FOR RESONANCE PROCESSING

Yutaka Washiyama and Seiji Okamoto, assignors to Kawai Musical Instruments Manufacturing Company, Limited
13 July 2004 (Class 84/603); filed in Japan 17 April 2001

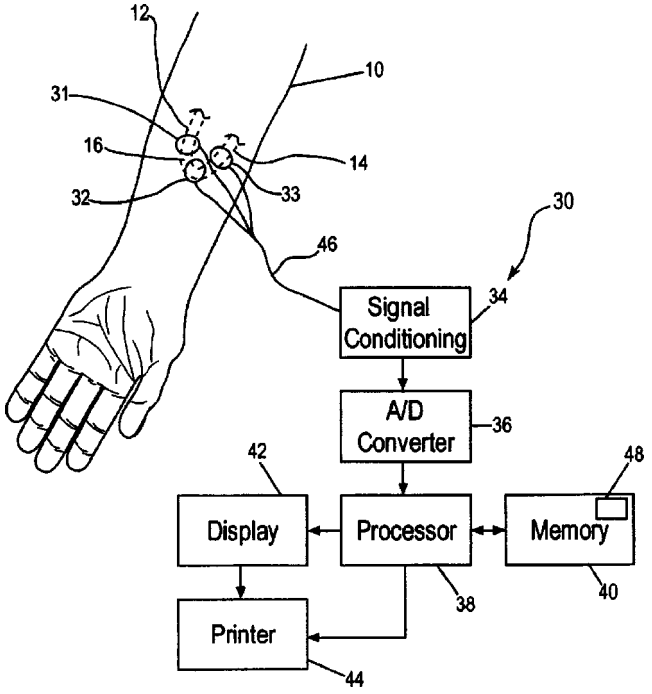
In 1983, Karplus and Strong published a simple string synthesis algorithm that uses a recirculating delay line with filtering. This was also patented (Unites States Patent 4,649,783) in 1987. Now, Kawai introduces a tiny wrinkle, the use of an FIR filter. But this was more than ably analyzed and discussed by Jaffe and Smith in 1983. Finally, the lack of DSP analysis of the algorithm is disappointing in the extreme.—MK

6,780,159

43.80.Jz ACOUSTIC DETECTION OF VASCULAR CONDITIONS

Richard H. Sandler and Hussein A. Mansy, assignors to Biomedical Acoustic Research Corporation
24 August 2004 (Class 600/504); filed 15 January 2002

Acoustic detection techniques are used in this apparatus to measure vibrations or sounds generated by blood flowing through AV (arterial-venous) access shunts, arteries, and/or veins and to process these measured vibrations or sound in order to diagnose the internal condition of the shunts, arteries and/or veins. The system generates spectral information from the vibration information, calculates a spectral parameter based on the spectral



information and detects vascular condition based on the spectral parameter. The method may be used to locate a vascular blockage and to assess the degree of blockage. The acoustic detection techniques may be applied to

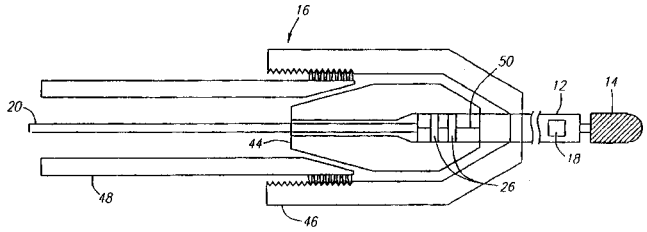
detect blockages, to detect stenosis within renal arteries, provide for early detection of abdominal aortic aneurysms, facilitate salvage of femoral-poplitical bypass graft, etc.—DRR

6,770,035

43.80.Qf ULTRASOUND IMAGING GUIDEWIRE WITH STATIC CENTRAL CORE AND TIP

David A. White and W. Martin Belef, assignors to Boston Scientific Corporation
3 August 2004 (Class 600/463); filed 1 April 2003

The device is an extension of United States Patents 5,951,480 and 6,171,250. It consists of an ultrasound imaging guidewire featuring a detachable guidewire body and a stationary central core. The goal of the device is to provide multiposition, ultrasonic imaging without loss of correct guidewire positioning and also to simplify the guidewire procedure. The guidewire consists of a static central core and an imaging guidewire. The acoustical scanning device can be rotated to yield 360° acoustical images of



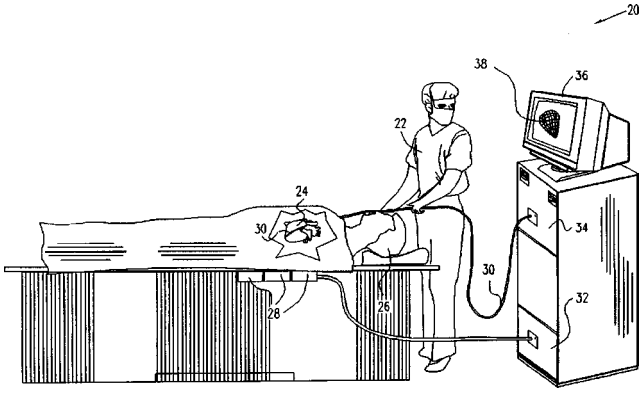
a site of interest within the patient's body. The imaging guidewire also includes a connector that allows the guidewire body to be disengaged from the static central core tip so that the body can be axially translated to obtain multiposition images. The translation may be done without losing the original guidewire positioning because the static central core maintains its position within the patient's body.—DRR

6,773,402

43.80.Qf LOCATION SENSING WITH REAL-TIME ULTRASOUND IMAGING

Assaf Govari and Ilya Beletsky, assignors to Biosense, Incorporated
10 August 2004 (Class 600/459); filed 19 February 2002

This apparatus for mapping surfaces of a cavity within the body of a patient includes an elongated probe with a longitudinal axis and a distal portion adapted for insertion in the cavity of a patient. A primary acoustic transducer on the distal portion of the probe emits acoustic waves within the body cavity. A number of secondary acoustic transducers, deployed along



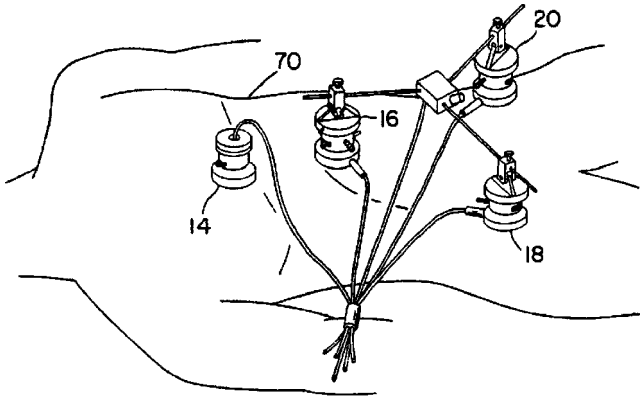
the longitudinal axis over the distal portion of the probe, are adapted to receive the acoustic waves after reflection from the cavity surfaces and to generate, in response to the received waves, electrical signals indicative of times of flight of the waves.—DRR

6,776,766

43.80.Qf METHOD AND APPARATUS FOR CHARACTERIZING GASTROINTESTINAL SOUNDS

Richard H. Sandler and Hussein A. Mansy, assignors to Rush–Presbyterian–St. Luke's Medical Center
17 August 2004 (Class 600/587); filed 9 October 2002

This apparatus for characterizing gastrointestinal sounds includes a microphone array that is positioned on the body. The microphone signals are digitized and a processor determines the spectra and duration. A character-



ization as to the status of the gastrointestinal tract is developed on the basis of the spectra and the duration of the sound or event.—DRR

6,780,154

43.80.Qf SEGMENTED HANDHELD MEDICAL ULTRASOUND SYSTEM AND METHOD

Robert P. Hunt *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
24 August 2004 (Class 600/446); filed 14 January 2003

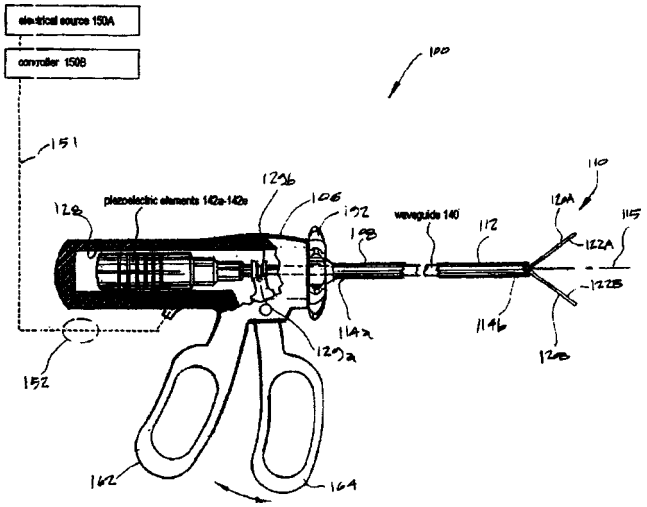
A portable medical ultrasound system is segmented, i.e., divided up into parts, so that its portability may be enhanced. Ultrasound data, such as image data in a video format, is wirelessly transmitted to a multi-use display device from a handheld ultrasound portable device. Any one of various multi-use display devices may be utilized, for example, PDAs, tablet computers, laptop computers, or desktop computers. The handheld portion of the system houses a transducer and an ultrasound processor in communication with the transducers. The multi-use display devices, used separately, receive data wirelessly from the ultrasound processor.—DRR

6,773,409

43.80.Sh SURGICAL SYSTEM FOR APPLYING ULTRASONIC ENERGY TO TISSUE

Csaba Truckai *et al.*, assignors to SURGRx LLC
10 August 2004 (Class 601/2); filed 19 September 2001

The device is a surgical instrument with paired jaws that deliver thermal energy to the engaged tissue to create an effective seal or weld in the tissue. The device applies highly compressive forces to engaged tissue together with the application of ultrasonic energy from two opposing sides of



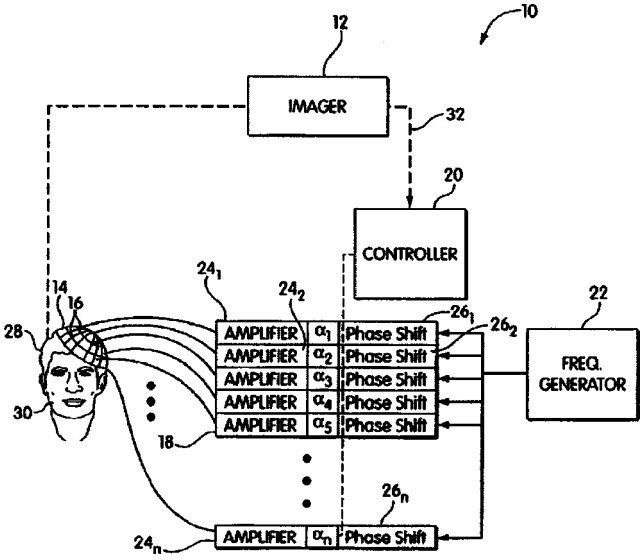
the tissue. The apparatus incorporates an ultrasound transmission assembly including piezoelectric elements coupled to an elongated waveguide. The reciprocatable waveguide assembly is designed to function in multiple ways: (a) by coupling ultrasound energy to both opposing jaws to deliver acoustic energy to both sides of engaged tissues to create uniform stress thermal welds, (b) by applying very high compressive forces to captured tissues over the length of the elongate jaws, and (c) by transecting the captured tissues together with delivery of energy to create the thermal weld. The patent also describes the possible uses of other energy sources to deliver thermal energy to tissue, e.g., microwave energy, rf energy, and laser energy, either in combination with ultrasound or independently.—DRR

6,770,031

43.80.Sh ULTRASOUND THERAPY

Kullervo Hynynen and Gregory T. Clement, assignors to Brigham and Women's Hospital, Incorporated
3 August 2004 (Class 600/437); filed 26 August 2002

This is a sort of a multi-channel system which delivers ultrasound signals to provide an image of at least a portion of a patient to provide identification of physical characteristics of different layers of the material



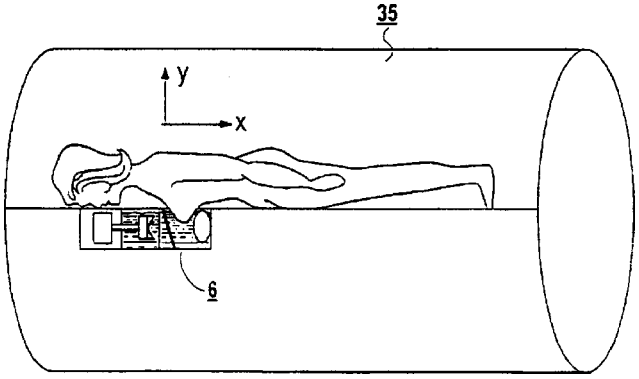
between the ultrasound sources and the region to be treated. A spectral propagation analysis determines phase and amplitude corrections for the sources depending upon the respective physical characteristics of the portions of the material disposed between each source and the target region. The patent asserts that this noninvasive surgical procedure enables better focusing of ultrasound beams onto tumorous or cancerous tissues that can then be ablated without significant damage to surrounding healthy tissue.—DRR

6,778,848

43.80.Sh DEVICE FOR THE ULTRASONIC THERAPY OF A FEMALE BREAST WITH PARALLEL SOUND DIRECTION

Mario Bechtold *et al.*, assignors to Siemens Aktiengesellschaft
17 August 2004 (Class 600/411); filed in Germany 14 October 1997

This device is contoured to provide ultrasonic therapy to the female breast. The breast is inserted into a container filled with a liquid mixture to promote good acoustic coupling. An ultrasonic transducer emits sound into



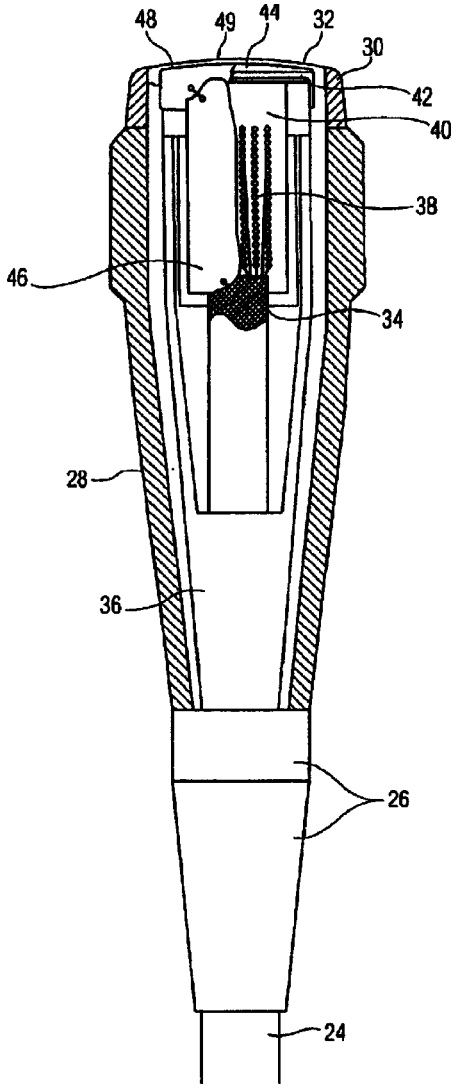
the breast in a limited region. The region, in the positive and negative directions of rotation, as related to the plane of the patient's body, covers angles of up to 50°. In the preferred embodiment, a compression pad and a positioning-fixing membrane are provided for fixing the position of the breast.—DRR

6,776,758

43.80.Vj RFI-PROTECTED ULTRASOUND PROBE

Michael Eugene Peszynski *et al.*, assignors to Koninklijke Philips Electronics N.V.
17 August 2004 (Class 600/437); filed 11 October 2002

The patent relates to the protection of ultrasonic probes from electromagnetic or radio frequency interference. The ultrasonic imaging system includes an electromagnetically protected ultrasound transesophageal echocardiography (TEE) probe, a display for showing ultrasound images, a housing containing a transducer connector, and a computer for processing the echo signals. The TEE probe features a probe housing including an image sensor, an outer acoustic lens, a cable interconnect, an internal metal shield that surrounds the sensor and cable interconnect, and a cable containing electrical connectors and electromagnetic cable shielding. The electrical



conductors are electrically connected to the cable interconnect at one end and, together with cable shielding, to a transducer at the other end.—DRR

6,770,033

43.80.Vj IMAGING METHOD AND DEVICE USING SHEARING WAVES

Mathias Fink *et al.*, assignors to Societe d'Elastographie Impulsionnelle pour les Systemes de Mesure de l'Elasticite (SEISME)
3 August 2004 (Class 600/443); filed in France 15 March 1999

Pulse shear waves in a viscoelastic medium are visualized by first transmitting ultrasonic compression waves at a very high rate to obtain a succession of images and then correlating the resulting images to determine movement of the medium where the shear wave is propagating.—RCW

6,773,399

43.80.Vj BLOCK-SWITCHING IN ULTRASOUND IMAGING

Xufeng Xi *et al.*, assignors to Zonare Medical Systems, Incorporated
10 August 2004 (Class 600/443); filed 20 October 2001

Selected sets of ultrasound transducer elements are used to produce a beam. The beam is scanned by shifting by more than one element (block-switching) between the steps in the scanning process. Loss of resolution is avoided by forming multiple receive beams in areas illuminated by individual transmit beams.—RCW

6,773,400

43.80.Vj NONINVASIVE TRANSCRANIAL DOPPLER ULTRASOUND FACE AND OBJECT RECOGNITION TESTING SYSTEM

Philip Chidi Njemanze, Owerri, Imo, Nigeria
10 August 2004 (Class 600/454); filed 1 April 2002

Two probes placed on the temples are focused on corresponding right and left cerebral arteries of a subject. A so-called laterality index is computed for the arteries. The laterality index and other information is displayed in real time along with a diagram being shown to the subject to elicit responses.—RCW

6,775,404

43.80.Vj APPARATUS AND METHOD FOR INTERACTIVE 3D REGISTRATION OF ULTRASOUND AND MAGNETIC RESONANCE IMAGES BASED ON A MAGNETIC POSITION SENSOR

Niko Pagoulatos *et al.*, assignors to University of Washington
10 August 2004 (Class 382/154); filed 15 March 2000

An intraoperative ultrasound imaging system is combined with a stereotactic system. In the combination, two-dimensional ultrasound and three-dimensional magnetic resonance images are interactively registered. The registration is based on tracking a magnetic sensor that is mounted on an ultrasound probe.—RCW

6,776,760

43.80.Vj MULTI-MODE PROCESSING FOR ULTRASONIC IMAGING

Vasilis Z. Marmarelis, assignor to Alfred E. Mann Institute for Biomedical Engineering at the University of Southern California
17 August 2004 (Class 600/448); filed 5 March 2003

Images that show different tissue types are produced by using matches within a database containing information about different characteristics of tissue types.—RCW

6,780,152

43.80.Vj METHOD AND APPARATUS FOR ULTRASOUND IMAGING OF THE HEART

Kutay F. Üstüner *et al.*, assignors to Acuson Corporation
24 August 2004 (Class 600/443); filed 26 June 2002

Heart images are formed from multiple sets of ultrasound data acquired and processed using different imaging parameters. The parameters, such as array position, temporal-frequency response, or transmit focal depth,

have either laterally or axially different spatial spectra. Sets of images during one cardiac cycle are temporally aligned with sets of images obtained using a different parameter set during another cardiac cycle.—RCW

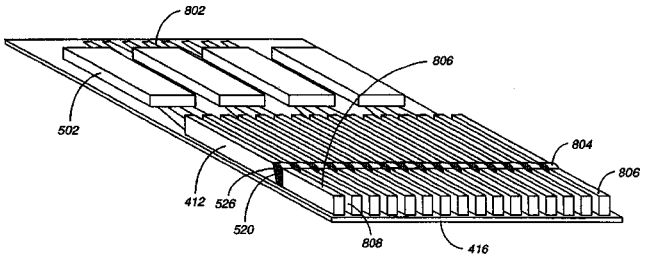
6,780,157

43.80.Vj METHOD AND APPARATUS FOR ULTRASONIC IMAGING

Douglas N. Stephens and Matthew O'Donnell, assignors to Volcano Therapeutics, Incorporated

24 August 2004 (Class 600/466); filed 18 June 2002

An ultrasound transducer array contains elements in which an acoustic discontinuity is cut to produce two sections. One section operates in a thick ness extensional mode. The other section operates in a length extensional mode. These modes of operation enable the transducer elements to provide both forward-looking and side-looking beams.—RCW



LETTERS TO THE EDITOR

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

Time delays in coupled multibubble systems (L)

Alexander A. Doinikov[a)]
Institute of Nuclear Problems, Belarus State University, 11 Bobruiskaya Street, Minsk 220050, Belarus

Richard Manasseh[b)]
Energy and Thermofluids Engineering, CSIRO Manufacturing and Infrastructure Technology, P.O. Box 56, Highett, Melbourne VIC 3190, Australia

Andrew Ooi[c)]
Department of Mechanical and Manufacturing Engineering, University of Melbourne, VIC 3010, Melbourne, Australia

(Received 5 July 2004; revised 5 October 2004; accepted 7 October 2004)

Damping constants of free oscillations of coupled gas bubbles change considerably when time delays due to the finite speed of sound in the surrounding liquid are taken into account. This effect is illustrated analytically by the example of a two-bubble system and numerically by the example of a bubble chain. By comparing the numerical data with experimental measurements, it is shown that allowing for time delays considerably improves agreement between theory and experiment.
 [DOI: 10.1121/1.1828573]

PACS numbers: 43.20.Ks, 43.30.Jx [AJS] Pages: 47–50

I. INTRODUCTION

The purpose of this paper is to draw attention to the effect of time delays on free oscillations of a system of coupled gas bubbles in a liquid. Time delay arises from the finite speed of sound propagation in the liquid, or, in other words, from the finite compressibility of the liquid. As a result, the acoustic pressure field that is experienced by each bubble in the system due to oscillations of all the other bubbles, and which provides acoustic coupling between the bubbles, is a time-retarded field. The literature on this topic is devoted mainly to investigating forced oscillations which are caused by an external acoustic field (Feuillade, 2001; Hsiao *et al.*, 2001); or time-averaged radiation forces between two bubbles which are known as secondary Bjerknes forces (Doinikov and Zavtrak, 1997; Mettin *et al.*, 2000; Doinikov, 2001). In these cases, the effect considered in the present paper does not manifest itself distinctly. However, there are many applications, such as feedback measurements of industrial aerators, emission of sound by submarines, oceanic bubble dynamics, etc., where it is the free oscillations of multibubble systems that are the subject of interest (Pumphrey and Crum, 1990; Boyd and Varley, 2001; Manasseh *et al.*, 2001; Chanson and Manasseh, 2003). The present paper shows that in the context of simulating such processes, time delay plays a very important role.

II. EQUATIONS OF FREE OSCILLATIONS OF A MULTIBUBBLE SYSTEM

Classically, small free-radial oscillations of N coupled bubbles are described by the following equations (Leighton, 1994):

$$\ddot{x}_n(t)+\omega_{n0}\delta_n\dot{x}_n(t)+\omega_{n0}^2x_n(t)=-\sum_{\substack{m=1\\m\neq n}}^{N}\frac{R_{m0}^2}{R_{n0}d_{nm}}\ddot{x}_m(t),\quad n=1,...,N, \tag{1}$$

where $x_n(t)$ is the small change in the radius of the nth bubble, ω_{n0}, δ_n, and R_{n0} are the resonant angular frequency, the total damping constant, and the equilibrium radius of the nth bubble, respectively, and d_{nm} is the distance between the centers of the nth and mth bubbles. However, these equations hold only for an incompressible liquid. For a compressible liquid, with a finite speed of sound propagation, they take the form (Fujikawa and Takahira, 1986; Mettin *et al.*, 2000)

[a)]Electronic mail: doinikov@inp.minsk.by
[b)]Electronic mail: richard.manasseh@csiro.au
[c)]Electronic mail: a.ooi@unimelb.edu.au

$$\ddot{x}_n(t)+\omega_{n0}\delta_n\dot{x}_n(t)+\omega_{n0}^2x_n(t)$$
$$=-\sum_{\substack{m=1\\m\neq n}}^{N}\frac{R_{m0}^2}{R_{n0}d_{nm}}\ddot{x}_m(t-d_{nm}/c),\quad n=1,...,N, \tag{2}$$

where c is the sound speed in the surrounding liquid. It should also be mentioned that in both Eqs. (1) and (2), the separation distances between bubbles are assumed to be large compared to the bubble radii, $d_{nm}\gg R_{n0}$. In the next sections, predictions of these two models are compared for the cases of a two-bubble system and a bubble chain.

III. TWO-BUBBLE SYSTEM

Let us first consider a system that consists of two interacting bubbles of equal size. This case can be analyzed analytically.

A. Without time delays

For the case of two equal bubbles, Eqs. (1) give

$$\ddot{x}_1(t)+b\dot{x}_1(t)+\omega_0^2x_1(t)=-\xi\ddot{x}_2(t), \tag{3a}$$

$$\ddot{x}_2(t)+b\dot{x}_2(t)+\omega_0^2x_2(t)=-\xi\ddot{x}_1(t), \tag{3b}$$

where $b=\omega_0\delta$, $\delta=\delta_r+\delta_t+\delta_v$ is the total damping constant which includes the radiation (δ_r), thermal (δ_t), and viscous (δ_v) dissipation, and $\xi=R_0/d$, where $R_0=R_{10}=R_{20}$ and $d=d_{12}$. Expressions for the damping constants δ_r, δ_t, and δ_v can be found in Clay and Medwin (1977), and will not be repeated here.

Solutions to Eqs. (3) are sought as

$$x_n(t)=a_n\exp(i\nu t),\quad n=1,2. \tag{4}$$

Substituting Eqs. (4) into Eqs. (3), and operating in the standard way, one obtains the following two eigenvalues of system (3):

$$\nu_1=\frac{ib+2\omega_0\sqrt{1+\xi-\delta^2/4}}{2(1+\xi)}, \tag{5a}$$

$$\nu_2=\frac{ib+2\omega_0\sqrt{1-\xi-\delta^2/4}}{2(1-\xi)}. \tag{5b}$$

Note that, mathematically, it is more correct to call $i\nu_1$ and $i\nu_2$ the eigenvalues of Eqs. (3). However, in the present paper it is more convenient to apply this term to ν_1 and ν_2, and we will do this below. Note also that Eqs. (3) have actually four eigenvalues but the other two are simply complex conjugates to $i\nu_1$ and $i\nu_2$ and therefore they are not given here to save space.

It is convenient to represent ν_n as $\nu_n=\omega_n+i\alpha_n$, where ω_n is then the eigenfrequency and α_n is the damping coefficient of the eigenmode corresponding to the eigenvalue ν_n. From Eqs. (5), one then obtains

$$\omega_1=\omega_0\frac{\sqrt{1+\xi-\delta^2/4}}{1+\xi},\quad \alpha_1=\frac{\omega_0\delta}{2(1+\xi)}, \tag{6a}$$

$$\omega_2=\omega_0\frac{\sqrt{1-\xi-\delta^2/4}}{1-\xi},\quad \alpha_2=\frac{\omega_0\delta}{2(1-\xi)}. \tag{6b}$$

These equations show that

$$\omega_1<\omega_2,\quad \alpha_1<\alpha_2, \tag{7}$$

i.e., the lower-frequency mode has a lower damping coefficient than the higher-frequency mode. This result is well known in the literature and reproduced here only for the purpose of further comparing with results of the next subsection.

B. With time delays

For the same case of two equal bubbles, Eqs. (2) give

$$\ddot{x}_1(t)+b\dot{x}_1(t)+\omega_0^2x_1(t)=-\xi\ddot{x}_2(t-\tau), \tag{8a}$$

$$\ddot{x}_2(t)+b\dot{x}_2(t)+\omega_0^2x_2(t)=-\xi\ddot{x}_1(t-\tau), \tag{8b}$$

where $\tau=d/c$. For $\tau\ll T_0$, where $T_0=2\pi/\omega_0$, $\ddot{x}_n(t-\tau)$ can be approximately represented by using a Taylor series as

$$\ddot{x}_n(t-\tau)\approx\ddot{x}_n(t)-\dddot{x}_n(t)\tau. \tag{9}$$

Introducing this approximation will increase the order of Eqs. (8). In order to perform eigenanalysis, it would be desirable to first reduce the order of Eq. (9) to be the same as the original equations, i.e., Eqs. (8). To do this, it is evident from Eqs. (8) that

$$\ddot{x}_n(t)\approx-b\dot{x}_n(t)-\omega_0^2x_n(t). \tag{10}$$

Recall that the spacing between the bubbles is assumed to be large compared to their size, i.e, $\xi\ll1$. It follows that the term $-\xi\ddot{x}_m(t-\tau)$ can be neglected in Eq. (10) as the right-hand sides of Eqs. (8), into which Eqs. (9) and (10) will be substituted, are already of the first order in ξ. Differentiating Eq. (10), one finds

$$\dddot{x}_n(t)\approx-b\ddot{x}_n(t)-\omega_0^2\dot{x}_n(t). \tag{11}$$

Substituting Eq. (11) into Eq. (9), one obtains

$$\ddot{x}_n(t-\tau)\approx(1+b\tau)\ddot{x}_n(t)+\tau\omega_0^2\dot{x}_n(t). \tag{12}$$

Finally, substitution of Eq. (12) into Eqs. (8) yields

$$\ddot{x}_1(t)+b\dot{x}_1(t)+\omega_0^2x_1(t)+\omega_0\delta_r\dot{x}_2(t)+\xi(1+b\tau)\ddot{x}_2(t)=0, \tag{13a}$$

$$\ddot{x}_2(t)+b\dot{x}_2(t)+\omega_0^2x_2(t)+\omega_0\delta_r\dot{x}_1(t)+\xi(1+b\tau)\ddot{x}_1(t)=0. \tag{13b}$$

Solutions to Eqs. (13) are sought as before in the form of Eqs. (4). This leads to

$$\nu^2-ib\nu-\omega_0^2=\pm[i\omega_0\delta_r\nu-(\xi+\delta\delta_r)\nu^2]. \tag{14}$$

Taking the plus sign in front of the right-hand side, one finds the first eigenvalue

$$\nu_1=\frac{i\omega_0(\delta+\delta_r)+2\omega_0\sqrt{1+\xi+\delta\delta_r-(\delta+\delta_r)^2/4}}{2(1+\xi+\delta\delta_r)}, \tag{15}$$

which yields

$$\omega_1=\omega_0\frac{\sqrt{1+\xi+\delta\delta_r-(\delta+\delta_r)^2/4}}{1+\xi+\delta\delta_r},$$

$$\alpha_1=\frac{\omega_0(\delta+\delta_r)}{2(1+\xi+\delta\delta_r)}. \tag{16}$$

With the minus sign, one obtains

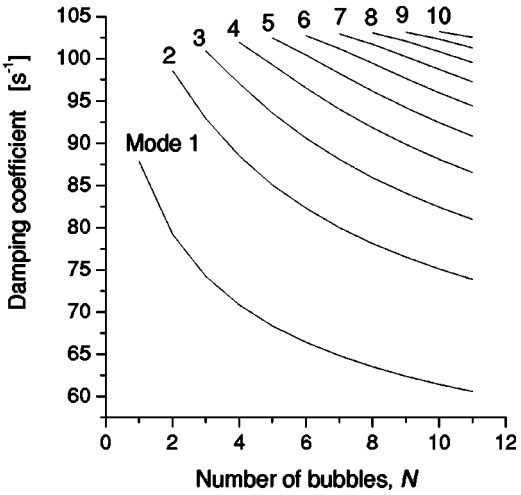


FIG. 1. Variation of damping coefficients with number of bubbles with no time delays. The host liquid is water, the gas within the bubbles is air, the equilibrium bubble radius $R_0 = 3.5$ mm, and the separation distance between the bubbles $\Delta z = 32.1$ mm.

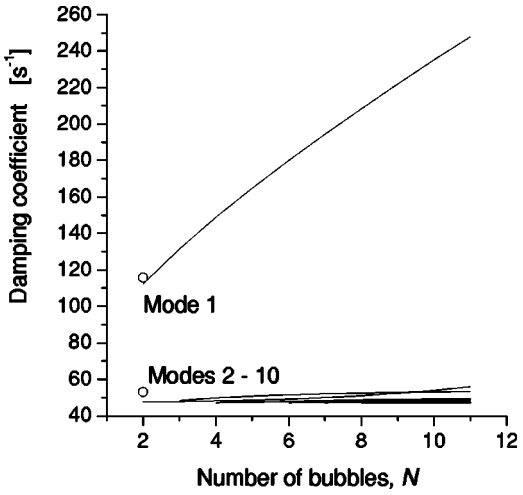


FIG. 2. Variation of damping coefficients with number of bubbles when time delays are present. Parameters are the same as in Fig. 1. The two open circles show predictions of the approximate equations (16) and (18).

$$\nu_2 = \frac{i\omega_0(\delta - \delta_r) + 2\omega_0\sqrt{1-\xi-\delta\delta_r-(\delta-\delta_r)^2/4}}{2(1-\xi-\delta\delta_r)}, \quad (17)$$

and correspondingly

$$\omega_2 = \omega_0 \frac{\sqrt{1-\xi-\delta\delta_r-(\delta-\delta_r)^2/4}}{1-\xi-\delta\delta_r},$$

$$\alpha_2 = \frac{\omega_0(\delta-\delta_r)}{2(1-\xi-\delta\delta_r)}. \quad (18)$$

One can see from these equations that

$$\omega_1 < \omega_2, \quad \alpha_1 > \alpha_2, \quad (19)$$

provided that $d > d_{cr} = \delta c/(\omega_0(1-\delta^2))$, i.e., there is a critical bubble spacing above which damping becomes higher for the lower-frequency mode. Thus, incorporating time delay reverses the situation with damping so that the lower-frequency mode has a higher damping coefficient and hence should subside faster.

IV. BUBBLE CHAIN

In this section, Eqs. (1) and (2) are applied to a bubble chain. It is assumed that all bubbles are of equal size and equally spaced. The eigenvalues of both systems are computed in the following manner. Equation (4) is substituted into, for example, system (1). This leads to a homogeneous system of linear algebraic equations in the unknowns a_n. The determinant of the system is then calculated and set equal to zero. In doing so, we get the characteristic equation of the system as a function of ν. The eigenvalues of the system are calculated as roots of the characteristic equation. The difference between systems (1) and (2) is that the characteristic equation of the former is a polynomial of degree $2N$ in ν, while that of the latter is a nonlinear equation in terms of the unknown ν. Calculations were conducted by using FORTRAN routines from the program package IMSL MATH/LIBRARY.

Numerical simulations were done for air bubbles in water. The equilibrium radius of the bubbles R_0 and the separation distance between the bubbles Δz were set to $R_0 = 3.5$ mm and $\Delta z = 32.1$ mm, respectively. These values correspond to experiments by Manasseh *et al.* (2004) as detailed in the next section. The damping co-efficients α_n of ten modes that were calculated by Eqs. (1) (i.e., without time delays) are shown in Fig. 1. This figure displays how the damping coefficients vary as the number of the bubbles in the chain is increased. It is seen that the higher the eigenfrequency, the higher the damping coefficient, and that all of them decrease as the number of the bubbles is increased. The damping coefficients calculated by Eqs. (2) (i.e., with time delays) are shown in Fig. 2. The two open circles were obtained from the approximate equations (16) and (18) and are given for comparison. The plots in Fig. 2 are clearly different from those in Fig. 1. The damping coefficient of the lowest frequency mode, mode 1, is much higher than those of all the other modes and increases rapidly with increasing number of the bubbles. Figure 3 shows a scaled-up plot of the damping coefficients for modes 2 to 10. It is seen that the magnitudes

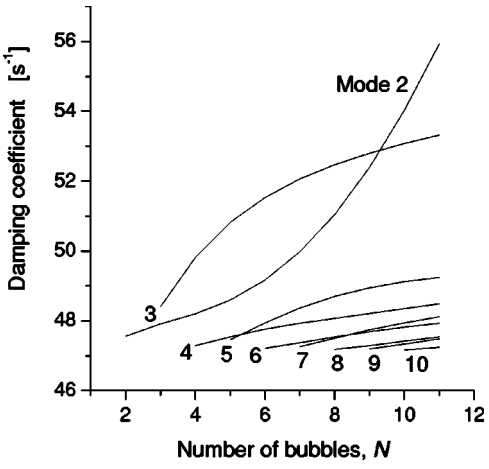


FIG. 3. Variation of damping coefficients with number of bubbles when time delays are present: Enlarged view for modes 2 to 10 in Fig. 2.

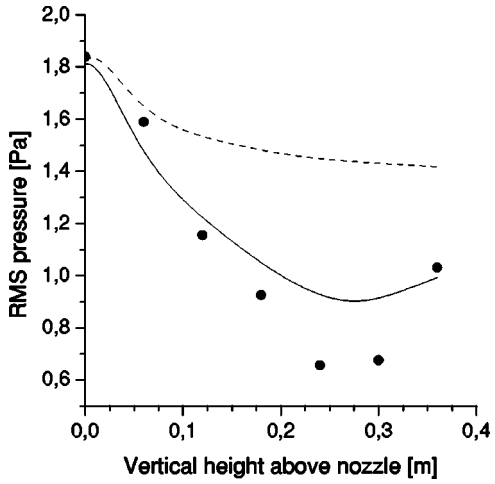


FIG. 4. Experimental and model vertical rms pressure along a vertical line 0.06 m from the nozzle axis for a 16-bubble chain. Experimental data (full circles) were adopted from Manasseh *et al.* (2004). The dashed line is the numerical result without time delay and the solid line is the numerical result with time delay.

of the damping coefficients are fairly close, especially for the higher-frequency modes. They do not appear to increase significantly with increasing N. In other words, comparing Figs. 1 and 2, one can say that time delays result in a strong suppression of the lowest frequency mode and an equalization of the damping of all the other modes.

V. COMPARISON WITH EXPERIMENT

The significance of time delays is well demonstrated by comparing predictions of Eqs. (1) and (2) with experiments by Manasseh *et al.* (2004) in which the distribution of acoustic pressure around a chain of rising bubbles was measured. The chain was produced in a water-filled tank by the nozzle. The nozzle was supplied with air under pressure producing bubbles in a highly repeatable manner. The acoustic pressure field was induced by free oscillations of the bubbles in the chain. The oscillations developed when the newly formed bubble at the base of the chain detached from the nozzle, creating sound on its detachment. The full circles in Fig. 4 show the profile of rms pressure along a vertical line 0.06 m from the nozzle axis that was experimentally obtained by Manasseh *et al.* for a bubble chain consisting of 16 bubbles. Manasseh *et al.* had measured a significantly anisotropic sound-pressure distribution around the chain: in the radial direction pressure fell off roughly as expected for a monopole point source, while in the vertical direction (parallel to the chain) pressure did decrease with distance from the nozzle but more gradually. Manasseh *et al.* used a coupled-oscillator model that neglected time delays; it qualitatively predicted the anisotropy between vertical and radial directions but failed to predict the correct vertical profile of rms pressure. Parameters for the numerical model based on the present time-delay formulation were chosen to correspond with the experimental data. The initial equilibrium radius of the bubbles and the separation distance between them were set to $R_0 = 3.5\,\mathrm{mm}$ and $\Delta z = 32.1\,\mathrm{mm}$, respectively. In addition, it is assumed that the initial radial velocity of the bottom bubble is $\dot{R}(t=0) = 9\,\mathrm{mm/s}$, while that of all the other bubbles is equal to zero. The solid line in Fig. 4 shows the numerical data obtained by solving Eqs. (2). The numerical prediction using Eqs. (1) with similar initial conditions is shown by the dashed line. It is seen that allowing for time delays provides a better agreement with experiment data.

VI. CONCLUSIONS

It has been shown theoretically by the examples of a two-bubble system and a bubble chain that time delays due to the finite compressibility of the surrounding liquid considerably change the damping coefficients of the eigenmodes of a coupled multibubble system, in particular reversing the trend of higher damping for higher frequencies, provided bubbles are further than a critical distance. The change is such that the lowest frequency mode acquires the highest damping coefficient, while the damping coefficients of all the other modes are much smaller and close to one another. It has also been shown by comparison of numerical and experimental results that the time-delay model provides a better agreement with experiment. This gives confidence that the predicted change in the behavior of the damping coefficients by the time-delay model is a physical phenomenon.

Boyd, J. W. R., and Varley, J. (**2001**). ''The uses of passive measurement of acoustic emissions from chemical engineering processes,'' Chem. Eng. Sci. **56**, 1749–1767.

Chanson, H., and Manasseh, R. (**2003**). ''Air entrainment processes of a circular plunging jet: Void-fraction and acoustic measurements,'' J. Fluids Eng. **125**, 910–921.

Clay, C. S., and Medwin, H. (**1977**). *Acoustical Oceanography* (Wiley-Interscience, New York).

Doinikov, A. A. (**2001**). ''Acoustic radiation interparticle forces in a compressible fluid,'' J. Fluid Mech. **444**, 1–21.

Doinikov, A. A., and Zavtrak, S. T. (**1997**). ''Radiation forces between two bubbles in a compressible liquid,'' J. Acoust. Soc. Am. **102**, 1424–1431.

Feuillade, C. (**2001**). ''Acoustically coupled gas bubbles in fluids: Time-domain phenomena,'' J. Acoust. Soc. Am. **109**, 2606–2615.

Fujikawa, S., and Takahira, H. (**1986**). ''A theoretical study on the interaction between two spherical bubbles and radiated pressure waves in a liquid,'' Acustica **61**, 188–199.

Hsiao, P.-Y., Devaud, M., and Bacri, J.-C. (**2001**). ''Acoustic coupling between two air bubbles in water,'' Eur. Phys. J. E **4**, 5–10.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic, London).

Manasseh, R., Nikolovska, A., Ooi, A., and Yoshida, S. (**2004**). ''Anisotropy in the sound field generated by a bubble chain,'' J. Sound Vib. **278**, 807-823.

Manasseh, R., LaFontaine, R. F., Davy, J., Shepherd, I. C., and Zhu, Y. (**2001**). ''Passive acoustic bubble sizing in sparged systems,'' Exp. Fluids **30**, 672–682.

Mettin, R., Luther, S., Kamphausen, S., and Lauterborn, W. (**2000**). ''Dynamics of delay-coupled spherical bubbles,'' in *Proceedings of 15th International Symposium on Nonlinear Acoustics, Goettingen, Germany, 1–4 Sept., 1999* (AIP Melville, NY, 2000), pp. 359–362.

Pumphrey, H. C., and Crum, L. A. (**1990**). ''Free oscillations of near-surface bubbles as a source of the underwater noise of rain,'' J. Acoust. Soc. Am. **87**, 142–148.

The overtone structure of acoustic fluctuations due to harmonic variations in sound speed (L)

Robert L. Field[a] and Jacob George
Naval Research Laboratory, Stennis Space Center, Mississippi 39529

(Received 14 November 2002; revised 23 June 2004; accepted 8 July 2004)

Explicit relations are derived between the acoustic pressure and a sound speed that is periodic. They are derived for the Pekeris waveguide, and extended to the range and depth-dependent adiabatic mode/WKB case. As expected, acoustic pressure is found to fluctuate at the sound speed fundamental and its overtones. A Pekeris model simulation qualitatively agrees with the intensity fluctuation spectrum from a range and depth-dependent parabolic equation (PE) simulation of a shelf edge experiment off the New Jersey coast. In this paper we show that the energy of the higher overtones can be significant and derive a modulation index that determines the factors that contribute to the overtone energy. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1789999]

PACS numbers: 43.30.Re, 43.30.Bp [WLS] Pages: 51–54

I. INTRODUCTION

Our purpose in this paper is to derive an explicit relationship between acoustic pressure and geotime for an ocean where the sound speed undergoes periodic variations such as generated by long wavelength internal tides. Since fluctuations in acoustic transmissions have been extensively studied from different points of view, a brief review of only the studies most germane to this work is given below.

The effect of geostrophic flow on sound speed and thereby on the intensity and phase of acoustic transmission has been analyzed by Jacobson *et al.*[1–4] These authors used ray theory, and considered long range transmission in the deep ocean over times of typically days. Intensity variations due to tidal depth changes found by Baer[3] are similar to those discussed below for shallow water. An expression for intensity variations derived by Porter[5] for the problem in Ref. 1 has the same form as Eq. (8) below for normal modes in shallow water. The discussion below goes further and explicitly derives the harmonic structure of acoustic fluctuations, which is the topic of the present paper.

In 2000 an acoustic experiment was performed in the New Jersey Bight off the East Coast of the United States.[6] In part of the experiment, the acoustic transmissions were along the continental shelf. As discussed in Ref. 6, the spectrum of the acoustic transmission time series at 854.5 Hz clearly showed the harmonics of the M_2 tide, including the fundamental frequency, $F_0(=2.1094\,\text{cpd})$ and the overtones, $F_1(=4.2188\,\text{cpd})$ and $F_2(=6.3281\,\text{cpd})$.

In this paper a geotime-dependent Pekeris equation is derived, and extended to the range- and depth-dependent adiabatic mode/WKB model. The spectrum from the Pekeris equation is compared to a PE simulation to show that the overtone structure holds in a range- and depth-dependent waveguide.

Although the harmonic structure of acoustic intensity fluctuations may be obvious from a mathematical point of view, one must ask if higher overtones can be generated from a well-defined sound speed, how energetic can they be, and if the higher overtone energy is significant, what are the factors that drive the spectrum to them? Without knowing the answer to these questions, measurements and model simulations are problematic. For example, for modeling, what time step should be chosen in order not to alias the higher harmonic energy?

In the paper we derive a modulation index for the Pekeris waveguide that determines the factors that drive the intensity spectrum to higher harmonics. Although in the paper we focus on a simple tidal variation in sound speed, the approach can be extended to more general ocean processes.

II. THE GEOTIME-DEPENDENT WAVEGUIDE

In this section, the expression for the geotime-dependent Pekeris waveguide is derived, and extended to the WKB case. The complex pressure for the trapped modes in the Pekeris waveguide is given by (Ref. 7, Eq. 5.146)

$$p^P(r,z)=\frac{\sqrt{2\pi}e^{i\pi/4}}{\rho}\sum_{n=1}^{N_{\max}}\Gamma_n^2\sin(k_{zn}^P z_0)\sin(k_{zn}^P z)\frac{e^{ik_n^P r}}{\sqrt{k_n^P r}}, \tag{1}$$

where $p^P(r,z)$, k_{zn}^P, k_n^P are the complex pressure, and vertical and horizontal wave numbers for the Pekeris waveguide, respectively. The factor Γ_n is a function of water column density, ρ, ocean bottom density, water depth, h, sound speed in water, $C(t)$, ocean bottom sound speed, acoustic angular frequency, ω, and the vertical wavenumber, k_{zn}^P.

Only the time dependence of $e^{ik_n^P r}$ is derived since it accounts for nearly all the fluctuation energy. The coefficient of this factor on the right-hand side of Eq. (1) accounts for only fractions of a decibel in fluctuation energy.

The vertical wave number for the Pekeris waveguide is given by (Ref. 7, Eq. 5.140)

$$k_{zn}^P=\frac{(n-1/2)\pi}{h}-\frac{\varphi_n}{2h}\equiv k_{zn}^R-\frac{\varphi_n}{2h}, \tag{2}$$

[a]Electronic mail: bob.field@nrlssc.navy.mil

where the superscript, R, denotes the corresponding wave number for the rigid waveguide. The phase factor, φ_n, for the nth mode, is an implicit function of the sound speed and time (Ref. 7, Eq. 5.138).

The square of the horizontal wave number for the Pekeris waveguide is $(k_n^P)^2=(\omega/C(t))^2-(k_{zn}^P)^2$. The sound speed as a function of time, $C(t)$, can be written as

$$C(t)=C_0(1+\Delta C/C_0)\equiv C_0+A\cos\Omega t, \tag{3}$$

where C_0 is the average value, A is a constant, and $\Omega=2\pi F_0$, where F_0 is the frequency of the ocean sound speed. Since $\Delta C/C_0\ll 1$ even for large sound speed fluctuations, $C(t)$ can be expanded in a binomial series, keeping only the zeroth- and first-order terms. Again using the binomial expansion, the Pekeris horizontal wave number can be written as

$$k_n^P=k_{n0}^R+\frac{1}{k_{n0}^R}\left(\frac{k_{zn}^R\varphi_n}{2h}-\frac{\varphi_n^2}{8h^2}-\frac{\omega^2\Delta C}{C_0^3}\right), \tag{4}$$

where k_{n0}^R is the value of k_n^R evaluated at $C=C_0$. The phase, $\varphi_n(C)$, can be expanded in a Taylor series about C_0 for each mode. Denoting $\varphi_n(C)$ by φ_n, $\varphi_n(C_0)$ by φ_{n0}, and $\varphi_n'(C_0)$ by φ_{n0}' (where the prime denotes the derivative with respect to C), one can write

$$k_n^P r=k_{n0}^{P1}r-I_n^P\cos\Omega t, \tag{5}$$

where

$$k_{n0}^{P1}\equiv k_{n0}^R+\left(k_{zn}^R-\frac{\varphi_{n0}}{4h}\right)\frac{\varphi_{n0}}{2hk_{n0}^R}, \tag{6}$$

is the first-order square root approximation of k_n^P evaluated at C_0 and

$$I_n^P\equiv\left[-\left(k_{zn}^R-\frac{\varphi_{n0}}{2h}\right)\frac{\varphi_{n0}'}{2h}+\frac{\omega^2}{C_0^3}\right]\frac{rA}{k_{n0}^R}. \tag{7}$$

The complex exponential term of Eq. (1) becomes

$$\begin{aligned}e^{ik_n^P r}&=e^{ik_{n0}^{P1}r}e^{-iI_n^P\cos\Omega t}\\&=e^{ik_{n0}^{P1}r}[\cos(I_n^P\cos(\Omega t))-i\sin(I_n^P\cos(\Omega t))]\end{aligned} \tag{8}$$

The acoustic pressure is frequency modulated by the time-dependent part of the sound speed with a modulation index given by I_n^P. Using Bessel identities for the terms in the square bracket, Eq. (8) becomes[8]

$$\begin{aligned}e^{ik_n^P r}=e^{ik_{n0}^{P1}r}\Bigg\{&\left[J_0(I_n^P)+2\sum_{l=1}^{\infty}(-1)^l J_{2l}(I_n^P)\cos(2l\Omega t)\right]\\&-i\left[2\sum_{l=0}^{\infty}(-1)^l J_{2l+1}(I_n^P)\cos[(2l+1)\Omega t]\right]\Bigg\}.\end{aligned}$$

Substituting this expression in Eq. (1) and changing the orders of summation we get the explicit, time-dependent expression relating the complex acoustic pressure to the sound speed,

$$\begin{aligned}p^P(r,z,t)=\frac{\sqrt{2\pi}e^{i\pi/4}}{\rho}\Bigg\{&\left[\sum_{n=1}^{N_{\max}}\Gamma_{n0}^2\frac{\sin(k_{zn0}^P z_0)\sin(k_{zn0}^P z)}{\sqrt{k_{n0}^P r}}e^{ik_{n0}^{P1}r}J_0(I_n^P)\right]\\&+\left[2\sum_{l=1}^{\infty}(-1)^l\cos(2l\Omega t)\sum_{n=1}^{N_{\max}}\Gamma_{n0}^2\frac{\sin(k_{zn0}^P z_0)\sin(k_{zn0}^P z)}{\sqrt{k_{n0}^P r}}e^{ik_{n0}^{P1}r}J_{2l}(I_n^P)\right]\\&-\left[i2\sum_{l=0}^{\infty}(-1)^l\cos[(2l+1)\Omega t]\sum_{n=1}^{N_{\max}}\Gamma_{n0}^2\frac{\sin(k_{zn0}^P z_0)\sin(k_{zn0}^P z)}{\sqrt{k_{n0}^P r}}e^{ik_{n0}^{P1}r}J_{2l+1}(I_n^P)\right]\Bigg\}.\end{aligned} \tag{9}$$

The time-independent terms Γ_{n0}^2, k_{zn0}^P, and k_{n0}^P are all evaluated at the average sound speed, C_0. Equation (9) clearly shows that the complex acoustic pressure can be expressed as a sum of harmonics of the fundamental ocean frequency Ω.

It must be pointed out that Eq. (1) is computationally more efficient than Eq. (9). The purpose of deriving Eq. (9) has been to explicitly relate sound speed to acoustic fluctuations. The main result of the derivation is the modulation index, I_n^P, given in Eq. (7). As I_n^P increases, the acoustic energy fluctuates more at the higher harmonics of F_0. For very weak modulation, $I_n\approx 0$, the higher-order Bessel functions vanish, the zeroth-order Bessel function is 1, and Eq. (9) reduces to Eq. (1). If the phase terms are set to zero in Eqs. (6) and (7), Eq. (9) reduces to the geotime-dependent rigid waveguide.

An extension of the above results to the depth and range-dependent adiabatic mode/WKB model is straightforward, and is sketched below. The expression for the acoustic pressure in the adiabatic mode/WKB model is very similar to that for the Pekeris waveguide given in Eq. (1) [Ref. 7, Appendix A, Eq. (A25) and Ref. 9]. The major difference is that the quantity $k_n^P r$ in the Pekeris case is replaced by the integral $\int_0^r dr'\,k_n(r')$ in the WKB model. For long-wavelength ocean processes it is reasonable to assume that the sound speed at a given range is a smooth function of time. Then the local horizontal wave number for the nth mode can also be expected to be a smooth function of time, and the wave

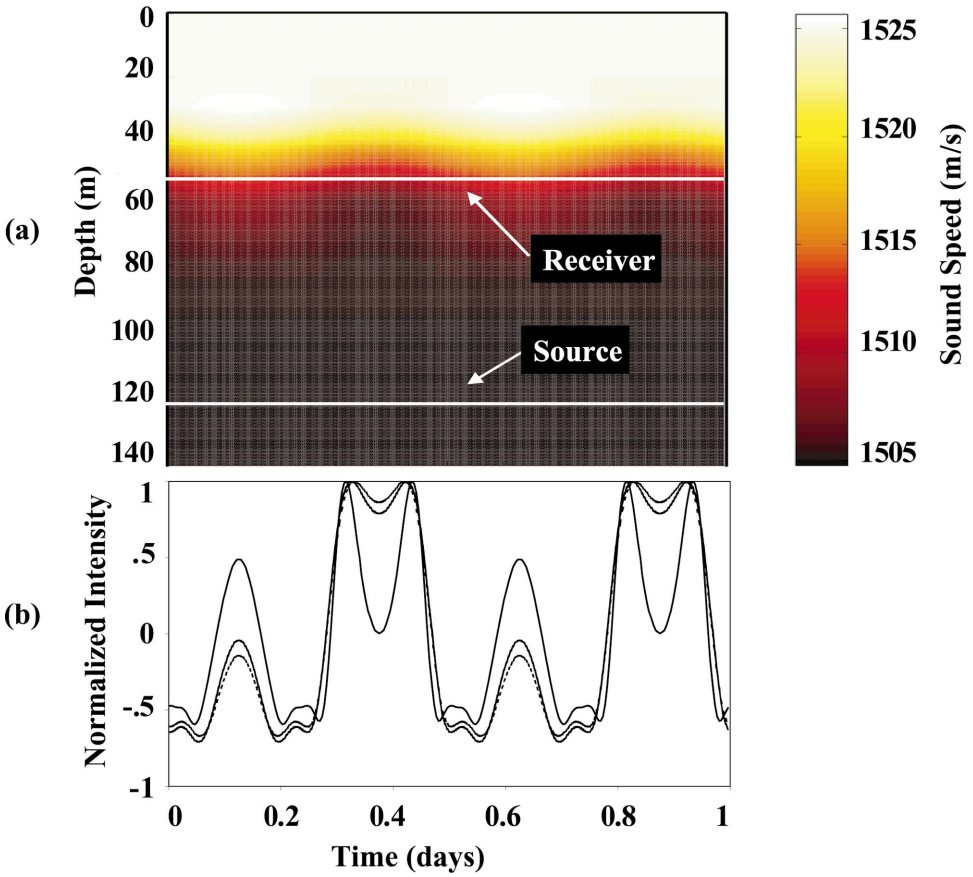


FIG. 1. (a) Sound speed as a function of depth and time used in the PE simulation. The amplitude of the sound speed variation, A, is a function of depth. The depths of the source and receiver are marked. The direction of acoustic propagation is into the plane of the figure. (b) Normalized intensity time series from the PE simulation (solid), Eq. (1) (dashed), and Eq. (9) (gray).

number can be expanded in a Fourier series.[10] The integral of the modal wavenumber over the range can be written as

$$\int_0^r dr' k_n(r',t) = A_{n0}(r)/2 + \sum_{m=1}^{\infty} A_{nm}(r)\cos(m\Omega t) + \sum_{m=1}^{\infty} B_{nm}(r)\sin(m\Omega t), \tag{10}$$

where $A_{nm}(r) \equiv \int_0^r dr' a_{nm}(r')$, $B_{nm}(r) \equiv \int_0^r dr' b_{nm}(r')$, and $a_{nm}(r)$ and $b_{nm}(r)$ are the Fourier coefficients. In the Pekeris case, the $\exp(ik_n r)$ term (except for a constant phase factor) is

$$\exp[-iI_n^P \cos(\Omega t)], \tag{11}$$

as given in Eq. (8). The corresponding term in the adiabatic/WKB model is

$$\exp\left[i\sum_{m=1}^{\infty} A_{nm}(r)\cos(m\Omega t) + i\sum_{m=1}^{\infty} B_{nm}(r)\sin(m\Omega t)\right]. \tag{12}$$

The only difference between expressions (11) and (12) is that the former contains just one sinusoidal function, while the latter contains a sum of sinusoidal functions. Therefore, the harmonic structure of acoustic fluctuations is expected to hold in range- and depth-dependent waveguides.

III. SIMULATION RESULTS

Results from a PE simulation of the experiment in Ref. 6 are compared with results from Eqs. (1) and (9). The parameters used are as follows: bathymetry at the source=132 m, bathymetry at the receiver=144 m, source–receiver range =9715 m, source depth=123 m, receiver depth=56 m, bottom speed=1800 m/s, bottom density=2.0 g/cm^3, oceanographic frequency, F_0=2.0 cpd, and acoustic frequency =854.5 Hz. In the Pekeris calculation the average sound speed at the thermocline, 1517 m/s, and the water depth at the source, 132 m, were used.

Figure 1(a) shows the synthetic sound speed as a function of depth and time used for the PE simulation. Figure 1(b) shows the corresponding geotime variations of the acoustic intensity ($|P|^2$) in linear units (not decibels), calculated by PE and by Eqs. (1) and (9). Each time series has its mean subtracted and the result normalized. Figure 2 shows

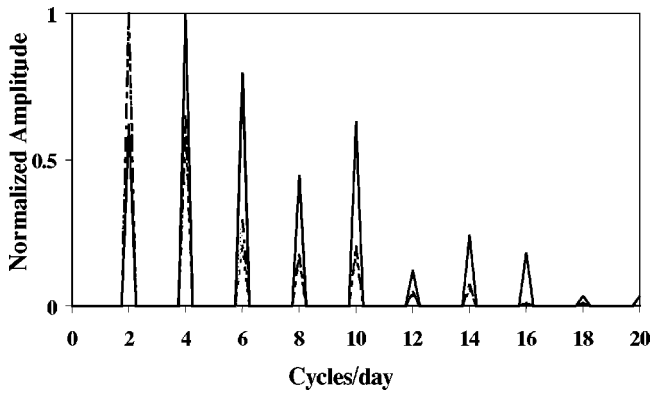


FIG. 2. Normalized spectra of a four-day intensity time series, PE simulation (solid), Eq. (1) (dashed), and Eq. (9) (gray).

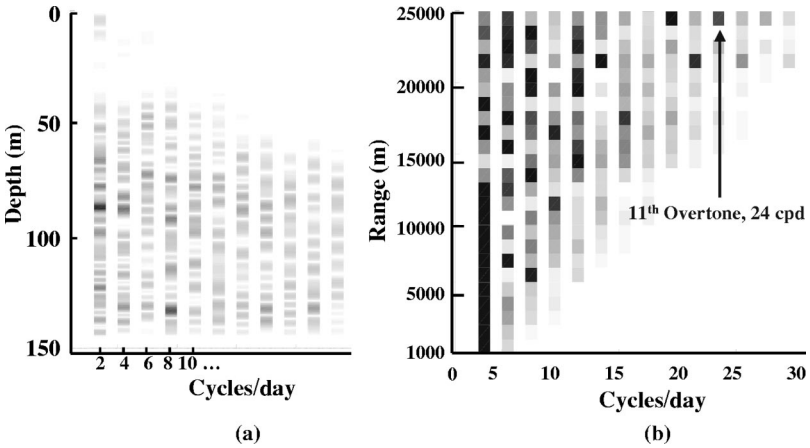


FIG. 3. (a) Normalized intensity spectra as a function of depth from the PE simulation. (Black=high intensity, white=low intensity.) Most of the acoustic energy is trapped between 50 m and the bottom (144 m). The harmonic structure holds regardless of depth. (b) Normalized intensity spectra as a function of range computed from Eq. (1). (Black=high intensity, white=low intensity.) As the range increases, the modulation becomes stronger, driving the spectra to higher harmonic components. At 25 km, significant overtones occur at and beyond 24 cpd, corresponding to time scales of less than one hour.

the spectra of Fig. 1(b) (extended to four days), revealing the harmonic structure.

Figure 3(a) shows the intensity spectra of the PE simulation as a function of depth at the range of 9715 m. The display is normalized by the maximum value. Figure 3(b) shows the normalized spectra at the receiver depth of 56 m as a function of range. As the range increases, the modulation index, I_n^P, becomes proportionally larger and produces higher harmonics as may be seen from the shapes of the Bessel functions in Eq. (9).

IV. SUMMARY AND CONCLUSIONS

Explicit relations have been derived between the acoustic pressure and a sound speed that is periodic in time. This was done for the Pekeris waveguide, and extended to the range- and depth-dependent adiabatic mode/WKB case. As expected, acoustic pressure has been found to fluctuate at the sound speed fundamental and its overtones. A Pekeris model simulation qualitatively agrees with the intensity fluctuation spectrum from a PE simulation of a shelf edge experiment in which the M_2 tide was dominant.

Acoustic fluctuations have been shown to depend on the strength of the modulation, denoted by the modulation index, I_n^P. Range was shown to be one of the primary drivers of this index, as measured by the higher overtone structure of the fluctuation spectrum.

ACKNOWLEDGMENT

This work is supported by the Office of Naval Research, Program Element PE 62435N.

[1] L. A. Stalworth and M. J. Jacobson, "Acoustic propagation in an isospeed channel with uniform tidal current and depth change," J. Acoust. Soc. Am. **48**, 382–391 (1970).
[2] E. R. Franchi and M. J. Jacobson, "Effect of hydrodynamic variations on sound transmission across a geostrophic flow," J. Acoust. Soc. Am. **54**, 1302–1311 (1973).
[3] R. N. Baer and M. J. Jacobson, "Sound transmission in a channel with bilinear sound speed and environmental variations," J. Acoust. Soc. Am. **54**, 80–91 (1973).
[4] K. G. Hamilton, W. L. Siegmann, and M. J. Jacobson, "Combined influence of spatially uniform currents and tidally varying sound speed on acoustic propagation in the deep ocean," J. Acoust. Soc. Am. **62**, 53–62 (1977).
[5] R. P. Porter, "Acoustic probing of space-time scales in the ocean," in *Ocean Acoustics*, edited by J. A. DeSanto (Springer-Verlag, Berlin, 1979).
[6] R. L. Field, J. Newcomb, J. Showalter, J. George, and Z. Hallock, "Acoustic fluctuations and their harmonic structure," *Impact of Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. Pace and F. Jensen (Kluwer Academic, New York, 2002), Sec. 3, pp. 271–278.
[7] G. V. Frisk, *Ocean and Seabed Acoustics, A Theory of Wave Propagation* (Prentice–Hall, Englewood Cliffs, NJ, 1994).
[8] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972), p. 361.
[9] I. Tolstoy and C. S. Clay, *Ocean Acoustics* (American Institute of Physics, Melville, NY, 1987), Eq. (6.104).
[10] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists* (Academic, New York, 1995), Chap. 14.

The energetics and the stability of axially moving Kirchhoff strings (L)

Li-Qun Chen[a)]
Department of Engineering Mechanics, Shanghai Jiaotong University, Shanghai 200030, China

Wei-Jia Zhao
Department of Mathematics, Qingdao University, Qingdao 266071, China

(Received 15 April 2004; revised 11 August 2004; accepted 5 September 2004)

The energetics and the stability of free transverse vibration of an axially moving string are investigated based on the Kirchhoff nonlinear model. The model is derived in a physically meaningful manner. The time-rate of the total mechanical energy associated with the vibration is calculated to show that the energy is not conserved. It is proved that there exists a conserved quantity that remains a constant during the nonlinear vibration. The conserved quantity is applied to verify the Lyapunov stability of the axially moving Kirchhoff string. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1810310]

PACS numbers: 43.40.Cw [JGM] Pages: 55–58

I. INTROUCTION

Axially moving strings can represent many engineering devices such as power transmission belts, plastic films, magnetic tapes, paper sheets, and textile fibers.[1,2] Energetics and stability are of considerable interest in the study of axially moving strings.

The total mechanical energy associated with axially moving string is not constant when the strings travel between two supports. Chubachi first discussed periodicity of the energy transfer in an axially moving string.[3] Miranker analyzed energetics of an axially moving string and derived an expression for the time rate of change of the string energy.[4] Wickert and Mote pointed out that Miranker's expression represents the local rate of change only because it neglected the energy flux at the supports, and they presented the temporal variation of the total energy related to the local rate of change through the application of the one-dimensional transport theorem.[5] They also calculated the temporal variation of energy associated with modes of moving strings. Renshaw examined the change of the total mechanical energy of two prototypical winching problems, which provided strikingly different examples of energy flux at a fixed orifice of an axially moving system.[6] Lee and Mote presented a generalized treatment of energetics of axially moving strings.[7] Renshaw, Rahn, Wickert, and Mote defined a conserved quantity for axially moving strings.[8] Zhu and Ni investigated energetics of axially moving strings with arbitrarily varying length.[9]

Both energy and conserved quantities are potentially useful for analysis of stability, design of controllers, and development and validation of numerical algorithms. However, all aforementioned research on energy and conserved quantities for axially moving strings is confined to linear models. Kirchhoff proposed a nonlinear integro-differential equation for nontranslating strings, which considers the significant tension variation due to the transverse displacement.[10] Mote developed a nonlinear differential equation governing transverse vibration of axially moving strings.[11] Based on Mote's nonlinear model of axially moving strings, Chen and Zu analyzed the energetics and defined a conserved quantity.[12] However, there has been no research on energetics and conserved quantities for the Kirchhoff nonlinear strings. Besides, the stability of the straight equilibrium configuration of a string with subcritical axial speed has not been proved rigorously, although it is well known. To address the lack of research in these aspects, this paper deals with the energetics and the stability of Kirchhoff nonlinear strings.

In this work, the energetics and the stability of axially moving Kirchhoff strings are investigated. A physical interpretation of the Kirchhoff model is presented so that the total energy associated with an axially moving string can be defined accordingly. The time-rate of the energy change is calculated under given boundary conditions. A conserved quantity is constructed. As an application, the conserved quantity is used to prove that the straight equilibrium configuration of an axially moving string is stable in the Lyapunov sense on the condition that the axial speed is less than the critical speed.

II. ON KIRCHHOFF'S MODEL

Consider a uniform axially moving string of linear density ρ, cross-sectional area A, and initial tension P_0. The string travels at the constant and uniform axial transport speed V between two boundaries separated by distance L. The distance from the left boundary is measured by fixed axial coordinate x, and time is denoted by t. The transverse displacement of the string is given by the variable $u(x,t)$ in the sense that $u(x,t)$ describes the displacement of the string element instantaneously located at x even though different material elements occupy that position at different times.

The equation of motion in the transverse direction can be derived from Newton's second law,

$$\rho(u_{,tt}+2Vu_{,tx}+V^2u_{,xx})=[(P_0+P(x,t))u_{,x}]_{,x}, \qquad (1)$$

where a comma preceding x or t denotes partial differentiation with respect to x or t, and the dynamic tension $P(x,t)$, resulted from the transverse displacement, is connected with

the perturbed stress $\sigma(x,t)$ by the relationship

$$P(x,t)=A\sigma(x,t). \tag{2}$$

The Lagrangian strain

$$\varepsilon=\tfrac{1}{2}u_{,x}^2 \tag{3}$$

is used to account for geometric nonlinearity due to small but finite stretching of the string. For a linear elastic string, the perturbed stress is proportional to the Lagrangian strain through the constitutive relation

$$\sigma=E\varepsilon, \tag{4}$$

where E is Young's modulus.

Substitution of Eqs. (2)–(4) leads to Mote's model of axially moving string[11]

$$\rho(u_{,tt}+2Vu_{,tx}+V^2u_{,xx})-P_0u_{,xx}-\tfrac{3}{2}EAu_{,x}^2u_{,xx}=0. \tag{5}$$

If the spatial variation of the dynamic tension is rather small, then one can use the averaged value of the dynamic tension

$$\overline{P}(t)=\frac{1}{L}\int_0^L P(x,t)\mathrm{d}x \tag{6}$$

to approximate the exact value given by Eq. (2). In this case, the dynamic tension is a function of time alone, and is independent of the spatial coordinate. Substitution of Eqs. (2)–(4) and (6) into Eq. (1) yields the Kirchhoff nonlinear model of axially moving strings:

$$\rho(u_{,tt}+2Vu_{,tx}+V^2u_{,xx})-P_0u_{,xx}-\frac{EA}{2L}u_{,xx}\int_0^L u_{,x}^2\,\mathrm{d}x=0. \tag{7}$$

The Kirchhoff model (7) can be derived from Carrier's nonlinear model of strings[13] and its counterpart of axially moving strings.[14,15] The derivation is based on the integration of the axial strain and the longitudinal displacement from the equation of longitudinal motion under the quasi-static assumption, which can be found in the cases of non-translating strings[16] and axially translating beam.[17] That derivation seems more rigorous mathematically, while the present manner is more explicatory physically.

III. ENERGETICS

Consider the total mechanical energy in a specified spatial domain, the span $(0,L)$. The total mechanical energy consists of the kinetic energy of all material particles and the potential energy due to the initial tension and the disturbed tension caused by the transverse motion

$$\mathcal{E}=\int_0^L\left\{\frac{\rho}{2}[V^2+(u_{,t}+Vu_{,x})^2]+[P_0+P(t)]\varepsilon\right\}\mathrm{d}x, \tag{8}$$

where $u_{,t}+Vu_{,x}$ is the absolute velocity of the material particle at x in the transverse direction and includes both local and convective components. Based on Eqs. (2)–(4) and (6), the energy can be expressed by the transverse displacement as

$$\mathcal{E}=\int_0^L\left\{\frac{\rho}{2}[V^2+(u_{,t}+Vu_{,x})^2]+\frac{1}{2}\left(P_0+\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\right)u_{,x}^2\right\}\mathrm{d}x. \tag{9}$$

The time rate of change of the energy is

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t}=\int_0^L\frac{\partial}{\partial t}\left\{\frac{\rho}{2}[V^2+(u_{,t}+Vu_{,x})^2]+\frac{1}{2}\left(P_0+\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\right)u_{,x}^2\right\}\mathrm{d}x, \tag{10}$$

where the order of differentiation and integration can be interchanged as the limit of integration is time independent. Some mathematical manipulations cast Eq. (10) into the form

$$\begin{aligned}\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t}=&\int_0^L(u_{,t}+Vu_{,x})\left[\rho u_{,tt}+2\rho Vu_{,tx}+\left(\rho V^2-P_0-\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\right)u_{,xx}\right]\mathrm{d}x\\&+\left[\frac{1}{2}V\left(P_0-\rho V^2+\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\right)u_{,x}^2+u_{,t}\left[\left(P_0-\rho V^2+\frac{EA}{2L}u_{,x}\int_0^L u_{,x}^2\,\mathrm{d}x\right)u_{,x}-\frac{1}{2}\rho Vu_{,t}\right]\right]_0^L+\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\int_0^L u_{,x}u_{,xt}\,\mathrm{d}x.\end{aligned} \tag{11}$$

Consider the string constrained by the motionless end,

$$u(0,t)=0,\quad u(L,t)=0. \tag{12}$$

Then $u_{,t}=0$ at both ends. Substitution of Eqs. (7) and (12) into Eq. (11) gives the time-rate of energy change

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t}=\frac{1}{2}V\left(P_0-\rho V^2+\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\right)u_{,x}^2\Big|_0^L+\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\int_0^L u_{,x}u_{,xt}\,\mathrm{d}x. \tag{13}$$

Equation (13) indicates that the time-rate of energy change is generally not zero. Therefore, the energy of the string is not conserved.

IV. CONSERVED QUANTITY

Although the energy is not a constant in nonlinear vibration of axially moving strings, there does exist an alternative quantity that is conserved under fixed boundary conditions. For nonlinear transverse vibration of the axially moving string, governed by Eq. (7) and constrained by condition (12), the quantity defined by

$$I=\int_0^L\left[\frac{\rho}{2}u_{,t}^2+\frac{1}{2}(P_0-\rho V^2)u_{,x}^2+\frac{EA}{8L}u_{,x}^2\int_0^L u_{,x}^2\,\mathrm{d}x\right]\mathrm{d}x \tag{14}$$

is invariant during the vibration. Evaluation of the temporal differentiation

$$\frac{\mathrm{d}I}{\mathrm{d}t}=\int_0^L\frac{\partial}{\partial t}\left[\frac{\rho}{2}u_{,t}^2+\frac{1}{2}(P_0-\rho V^2)u_{,x}^2+\frac{EA}{8L}u_{,x}^2\int_0^L u_{,x}^2\,\mathrm{d}x\right]\mathrm{d}x \tag{15}$$

leads to

$$\frac{\mathrm{d}I}{\mathrm{d}t}=\int_0^L u_{,t}\left[\rho u_{,tt}+2\rho V u_{,tx}+\left(\rho V^2-P_0-\frac{EA}{2L}\int_0^L u_{,x}^2\,\mathrm{d}x\right)u_{,xx}\right]\mathrm{d}x+\left[u_{,t}\left(\frac{EA}{2L}u_{,x}\int_0^L u_{,x}^2\,\mathrm{d}x-\rho V u_{,t}\right)\right]_0^L. \tag{16}$$

Hence Eqs. (7) and (12) result in $\mathrm{d}I/\mathrm{d}t=0$. It should be pointed out that I is not conserved for moving boundary conditions, and I is invariant only for an axially moving string with two fixed ends. It should be remarked that the conserved quantity defined here is based on the same principle as its counterpart in the linear case.[8]

V. STABILITY

The conserved quantity will be used to demonstrate the stability of the straight equilibrium configuration of a string moving with subcritical axial speed. Actually, the following proposition will be proved.

For the solution $u(x,t)$ of Eq. (7) under the boundary condition (12) and the initial conditions

$$u(x,0)=a(x),\quad u_{,t}(x,0)=b(x), \tag{17}$$

if the axial speed is under the critical speed, namely,

$$V<\sqrt{\frac{P_0}{\rho}}, \tag{18}$$

then for $t\in[0,+\infty]$,

$$\|u(x,t)\|\leqslant\frac{5\rho L^2}{3(P_0-\rho V^2)}\|b(x)\|+\frac{5}{3}\|a(x)\|+\frac{5EA}{12(P_0-\rho V^2)L^2}\|a(x)\|^2, \tag{19}$$

where the norm for a function f on $[0,L]$ is defined by

$$\|f\|=\int_0^L\left[\frac{1}{L}f^2(x)+Lf'^2(x)\right]\mathrm{d}x. \tag{20}$$

In fact, the Holder inequality and the homogeneous boundary conditions give

$$\int_0^L u^2(t,x)\,\mathrm{d}x=\int_0^L\left[\int_0^x u_{,x}\,\mathrm{d}x\right]^2\mathrm{d}x\leqslant\int_0^L\left[\sqrt{x}\int_0^x u_{,x}^2\,\mathrm{d}x\right]\mathrm{d}x=\frac{2L^2}{3}\int_0^L u_{,x}^2\,\mathrm{d}x. \tag{21}$$

Notice that under the condition (18), $P_0-\rho v^2>0$. Equations (20), (21), and (14) result in

$$\|u(x,t)\|^2=\int_0^L\left[\frac{1}{L}u^2(x,t)+Lu_{,x}^2(x,t)\right]\mathrm{d}x\leqslant\frac{5L}{3}\int_0^L u_{,x}^2\,\mathrm{d}x\leqslant\frac{5L}{3}\frac{2}{P_0-\rho V^2}\int_0^L\left[\frac{\rho}{2}u_{,t}^2+\frac{1}{2}(P_0-\rho V^2)u_{,x}^2+\frac{EA}{8L}u_{,x}^2\int_0^L u_{,x}^2\,\mathrm{d}x\right]\mathrm{d}x=\frac{10LI}{3(P_0-\rho V^2)}. \tag{22}$$

As I is a constant during the vibration, its value can be calculated at the initial time

$$I=\int_0^L\left[\frac{\rho}{2}b^2(x)+\frac{1}{2}(P_0-\rho V^2)a'^2(x)+\frac{EA}{8L}a'^2(x)\int_0^L a'^2(x)\mathrm{d}x\right]\mathrm{d}x. \tag{23}$$

The definition of the norm leads to

$$\int_0^L\frac{\rho}{2}b^2(x)\mathrm{d}x\leqslant\frac{\rho L}{2}\|b(x)\|,$$

$$\int_0^L\frac{1}{2}(P_0-\rho V^2)a'^2(x)\mathrm{d}x\leqslant\frac{1}{2L}(P_0-\rho V^2)\|a(x)\|, \tag{24}$$

$$\int_0^L\frac{EA}{8L}a'^2(x)\int_0^L a'^2(x)\mathrm{d}x\,\mathrm{d}x=\frac{EA}{8L}\left[\int_0^L a'^2(x)\mathrm{d}x\right]^2\leqslant\frac{EA}{8L}\|a(x)\|^2.$$

Hence the conserved quantity (14) has the estimation

$$I\leqslant\frac{\rho L}{2}\|b(x)\|+\frac{1}{2L}(P_0-\rho V^2)\|a(x)\|+\frac{EA}{8L}\|a(x)\|^2. \tag{25}$$

Then inequality (19) holds as the result of inequalities (22) and (25).

Inequality (19) means that the resulting vibration of a small initial disturbance will be small. Hence, the straight equilibrium configuration of an axially moving Kirchhoff string is stable in the Lyapunov sense.

VI. CONCLUSIONS

The Kirchhoff model is formulated for axially moving strings. The derivation gives the physical interpretation of the model, as the dynamic tension is approximated by its spatial averaged value. The total mechanical energy is defined based on the interpretation, and is demonstrated not conserved by evaluating the time-rate of change. A conserved quantity is constructed for axially moving Kirchhoff nonlinear strings. The quantity is used to prove the Lyapunov stability of free transverse vibration.

ACKNOWLEDGMENTS

The research is supported by the National Natural Science Foundation of China (Project No. 10172056).

[a]Electronic mail: lqchen@online.sh.cn

[1] J. A. Wickert and C. D. Mote, Jr., ''Current Research on the Vibration and Stability of Axially Moving Materials,'' Shock Vib. Dig. **20**(5), 3–13 (1988).

[2] L. Q. Chen, ''Analysis and Control of Transverse Vibrations of Axially Moving Strings,'' Appl. Mech. Rev. (accepted).

[3] T. Chubachi, ''Lateral Vibration of Axially Moving Wire or Belt Materials,'' Bull. Jpn. Soc. Mech. Eng. **1**(1), 24–29 (1958).

[4] W. L. Miranker, ''The Wave Equation in a Medium in Motion,'' IBM J. Res. Dev. **4**(1), 36–42 (1960).

[5] J. A. Wickert and C. D. Mote, Jr., ''On the energetics of axially moving continua,'' J. Acoust. Soc. Am. **85**, 1365–1368 (1989).

[6] A. A. Renshaw, ''The Energetics of Winched Strings,'' J. Vibr. Acoust. **119**, 643–644 (1997).

[7] S. Y. Lee and C. D. Mote, Jr., ''A Generalized Treatment of the Energetics of Translating Continua, Part 1: Strings and Second Order Tensioned Pipes,'' J. Sound Vib. **204**, 717–734 (1997).

[8] A. A. Renshaw, C. D. Rahn, J. A. Wickert, and C. D. Mote, Jr., ''Energy and Conserved Functionals for Axially Moving Materials,'' J. Vibr. Acoust. **120**, 634–636 (1998).

[9] W. D. Zhu and J. Ni, ''Energetics and Stability of Translating Media with an Arbitrarily Varying Length,'' J. Vibr. Acoust. **122**, 295–304 (2000).

[10] G. Kirchhoff, *Lectures on Mathematical Physics* (in German) (Mechanik, Leipzig, Germany, 1876).

[11] C. D. Mote, Jr., ''On the Nonlinear Oscillation of an Axially Moving String,'' J. Appl. Mech. **33**, 463–464 (1966).

[12] L. Q. Chen and J. W. Zu, ''Energetics and Conserved Functional of Axially Moving Materials Undergoing Transverse Nonlinear Vibration,'' J. Vibr. Acoust. **126**, 452–455 (2004).

[13] G. F. Carrier, ''On the Nonlinear Vibration Problem of the Elastic String,'' Q. Appl. Math. **3**, 157–165 (1945).

[14] A. L. Thurman and C. D. Mote, Jr., ''Free, Periodic, Nonlinear Oscillation of an Axially Moving Strip,'' J. Appl. Mech. **36**, 83–91 (1969).

[15] H. Koivurova and E. M. Salonen, ''Comments on Nonlinear Formulations for Travelling String and Beam Problems,'' J. Sound Vib. **225**, 845–856 (1999).

[16] A. H. Nayfeh and D. T. Mook, *Nonlinear Oscillations* (Wiley, New York, 1979).

[17] J. A. Wickert, ''Non-linear Vibration of a Traveling Tensioned beam,'' Int. J. Non-Linear Mech. **27**, 503–517 (1992).

Discrimination of first- and second-order regular intervals from random intervals as a function of high-pass filter cutoff frequency (L)

William A. Yost,[a] Dan Mapes-Riordan, Raymond Dye, Stanley Sheft, and William Shofner

Parmly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

(Received 1 July 2004; revised 18 October 2004; accepted 19 October 2004)

This study extends the work of Kaernbach and Demany [J. Acoust. Soc. Am. **104**, 2998–2306 (1998)] in which regular interval stimuli (RIS) click trains with first-order intervals could be discriminated from random-interval click trains, but RIS with second-order intervals could not. Kaernbach and Demany concluded that their results cast doubt on autocorrelation as a method of analysis for such stimuli. The present study investigated the same stimuli, but for a variety of filter conditions. The results suggest that while RIS click trains with first-order intervals are more easily discriminated from random-interval stimuli than second-order interval RIS click trains, discrimination based on second-order intervals is possible except when the stimuli are high-pass filtered above 8 kHz, i.e., above the spectral region of phase locking. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1830671]

PACS numbers: 43.66 Ba, 43.66 Hg, 43.66 Mk [GDK] Pages: 59–62

I. INTRODUCTION

Kaernbach and Demany (1998) originally, and then Kaernbach and Bering (2001) and Pressnitzer *et al.* (2002, 2004), generated a set of regular interval stimuli (RIS) consisting of click trains such that the interclick interval (ICI) of the click trains contained either a constant interval between successive pulses in the click trains (first-order intervals) or constant intervals between every other pulse (second-order intervals). The long-term spectra and autocorrelation functions are similar for both first-order and second-order interval RIS click trains, especially in regard to the autocorrelation features that have been used to account for the perception of RIS sounds (Yost, 1996, 1997). By high-pass filtering the RIS click trains above 6 kHz, Kaernbach and Demany (1998) eliminated the ability of listeners to use resolved harmonics, forcing the listeners to use temporal cues such as could be modeled by autocorrelation. Their data showed that only first-order RIS click trains could be processed by the listeners. They argued that these results were inconsistent with an autocorrelation approach for processing such stimuli. Pressnitzer *et al.* (2002, 2004) showed that RIS click trains very similar to the ones used by Karenbach and Demany (1998), high-pass filtered at 3 kHz with second-order intervals, appear to produce a pitch, and the pitch appears to be associated with the most common interspike interval of single units in the ventral cochlear nucleus. While their data do not fully support an autocorrelation approach for processing these RIS click stimuli, they did show in their 2004 paper that second-order interval stimuli can produce a pitch.

What has not been determined is how the discriminability of these RIS click trains, as originally studied by Kaernbach and Demany (1998), varies as a function of filtering. Pressnitzer *et al.* (2002, 2004) implied that the pitch strength of the second-order interval click trains was less salient than that of the first-order interval click trains. However, they did not measure pitch strength. Discrimination experiments like some of those performed by Kaernbach and Demany (1998) and Kaernbach and Bering (2001) have been used to indirectly measure pitch strength (Yost, 1997). The 6-kHz high-pass filter used by Kaernbach and Demany (1998) places the temporal information in RIS click trains above the spectral region where phase locking is assumed to operate. Kaernbach and Bering (2001) had one condition in which the click-train stimuli were high-pass filtered at 2 kHz. In this condition the second-order interval stimuli were discriminable from their noise foils at only about 64%, while the first-order intervals were discriminable at about 95%. The present study used similar conditions to those employed by Kaernbach and Demany (1998) to investigate RIS click-train discrimination in several spectral regions where the spectral ripples of the RIS click trains are likely unresolved, but where phase locking is probable (i.e., below approximately 6 kHz; see for instance Johnson, 1980). As such, the RIS click-train discrimination experiments can determine if the pitch strength of second-order interval click trains is different from that of first-order interval click trains in spectral regions in which phase locking is possible.

II. STIMULI

Click-train RIS sounds are a sequence of ICI values, each value representing the duration between the prior and current click. Random-interval click trains were created using the same method employed by Kaernbach and Demany (1998) by randomly assigning each ICI a duration from a uniform distribution between zero and twice the mean ICI (d ms) with a resolution of 25 μ s. No more than two consecu-

[a]Electronic mail: wyost@luc.edu

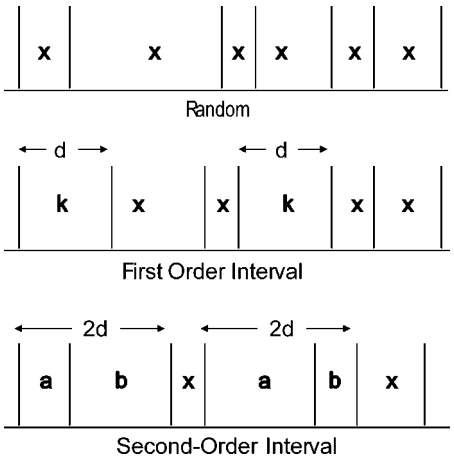


FIG. 1. RIS click trains; top: random intervals (x), middle: first-order intervals (kxx), bottom: second-order intervals (abx), d is mean duration.

tive ICIs could be larger or smaller than d, in order to prevent the occurrence of long sequences of short or long duration ICIs which could result in a timbre cue.

First-order ICI sequences were created from random ICI sequences by replacing every third random ICI with a regular interval equal to d. These ICI sequences have been termed ''kxx'' by Kaernbach and Demany (1998), where k represents the regular interval of duration d and x represents the random intervals.

Second-order ICI sequences were created from random ICI sequences by replacing the first two of every three random ICIs with ICIs whose sum is 2d. The duration of the first of the three intervals, denoted a, was randomly drawn from a uniform distribution between zero and 2d. The duration of the second interval, denoted b, was equal to $2d-a$. Thus, the sequences contain second-order regular intervals of duration $a+b=2d$. These sequences have been termed ''abx'' by Kaernbach and Demany. Figure 1 provides a schematic diagram of a random-order interval click train and its kxx and abx versions. Both the kxx and abx stimuli produce a very similar pitch sensation when the pitch can be detected, and the sensation has been labeled ''rattle'' pitch by Kaernbach and Demany (1998).

The kxx and abx stimuli have similar, but not identical, spectra and autocorrelation functions. The spectra have spectral peaks space at 1/d (for kxx) and 1/2d (for abx). The normalized autocorrelation functions have a major peak (AC1) at d (for kxx) or 2d (for abx), and the height of AC1 is 0.33 for both kxx and abx stimuli. These spectral and autocorrelation function features are those used by most spectral and temporal models of pitch processing. Autocorrelation models (see Yost, 1996, 1997) predict that the pitch of the RIS click trains is associated with the reciprocal of d or 2d, and the strength of the pitch is determined by the height of AC1. These models would predict that the pitch strength of the kxx and abx RIS click trains would be the same. Pressnitzer *et al.* (2004) point out that the autocorrelation functions of their random-interval click trains also have a very small broad peak near lag d, and their data suggest that their random-interval click trains have a very weak pitch.

The amplitudes of the clicks were identical and were adjusted to create a click train with a spectrum level of 30 dB. Each stimulus was temporally windowed using a 5-ms squared cosine function. The duration of d was 2, 4, or 8 ms for kxx stimuli and 1, 2, and 4 ms for abx stimuli, resulting in fundamental frequencies, f_0, of 500, 250, or 125 Hz (f_0 is the reciprocal of d for kxx stimuli and the reciprocal of 2d for abx stimuli). The total stimulus duration was kept constant at 480 ms. Stimuli were presented in a wideband condition (stimuli low-pass filtered at 15 kHz) and in several high-pass-filter conditions. The high-pass-filter cutoff frequencies were 8 and 16 times f_0 (all stimuli were low-pass filter at 15 kHz). In the high-pass-filter conditions, a low-pass-filtered Gaussian noise was added to mask distortion components that could be used as discrimination cues. The low-pass-filter cutoff frequency was one-half octave below the high-pass cutoff frequency applied to the click train. The spectrum level of the masking noise was 20 dB; 10 dB below the spectrum level of the click-train stimuli. The relationship between interval order, f_0, d, and high-pass cutoff frequency is summarized in Table I.

III. PROCEDURE AND LISTENERS

Discrimination performance was measured using a cued, two-alternative, forced-choice (2AFC) procedure. Each trial consisted of a random-interval click train, followed by two other stimuli presented in random order, a random-interval click train and one of the regular-interval click trains. A 300-ms silent gap was inserted between each stimulus. Listeners were asked whether the second or third stimulus was different from the other two stimuli (i.e., to pick the regular-interval click train). Eighteen conditions were tested. A block of trials consisted of the three filter conditions (wideband, $8f_0$, $16f_0$) at a single order (first-kxx or second-abx) and at a single fundamental frequency (125, 250, or 500 Hz), that is, one row in Table I. For each f_0 and interval order, 40, wideband trials were followed by 40, $8f_0$ trials, which were then followed by 40, $16f_0$ trials for a total of 120 trials per block. This was done to provide listeners a clear impression of the pitch in the initial wideband stimulus trials in an attempt to maximize performance in the high-pass-filtered conditions. Not counting practice blocks, a block of 120 trials was repeated ten times throughout the course of the experiment, resulting in 400 trials per condition for each listener.

Four normal-hearing listeners participated in this experiment, including one of the authors (DM). The first block of trials for each listener was considered practice and was not included in the final results.

IV. RESULTS

The mean results across the four listeners are shown in Fig. 2 as percent-correct responses per stimulus condition. With a couple of exceptions the mean results are representative of individual listener performance. One listener had difficulty in all high-pass-filter conditions except the first-order, $d=2$-ms conditions. This poor performance is in contrast to this listener's relatively good wideband performance. Another listener (one of the authors, DM) was the only partici-

TABLE I. Summary of test conditions.

Interval order	Fundamental frequency, f_0 (Hz)	Mean interval duration, d (ms)	High-pass cutoff frequency (Hz)		
			Wideband	$8f_0$	$16f_0$
1-kxx	125	8	···	1000	2000
1-kxx	250	4	···	2000	4000
1-kxx	500	2	···	4000	8000
2-abx	125	4	···	1000	2000
2-abx	250	2	···	2000	4000
2-abx	500	1	···	4000	8000

pant who consistently discriminated [$P(C)>75\%$] all of the high-pass filtered, second-order stimuli except the 8-kHz cutoff frequency condition where his performance was at chance consistent with the performance of the other listeners. As expected, wideband conditions provided an upper bound to discrimination performance with decreasing discrimination performance with increasing high-pass-filter cutoff frequency. Except for the 8-kHz high-pass conditions, the average listener could discriminate the kxx and the abx stimuli from their respective noise foils, but performance for the abx discriminations was always poorer than for the kxx conditions. Also, average performance decreased with decreasing fundamental frequency across all high-pass-filtered conditions.

V. DISCUSSION AND CONCLUSION

High-pass-filtered, second-order regular versus random-interval discrimination is possible, yet is more difficult than first-order discrimination except for the 8-kHz cutoff-frequency condition, where second-order discrimination is impossible. This implies that the pitch strength of first-order interval RIS click trains is more salient than that of second-order interval RIS click trains. While we used the exact method described by Kaernbach and Demany (1998) and Kaernbach and Bering (2001) to generate the click trains, we did not use the exact same psychophysical procedures, nor were all of the stimulus conditions exactly similar. Our procedure and stimulus conditions are most like the one used in experiments 3 in both of the papers by Kaernbach and colleagues. Even so, there are small differences among the studies including stimulus duration; listener training; number of random, x, intervals following the fixed, k, interval; d; and details of the low-pass noise used to mask distortion products.

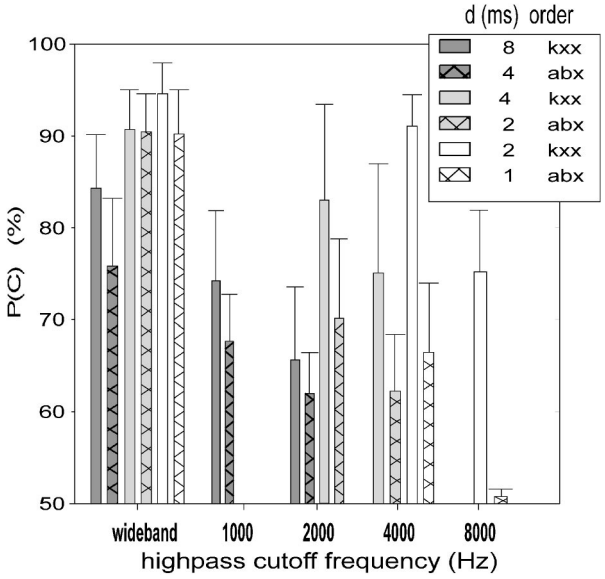


FIG. 2. Mean results (across four listeners) showing percent correct, $P(C)$, discrimination as a function of high-pass filter cutoff (in Hz) for random versus first-order (kxx) and second-order (abx) RIS click-train discrimination; bars are mean percent correct; error bars are standard errors. $P(C)=50\%$ is chance performance.

Our results at 8 kHz are consistent with those of Kaernbach and Demany (1998), but the results at all other cutoff frequencies extend their findings in that performance was above chance for the discrimination of second-order intervals (abx) from their random-interval foils. Our listeners in the kxx conditions did not seem to be as sensitive as the subjects in either of the Kaernbach studies, but were slightly more sensitive in the comparable abx condition in the Kaernbach and Bering (2001) study. These small performance differences are probably accounted for by individual differences and/or by the small procedural and stimulus differences among the studies. The current results suggest that the auditory system may be sensitive to second-order intervals as long as they occur below 6 kHz, where phase locking is possible.

Kaernbach and Demany (1998) used their results as evidence against autocorrelation as a means of explaining auditory temporal processing. To the extent that listeners can process second-order interval statistics, the current results suggest that processes like autocorrelation may be useful for explaining temporal processing phenomena for stimuli that contain temporal information in spectral regions below about 6 kHz, where phase locking occurs. However, autocorrelation functions for the kxx and abx stimuli are not identical in all respects, and there is some weak temporal regularity in the random-interval click trains used in these experiments (Pressnitzer *et al.*, 2004). That is, while the AC1 peak is at d or 2d and its normalized height is one-third (these are the features of autocorrelation used to account for pitch and pitch strength of RIS sounds; Yost, 1996, 1997) for the kxx and abx stimuli, other more subtle features of the functions differ (see Pressnitzer *et al.*, 2004). Thus, the ability for listeners to discriminate abx from the random-interval click trains may be based on some other aspects of the stimulus fine structure than those typically used to account for RIS

processing. A hint of this possibility can be seen for the wideband stimuli of Fig. 2. For most conditions performance for the abx stimuli is still poorer than for the kxx stimuli, even when many of the spectral components are resolved. Since these spectral components are almost the same for both kxx and abx stimuli, it might be that whatever temporal feature allows the abx stimuli to be discriminated from the random-interval click trains when the spectral features are unresolved is also used (at least to some extent) when the spectral features are resolved.

The current uses of autocorrelation models predict that the discriminability of the first- and second-order RIS click trains from their random-interval noise foils would be similar, and therefore that they would have similar pitch strengths. Since this did not occur, the current use of autocorrelation cannot be used to completely account for these data. Pressnitzer *et al.* (2004) attempted to account for the perceived pitch of these stimuli, but not pitch strength. It might be possible to use the strength of the interval statistics obtained in their work to derive predictions for pitch strength, but the current data and literature do not provide enough information to make quantitative predictions at this time. Thus, existing explanations of the pitch and pitch strength of RIS stimuli will need modification or new models will be required in order to account for the results of these RIS click-train experiments, but such models will need to account for the pitch and pitch strength of second-order, as well as first-order interval RIS sounds.

ACKNOWLEDGMENTS

This research was supported by a program project grant from the National Institutes of Deafness and other Communication Disorders, and was part of Dan Mapes-Riordan's Ph.D. dissertation.

Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.

Kaernbach, C., and Bering, C. (**2001**). ''Exploring the temporal mechanisms involved in the pitch of unresolved harmonics,'' J. Acoust. Soc. Am. **110**, 1039–1048.

Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of auditory temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.

Pressnitzer, D., de Cheveigné, A., and Winter, I. M. (**2002**). ''Perceptual pitch shift for sounds with similar waveform autocorrelation,'' ARLO **3**, 1–6.

Pressnitzer, D., de Cheveigné, A., and Winter, I. M. (**2004**). ''Physiological correlates of the perceptual pitch shift for sounds with similar autocorrelation,'' ARLO **5**, 1–6.

Yost, W. A. (**1996**). ''Pitch of iterated rippled noise,'' J. Acoust. Soc. Am. **100**, 511–519.

Yost, W. A. (**1997**). ''Pitch strength of iterated rippled noise when the pitch is ambiguous,'' J. Acoust. Soc. Am. **101**, 1644–1650.

Predicting the eigenmodes of a cavity containing an array of circular pipes

Benjamin Fenech[a] and Finn Jacobsen[b]
Acoustic Technology, Ørsted·DTU, Technical University of Denmark, Ørsted Plads, Building 352, DK-2800 Lyngby, Denmark

(Received 7 March 2004; revised 26 July 2004; accepted 10 September 2004)

An array of pipes inside a cavity, as found, for example, in a shell-and-tube heat exchanger, changes the eigenfrequencies of the cavity. It can be tedious to determine the shifted eigenfrequencies with a finite-element model. Based on previous work by Meyer and Neumann, Parker proposed a simple relationship for predicting the shifted eigenfrequencies. In this paper, results obtained from this relationship are compared with eigenfrequencies obtained from very accurate finite element simulations. From the results it can be concluded that Parker's relationship gives fairly good predictions of the eigenfrequencies for the first few modes in a cavity with pipes arranged in a rectangular configuration. The predictions are not so accurate for pipes arranged in a diamond configuration, and a modified version of the relationship is suggested for this configuration. If the number of pipes in the cavity is small, the simple relationship is no longer valid. [DOI: 10.1121/1.1836882]

PACS numbers: 43.20.Ks, 43.58.Ta [MO]. Pages: 63–67

I. INTRODUCTION

A shell-and-tube heat exchanger with fluid flow past an array of pipes can be a powerful source of noise and vibration.[1,2] A common excitation mechanism in such a situation is vortex shedding of the fluid flow,[3] caused by, for example, heat fins on the pipes. The excitation interacts with the natural acoustic modes of the shell enclosing the tubes and the vibration modes of the pipes, thus producing an acoustic and vibrational response. One of the tasks of the noise control engineer is to calculate these natural modes, and compare them with measurements in order to identify the dominant modes. In this work only the acoustic eigenmodes of the shell (or cavity) enclosing the tubes are considered.

Calculating the eigenfrequencies of a cavity is a simple task: either analytically for simple geometries, or using a finite element model for more complex geometries. However, in a shell-and-tube heat exchanger, the tubes are part of the system, and they alter the eigenvalue problem. This phenomenon has been described analytically[4,5] for single obstructing objects in a cavity, in terms of scattering and diffraction effects. However, not only are the resulting integral/differential equations of little use in practice; but the problem gets considerably more complicated when a large number of objects are considered, among other things because of multiple scattering effects.[4,5]

To model the complete heat exchanger (including the array of tubes) numerically can be a tedious task, due to the large number of small pipes, even though the usual regular pattern of the pipes may speed up the process. More importantly, the small diameter of the pipes (compared to the outer dimensions of the shell) limits the maximum element size that can be used. This will often lead to a very large number of elements, and a corresponding demand on computational time and power.

An alternative could be to model the shell without the tubes, while adjusting some parameter to take account of the effect of the tubes. Parker has proposed a very simple expression in terms of an effective speed of sound,[6] referring to the concept of added (virtual) mass as used by Meyer and Neumann.[7] The underlying reasoning of this model is not explained very clearly. However, both Parker[6] and Blevins[2] have presented experimental results that confirm this relationship.

Our purpose in this paper is to present some numerical simulations that have been performed in order to test Parker's expression and get an idea of its accuracy and limitations. To simplify the calculations, the problem has been reduced to two dimensions by taking a perpendicular cross section along an inline shell-and-tube heat exchanger, such that the acoustic velocity is in the plane perpendicular to the pipes' axes. Moreover, no external fluid flow is considered.

II. THE EFFECTIVE SPEED OF SOUND

The concept of an effective speed of sound was originally proposed by Meyer and Neumann, who developed a relationship to predict the reduced speed of sound in an ''acoustic lens.''[7] They stated that an alternating flow around an array of fixed, rigid objects has an increased inertia, in the same way as an oscillating object in a fluid has an added (virtual) mass. The added mass or inertia depends in both cases on the shape of the rigid object; for a circular cylinder in unrestricted two-dimensional (2-D) flow, the added mass equals the mass of the fluid displaced by the cylinder. For an array of circular pipes, the relationship becomes[6]

[a]Current address: Aerodynamics & Flight Mechanics Research Group, Aeronautics and Astronautics Department, School of Engineering Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom. Electronic mail: bfenech@gmail.com
[b]Electronic mail: fja@oersted.dtu.dk

$$\frac{c_e}{c} = \frac{1}{\sqrt{1+\sigma}}, \tag{1}$$

where c_e is the effective speed of sound in the cavity with the pipes, c is the speed of sound, and σ is the volume fraction, i.e., the percentage of the volume occupied by the pipes. Note that this relationship is only valid for pipes of circular cross section. In formulating this relationship, it was assumed that the compressibility of the fluid does not change when an array of acoustically hard objects is introduced. This assumption is questionable, especially when applied to modes in a cavity, where an object can sit in a region of no flow (pressure maxima). In this case, it is more likely that the presence of the object increases the stiffness, rather than the mass of the fluid, as suggested by Morse and Ingard.[4] Another factor not taken into account is that, since the added mass depends on the flow pattern, it is affected by other nearby boundaries (Massey[8]).

III. RESULTS

The simplified problem has been modeled as a number of circular bodies in a rectangular cavity. All the boundaries were treated as rigid. The numerical calculations were done using the eigenvalue solver of MSC Actran, a commercial finite element package. This solver uses a block Lanczos method with a Sturm sequence check, and uses Actran's sparse solver as an internal linear solver.[9] Triangular quadratic elements were used. A number of pretests showed that an element size equal to the radius of the pipes gave sufficiently accurate results, even though geometrically the modeled bodies were not perfectly circular. With such an element size, the number of elements per wavelength was at least five in the frequency range of interest. All post-processing was done using Matlab.

Two different, widely used pipe configurations were examined: a rectangular and a diamond configuration; see Fig. 1. The geometries shown are fixed in such a way as to reduce the number of variables of the problem. The simulations were done in terms of four nondimensional parameters: the volume fraction (σ), the cavity aspect ratio (a/b), the largest number of pipes in the x direction (n_x), and the normalized eigenfrequencies of the system (fa/c).

A. Number of pipes

Equation (1) applies for an array of pipes, and should not depend on the number of objects in the array. This was checked in the first simulation. The cavity aspect ratio a/b, pipe spacing d, and volume fraction σ were all fixed, while the number of pipes in the cavity was varied. This was done for the two different pipe configurations, and for three different volume fractions: 3%, 12.5%, and 50%. For a fixed volume fraction, the cavity dimensions increase as the number of pipes increases. Figure 2 shows a plot for a rectangular configuration of pipes with a volume fraction of approximately 50%. Each curve represents a particular mode of the first 50 modes, sorted by increasing eigenfrequencies.

The plot can be divided into two regions: one where the eigenfrequencies depend on the number of pipes in the cavity, and another one where the normalized eigenfrequencies are almost independent of the number of pipes in the cavity. The division occurs between $n_x=10$ and $n_x=15$. The same trend has been observed for a diamond configuration of pipes (not shown). For smaller volume fractions, the dependence of the eigenfrequencies is less pronounced; for $\sigma=3\%$, practically all the modes are represented by horizontal lines along the whole abscissa.

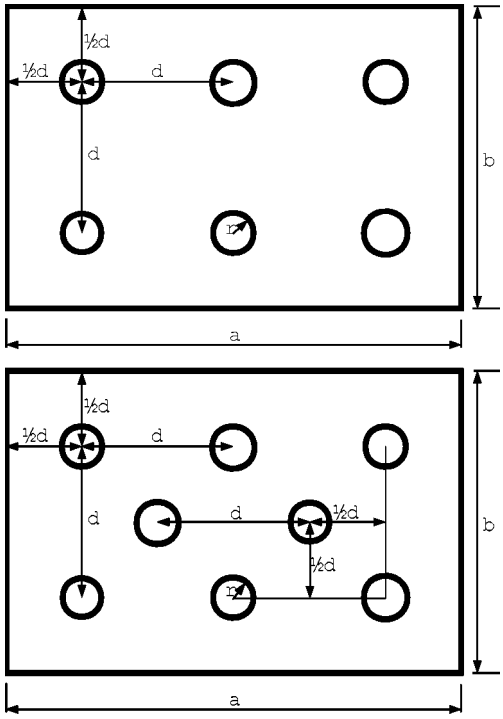


FIG. 1. Cross section of the heat exchanger. (a) Rectangular configuration. (b) Diamond configuration.

B. Volume fraction

The horizontal curves shown in Fig. 2 for a large number of pipes suggest the existence of a simple relationship for

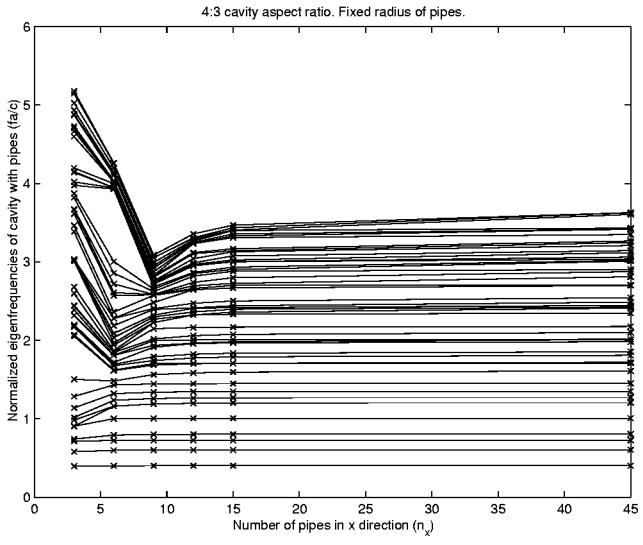


FIG. 2. Effect of the number of pipes on the modes of a cavity. Rectangular configuration. The volume fraction was approximately 50%.

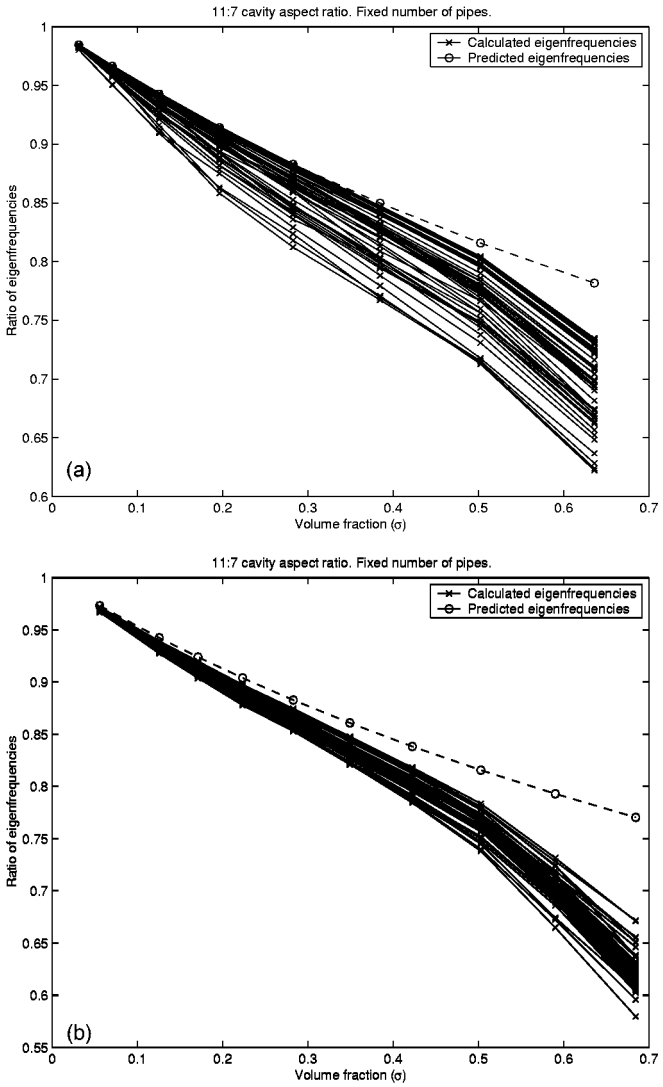


FIG. 3. The ratio of eigenfrequencies as a function of the volume fraction: calculated (ACTRAN) and predicted [using Eq. (2)]. (a) Rectangular configuration. (b) Diamond configuration.

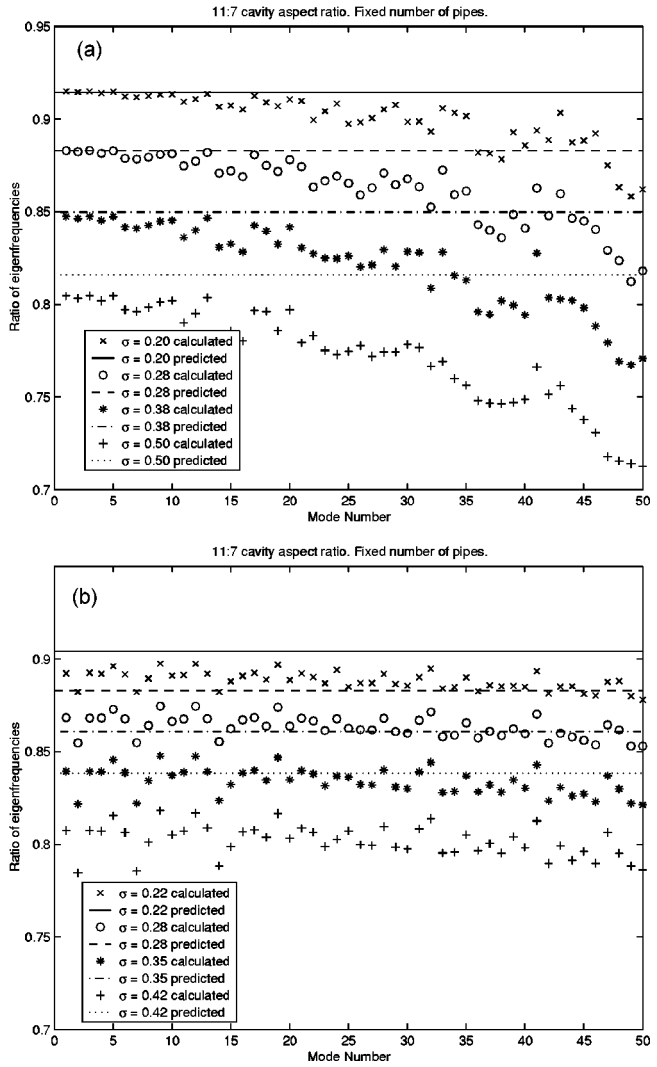


FIG. 4. The ratio of eigenfrequencies as a function of the mode number: calculated (ACTRAN) and predicted [using Eq. (2)]. (a) Rectangular configuration. (b) Diamond configuration.

predicting the shifting of eigenfrequencies. An additional set of simulations were thus carried out to test the accuracy of Eq. (1). For these simulations, the number of pipes in the cavity was kept constant, whereas the volume fraction was varied. A similar, but slightly modified cavity aspect ratio of 11:7 was used in the subsequent tests to avoid modal degeneracy for the first 50 modes.

Equation (1) can be rewritten in terms of the eigenfrequencies of the cavity with and without the pipes,

$$\frac{f_{m,n\text{pipes}}}{f_{m,n\text{nopipes}}} = \frac{1}{\sqrt{1+\sigma}}. \tag{2}$$

The eigenfrequencies of the cavity without pipes, $f_{m,n\text{nopipes}}$, can be calculated from

$$f_{m,n\text{nopipes}} = \frac{c}{2}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}, \quad m,n = 0,1,2,3,\ldots. \tag{3}$$

A significant number of eigenfrequencies was generated using Eq. (3), and sorted in ascending order, and the first 50 were compared with the eigenfrequencies obtained from the numerical simulations to obtain the ratio $f_{m,n\text{pipes}}/f_{m,n\text{nopipes}}$. Eight different volume fractions were tested for both a rectangular and a diamond pipe configuration; see Figs. 3(a) and 3(b). The figures also show the curve corresponding to Eq. (2).

According to the numerical simulations, an array of pipes decreases the eigenfrequencies of a cavity, and the effect increases as the volume fraction occupied by the pipes grows. For a rectangular array of pipes with a volume fraction of up to approximately 35%, the relationship suggested by Parker compares fairly well with the upper curves in the plot. These curves correspond to the first few modes (this can be verified in Fig. 4, which is described in the next paragraph). For higher volume fractions, the eigenfrequencies calculated numerically are somewhat lower than predicted. The relationship fails to predict the behavior of higher-order modes, which deviate gradually from the behavior of the first few modes. A diamond configuration of pipes shows the same trends, except that the relationship is less accurate in predicting the shifting of the eigenfrequencies, even for small volume fractions.

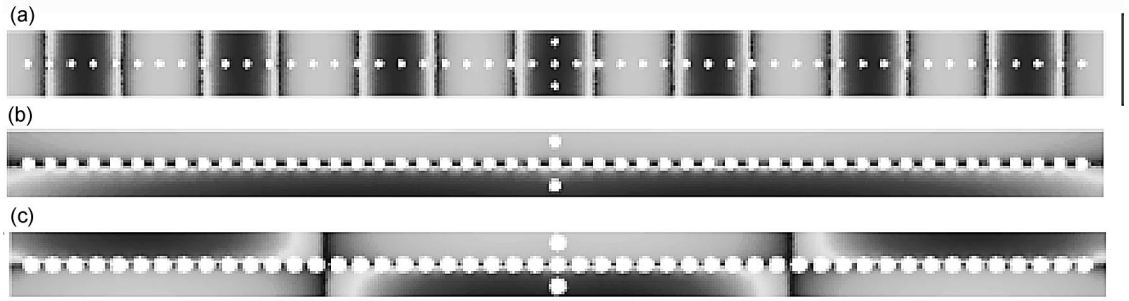


FIG. 5. Mode order shifting. The 14th mode (modes sorted by increasing eigenfrequencies) in the same cavity, with the same configuration of pipes. The mode shape is different for different radii. (a) Radius=0.02 m. Mode (14,0). (b) Radius=0.03 m. Mode (0,1). (c) Radius=0.04 m. Mode (2,1).

The accuracy of the predictions for higher-order modes is illustrated in Fig. 4, which shows the ratio of eigenfrequencies as a function of the mode number for four different volume fractions; those are considered to be most relevant in practice. For a given volume fraction, the predicted frequency ratio is represented by a horizontal curve. For a rectangular configuration of pipes with volume fractions of 20%, 28%, and 38%, the predicted frequency ratio compares very well with the numerical calculations for the first ten modes. For higher-order modes, the deviations from predicted values are clearly observable. The plot for a diamond configuration of pipes differs in two aspects: first, the predictions are less accurate even for the first few modes at small volume fractions; second, the calculated eigenfrequencies do not ''roll off'' appreciably for higher-order modes. A further characteristic, which is present in both plots, is that different modes are grouped together, and these patterns persist for different volume fractions. This suggests that the shift of eigenfrequencies is also a function of the mode shape, as stated, for example, by Morse and Ingard.[4]

IV. DISCUSSION

From Fig. 2 it is clear that Eq. (2) is not valid for the general case of a number of rigid objects in a rigid cavity. For a small number of pipes, the shift in eigenfrequencies is dependent on the number of pipes. This dependence can be explained by referring to a simplified approximate relationship for the shift of eigenfrequencies by a rigid scattering object developed by Morse and Ingard,[4]

$$K_N^2 \approx \eta_N^2 - \frac{k^2\sigma}{\Lambda_N}[\langle \rho^2 c^2 U_N^2\rangle - \langle \phi_N^2\rangle], \tag{4}$$

where K_N^2 are the eigenfrequencies of the cavity with the scattering object, η_N^2 are the eigenfrequencies of the empty cavity, $\langle U_N^2\rangle$ is the spatially averaged mean square velocity amplitude, $\langle \phi_N^2\rangle$ is the spatially averaged mean square pressure amplitude, and Λ_N is a normalization constant. The ''new'' eigenfrequencies depend on two correction terms: one decreasing and the other one increasing the ''old'' eigenfrequencies, depending on where the object is sitting, i.e., near a velocity or a pressure node. This prediction has been validated by modeling a cavity with a single line of pipes: the eigenfrequencies were increased when the pipes sat on a velocity node, and decreased when the pipes were on a pressure node. The fact that this up–down behavior breaks down for a large number of pipes may be explained by looking at the mode shapes: the pipes are almost always sitting on a pressure node, which makes the second correction term negligible. A line of pipes seems to either ''pull'' a nodal line, or force some particular modes (having nodal lines along the existing line of pipes) to occur at a much lower frequency, thus changing the mode ordering completely. This effect can be seen in Figs. 5(a)–5(c), which show the 14th mode in a particular cavity having the same configuration but with different pipe diameters. For the smallest diameter [Fig. 5(a)], the 14th mode corresponds to (14,0), the same as one would predict for an empty cavity. However, if the diameter is increased by a factor of 1.5, shown in Fig. 5(b), mode (0,1) occurs at a lower frequency than mode (14,0) and becomes the 14th mode. When the initial diameter is doubled, even mode (2,1) occurs at a lower frequency than mode (14,0), as shown in Fig. 5(c).

For a rectangular array of pipes, Eq. (2) gives better predictions for the first few modes than for higher-order modes. This is, in general, acceptable, since, in practice, the first modes are usually the most troublesome. However, the same relationship is not good for a diamond configuration. In formulating Eq. (1), it was assumed that adjacent pipes do not affect the flow pattern, so that a simple relationship for the added mass could be used. In a diamond configuration, the effect of adjacent pipes on the flow pattern might be more significant than in a rectangular configuration. If this is the case, a corrected value of the added mass for pipes arranged in a diamond configuration can yield a better prediction expression than Eq. (2). Such a correction can be determined either experimentally or perhaps analytically. In this project, a simpler approach was taken. Referring to Fig. 3, Eq. (2) gives an underestimation of the ratio of eigenfrequencies. Thus, as an alternative relationship, the reciprocal with a changed sign was considered, i.e., $f_{m,n\text{pipes}}/f_{m,n\text{nopipes}} = \sqrt{1-\sigma}$. It turned out that, for a diamond configuration of pipes, the data fell in between the two relationships. It was concluded that, for such a configuration, a weighted combination of the two equations,

$$\frac{f_{m,n\text{pipes}}}{f_{m,n\text{nopipes}}} = \frac{2\times\sqrt{1-\sigma}}{5} + \frac{3}{5\times\sqrt{1+\sigma}}, \tag{5}$$

gives more accurate predictions than Eq. (2). This result is a purely empirical one. Figure 6 shows the improved predic-

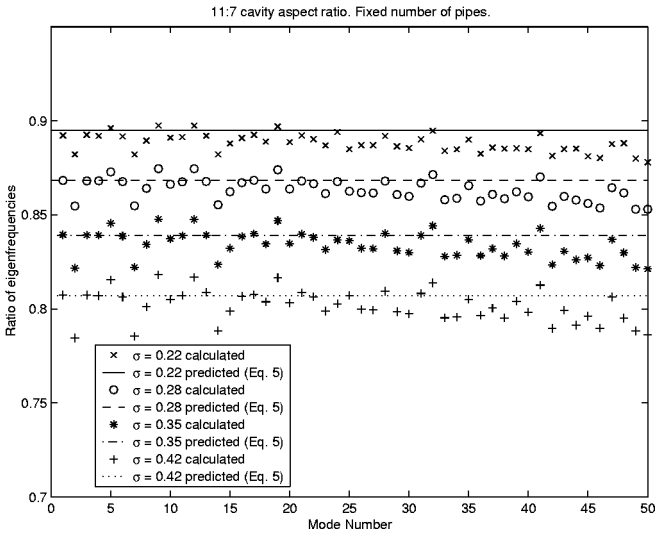


FIG. 6. Predictions using Eq. (5) compared to ACTRAN simulations, for a diamond configuration of pipes.

tions of eigenvalues for a cavity with a diamond configuration using Eq. (5) [compare with Fig. 4(b)].

In both configurations, higher-order modes deviate from the behavior of the first few modes. This is more significant for the rectangular configuration. The most likely reason for this is the change in mode ordering. Whenever a change in the mode ordering occurs, the ratio of eigenfrequencies $f_{m,n\text{pipes}}/f_{m,n\text{nopipes}}$ becomes invalid. Furthermore, even if the ratio is corrected so that the two frequencies correspond to the same mode shape, it still deviates considerably from the predicted ratio. Further tests have shown that changes in mode ordering are less likely to occur with a diamond configuration, thus explaining the collapsed curves for this configuration in Fig. 3.

Further simulations have been carried out, including cavities with irregularly spaced pipes, and tests for a cavity having an aspect ratio of 37:5. The results were similar to those presented here, and are therefore not included.

V. CONCLUSIONS

An array of acoustically rigid pipes affects the eigenfunctions and eigenvalues of a cavity. For a sufficiently large number of pipes, the eigenfrequencies are decreased, provided that the pipes are not intentionally placed at velocity nodes. The reduction depends on the volume fraction occupied by the pipes. Fairly good predictions of the magnitude of these reductions are possible using a very simple relationship having only one variable: the volume fraction. This relationship is reasonably accurate for the first few modes (say the first ten, sorted by increasing eigenfrequencies), for a rectangular configuration of pipes having volume fractions up to around 35%. With a diamond configuration of pipes, more accurate results can be obtained by using a slightly different, empirically derived relationship.

For higher-order modes, the fact that the order of the modes is likely to change means that the validity of the simple expression breaks down. This drawback is not so pronounced for a diamond configuration of pipes, and for cavities with a high aspect ratio. Furthermore, this phenomenon makes Eq. (2) useless for predicting the eigenfrequencies of a cavity with only a single line of pipes present.

Finally, pressure nodal lines tend to occur along lines of pipes. This can have implications on noise abatement techniques used in such applications.

ACKNOWLEDGMENTS

This work has been carried out in collaboration with the Danish company Ødegaard & Danneskiold-Samsøe A/S. The authors acknowledge the significant contribution from Graeme Keith to all the stages of the project.

[1] R. D. Blevins, "The effect of sound on vortex shedding from cylinders," J. Fluid Mech. **161**, 217–237 (1985).
[2] R. D. Blevins, *Flow-Induced Vibration*, 2nd ed. (Krieger Publishing, Florida, 1994).
[3] M. S. Howe, *Acoustics of Fluid–Structure Interactions* (Cambridge University Press, Cambridge, 1998).
[4] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[5] D. Ouellet, J.-L. Guyader, and J. Nicolas, "Sound field in a rectangular cavity in the presence of a thin, flexible obstacle by the integral equation method," J. Acoust. Soc. Am. **89**, 2131–2139 (1991).
[6] R. Parker, "Acoustic resonances in passages containing banks of heat exchanger tubes," J. Sound Vib. **57**, 245–260 (1978).
[7] E. Meyer and E. G. Neumann, *Physical and Applied Acoustics* (Academic, New York, 1972).
[8] B. Massey and J. Ward-Smith, *Mechanics of Fluids*, 7th ed. (Stanley Thornes Publishers Ltd., City, 1998).
[9] Free Field Technologies, MSC, ACTRAN *2003 User's Guide*, 2002.

A hybrid active/passive exhaust noise control system for locomotives

Paul J. Remington
BBN Technologies, 10 Moulton Street, Cambridge, Massachusetts 02138

J. Scott Knight, Doug Hanna, and Craig Rowley
BBN Technologies, 11 Main Street, Mystic, Connecticut 06355

(Received 2 August 2004; revised 6 October 2004; accepted 30 October 2004)

A prototype hybrid system consisting of active and passive components for controlling far-field locomotive exhaust noise has been designed, assembled, and tested on a locomotive. The system consisted of a resistive passive silencer for controlling high-frequency broadband noise and a feedforward multiple-input, multiple-output active control system for suppressing low-frequency tonal noise. The active system used ten roof-mounted bandpass speaker enclosures with 2–12-in. speakers per enclosure as actuators, eight roof-mounted electret microphones as residual sensors, and an optical tachometer that sensed locomotive engine speed as a reference sensor. The system was installed on a passenger locomotive and tested in an operating rail yard. Details of the system are described and the near-field and far-field noise reductions are compared against the design goal.
 [DOI: 10.1121/1.1836851]

PACS numbers: 43.50.Ki, 43.50.Gf, 43.50.Lj [KAC] Pages: 68–78

I. INTRODUCTION

A. Background

Diesel-electric locomotives are generally noisier than comparable electric locomotives. Since the U.S. railroad industry is powered primarily by diesel electric locomotives, noise from these units is a substantial barrier to the introduction of high-speed passenger rail service in the U.S. As indicated in Fig. 1, the primary sources of diesel-electric locomotive noise are the engine exhaust and the cooling fans. The data in the figure are based on measurements made in the frequency range from 40 Hz to 10 kHz at 100 ft. to the side of a 3000-HP locomotive.[1] Rolling-noise data are based on measurements at 100 ft. from track centerline with the locomotive coasting by. It is clear from the figure that both engine exhaust and the cooling fans must be reduced before significant reductions in diesel-electric locomotive noise can be achieved.

This program focused on controlling the engine exhaust noise through a combination of passive and active treatments. Because of the limited space inside the locomotive engine compartment, passive silencers alone cannot provide sufficient noise reduction, especially at low frequency. A purely active system, on the other hand, would become quite costly and complex at high frequency, where significant broadband noise attenuation is needed. A variety of approaches has been applied to the active control of exhaust noise. The earliest work was carried out by Chaplin.[2,3] More recent work includes active valving of exhaust,[4,5] active liners for resistive mufflers,[6] low-backpressure designs with exhaust noise shaping,[7,8] specialized transducer design for exhaust systems,[9] and specialized active systems for locomotives[10] to name just a few. Two alternate approaches were considered for this active system: an active liner employing a feedback controller and actuators and sensors contained within the exhaust system, and a feedforward system employing actuators and sensors external to the exhaust system. The latter approach was selected for a number of reasons, the most compelling being the lack of available actuators and sensors for the extreme high temperatures in the locomotive exhaust.

The system that was developed is illustrated schematically in Fig. 2. The controller in the active system is driven by a reference signal from an engine tachometer. The controller in turn drives an array of loudspeakers in enclosures surrounding the exhaust stack. Control microphones on the roof of the locomotive monitor the performance and are used by the controller along with the tachometer signal to adapt the control filter. The active system surrounds a specially developed resistive exhaust silencer. A resistive silencer was chosen because such silencers typically have low backpressure, a critical requirement in locomotive silencers. High backpressure can have a significant negative impact on efficiency, which is an overriding economic consideration with locomotives. Techniques for the design of resistive silencers are well known,[11] although finite-element analysis was utilized here to evaluate alternate designs in order to deal with the complex geometry of the silencer.

While a number of studies of active exhaust noise control system have been carried out as indicated above, this study is unique in that passive and active systems have been specifically designed to work together and to function in the locomotive environment where operating conditions are severe and the sound levels to be controlled are uniquely intense at low frequency.

B. Objective

The goal of this program is to design, build, and test a prototype exhaust noise reduction system that will reduce the exhaust noise signature of a passenger locomotive by 10 dBA. An additional requirement is that the hardware be suf-

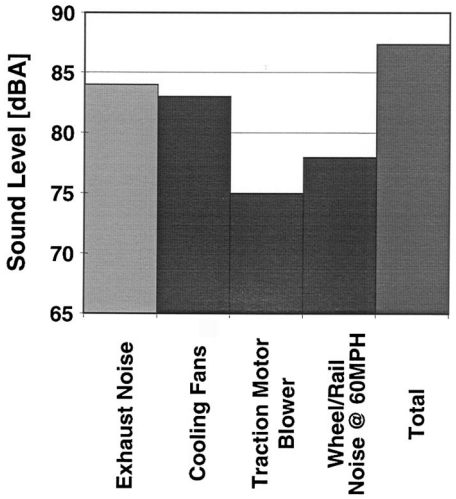


FIG. 1. Noise sources on an SD40-2 diesel-electric locomotive measured at 100 ft. with the locomotive running at throttle 8 at full load.

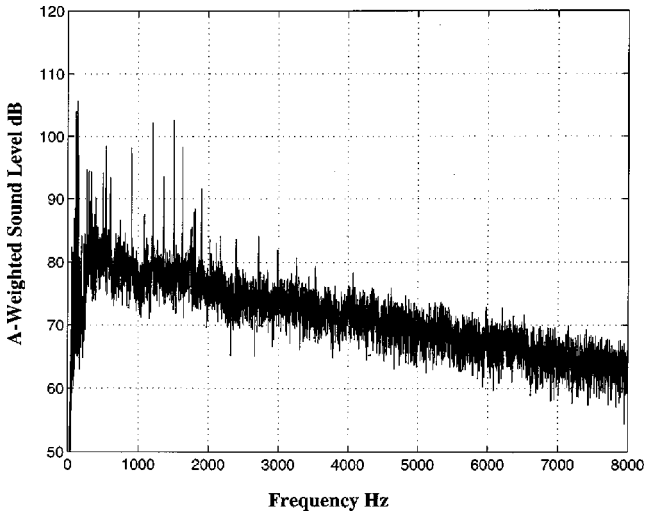


FIG. 3. The A-weighted exhaust noise spectrum 3 ft. aft of the exhaust stack at throttle 8 at full load.

ficiently durable and compatible with the locomotive that the system could be run in normal service for a period of 6 months. The 10-dBA goal is based on the relative magnitude of the source strengths on the locomotive as shown in Fig. 1. A 10-dBA reduction will bring exhaust noise down to the level of the traction motor blowers, the next highest source after the cooling fans. Of course, all of this is done in anticipation of the cooling fans ultimately receiving similar treatment.

II. THE NOISE REDUCTION REQUIREMENTS

Since both a passive silencer and an active system are to be designed, the minimum noise reduction required from each and the frequency range over which each must function must be determined. It is also necessary to specify, for the active system, if control of only tonal noise is sufficient or if broadband control is required.

Figure 3 shows the spectrum of the A-weighted sound-pressure level (SPL) measured approximately 3 ft. from the edge of the exhaust stack of a 3000-HP passenger locomotive operating in throttle 8 in self-load.[12] Note that in modern-day diesel-electric locomotives there are eight throttle settings, throttle 1 through throttle 8, and two idle settings, idle and high idle. Throttle 1 is the lowest speed and power setting, throttle 8 is the highest and, as the throttle setting increases from 1 through 8, the engine speed and power increase incrementally. High idle is simply a higher idle speed than normal idle. Note also that locomotives of this type often have a self-loading feature in which the power generated by the main alternator can be directed to the dynamic brake resistors. This feature was used for all of the loaded tests in this program. Unfortunately, these resistors need to be cooled and the dynamic brake fan that cools them can be quite noisy. However, by measuring close enough to the exhaust stack, contamination from this fan can be minimized.

The data were acquired using a 1/2-in. condenser microphone, recorded on a DAT recorder, and analyzed using a two-channel (800 line) FFT analyzer. Because the data were measured very near to the exhaust stack, most of the noise in Fig. 3 comes from the diesel engine exhaust. Although there may be some contributions from the dynamic brake fans and the radiator cooling fans, earlier studies[1] would tend to indicate that these other sources would only be significant in a few frequency bands and then only for measurements far from the locomotive.

Table I examines whether reduction of only the tones in the exhaust signature is sufficient, or whether broadband noise reduction is required. The table focuses on the low frequencies where the active system would function. It shows the estimated noise reduction that would occur if an

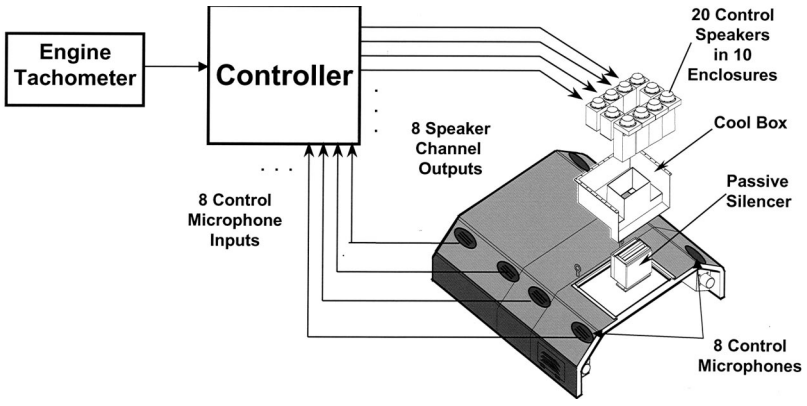


FIG. 2. Schematic drawing of the active–passive exhaust noise control system.

TABLE I. Total noise reduction in the control band for various control bands and tonal noise reductions.

Frequency range for tonal noise reduction (control band) [Hz]	Amount of tonal reduction [dB]	Total noise reduction in the control band [dBA]
0–250	Total removal	17
0–250	20	15
0–250	10	9
0–500	Total removal	6
0–500	20	6
0–500	10	5

active noise control system were used to suppress tones to varying degrees in various frequency bands. This table shows that controlling only the tones from 0 to 250 Hz will result in significant overall noise reduction and broadband noise control is not needed in that frequency band. If the control band is extended to 500 Hz, the table shows that just controlling the tones is not enough. Consequently, it is necessary to design the active component of the system to control just the tones from 0 to 250 Hz and rely on the passive silencer to control the noise above 250 Hz.

Figure 4 shows the A-weighted overall reduction in exhaust noise that would result if the following were to occur:

(i) active reduction of all tones below 250 Hz by 10 dB;
(ii) broadband passive noise reduction from 250 to 500 Hz by 5 dB; and
(iii) broadband passive noise reduction from 500 Hz to a selected higher frequency by 15 dB.

Figure 4 shows the increase in overall A-weighted exhaust noise reduction as the highest frequency at which the silencer is effective is increased. The curve crosses 10 dBA at ~5.5 kHz, indicating that the passive silencer should provide 15 dB of insertion loss up to at least 5.5 kHz. Note that if no low-frequency active control were employed the overall noise reduction from the passive silencer would be limited to ~5 dBA, and that active control of tones in the exhaust below 250 Hz with no broadband control at the higher frequencies would result in less than 1 dBA of noise reduction. Consequently, the control of overall exhaust noise requires an active system to control tones below 250 Hz and a passive silencer to control the high frequencies from 250 Hz to 5.5 kHz.

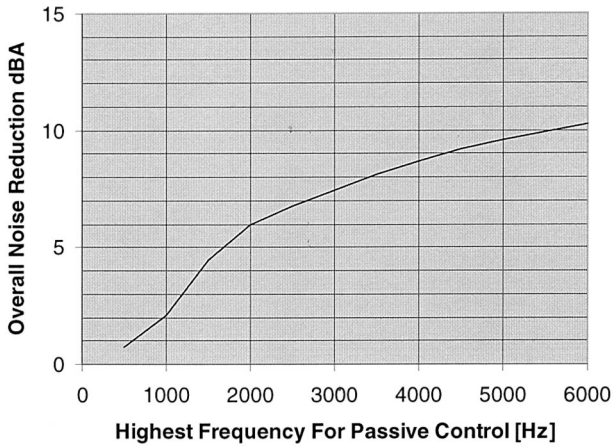


FIG. 4. A-weighted overall noise reduction as a function of the highest frequency of effective broadband noise control.

III. DESIGN OF THE ACTIVE SYSTEM

A. Actuator and sensor number and placement

To examine the number of control actuators and residual or control microphones needed for effective control of exhaust noise, a first-order model of the sound propagation from the exhaust stack was employed for the configuration of actuators and sensors shown in Fig. 2. The model used simple spherical spreading from a point source at the center of the exhaust and from an array of actuators placed around the exhaust stack as shown in the figure. The mean square sound pressure at 90 points 30 m from the center of the exhaust stack was used as a measure of system performance. The equations for the residual sound pressure, $\boldsymbol{\varepsilon}$, are given by

$$\boldsymbol{\varepsilon} = \mathbf{p_{ff}} - [\mathbf{H_{ff}}]\mathbf{v_s}, \tag{1}$$

where $\mathbf{p_{ff}}$ is the vector of sound pressures at the far-field locations due to the uncontrolled exhaust, $\mathbf{v_s}$ is the vector of volume velocities of the control actuators, and $[\mathbf{H_{ff}}]$ is the matrix relating the control actuator volume velocities to the far-field sound pressure. In addition, the equations for the residual sound pressures in the control microphones, $\boldsymbol{\varepsilon}_c$, are given by

$$\boldsymbol{\varepsilon}_c = \mathbf{p}_c - [\mathbf{H}_c]\mathbf{v_s}, \tag{2}$$

where $\mathbf{p}_c$ is the vector of sound pressures at the control microphone locations due to the uncontrolled exhaust, and $[\mathbf{H}_c]$ is the matrix relating the control actuator volume velocities to the control microphone sound pressures. The number of significant singular values in $[\mathbf{H_{ff}}]$ was examined for an array of 32 control actuators spaced at a realistic distance from the exhaust stack, allowing for the physical size of the exhaust opening and the anticipated size of the actuators. With the actuators located as close as possible to the exhaust stack, it was estimated that eight actuators would be sufficient to obtain good performance out to 250 Hz. As expected, increasing the distance of the actuators from the exhaust increased the number of significant singular values and, as a consequence, the number of control actuators. The number and placement of control microphones was then examined, assuming that they would be placed close to the edge of the locomotive hood. From Eq. (2), the residual sound pressure in the control microphones will be minimized in a least-squares sense if

$$\mathbf{v_s} = [\mathbf{H}_c]^{\#}\mathbf{p}_c, \tag{3}$$

where $[\mathbf{H}_c]^{\#}$ means the pseudoinverse of the matrix, $[\mathbf{H}_c]$. By using this result in Eq. (1), it is possible to estimate the reduction in the far-field sound if the sound pressure in the control microphones is minimized

$$\boldsymbol{\varepsilon} = \mathbf{p_{ff}} - [\mathbf{H_{ff}}][\mathbf{H_c}]^{\#}\mathbf{p}_c. \tag{4}$$

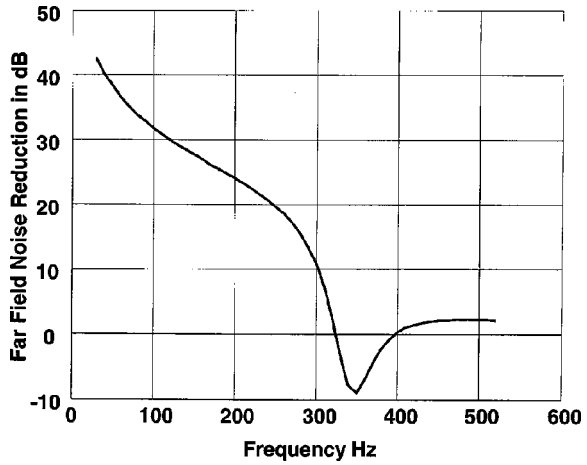


FIG. 5. Reduction in the mean square, angular averaged, sound pressure at 30.4 meters for the sensor–actuator configuration of Fig. 6.

Equation (4) was used to determine the change in the far-field sound pressure for selected control microphone placement locations. Equation (3) was used to estimate the volume velocity requirements of the control actuators. Figure 5 shows the estimated noncausal noise reduction achievable using the actuator and control microphone configuration shown in Fig. 6. Noncausal calculations typically overstate the performance expected from the control system; however, Fig. 5 does indicate that the system as configured in Fig. 6 should be able to provide some performance out to $\sim$250 Hz.

B. Actuator requirements

For the active system to be effective the control speakers must provide sufficient volume velocity to control the sound produced by the locomotive exhaust. Measurements of the sound pressure obtained during tests carried out on a 3000-HP passenger locomotive[12] were used to estimate the required volume velocity. The data were obtained from measurements made with a 1/2-in. condenser microphone placed 0.46 m above and 0.88 m aft of the exhaust stack. The locomotive was operated in self-load. The data were recorded on a DAT recorder and analyzed using a two-channel (800 line) FFT analyzer. We estimated the volumetric velocity in each

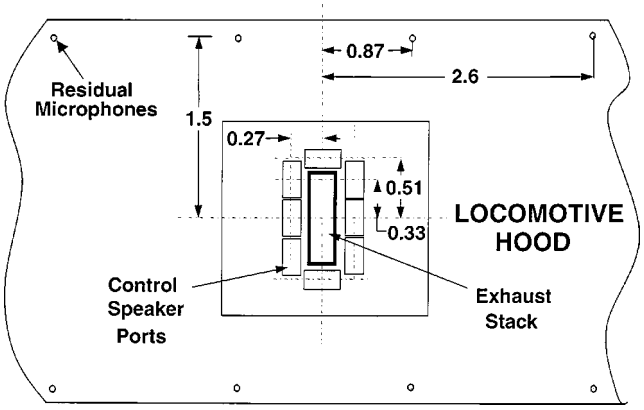


FIG. 6. Plan view of the locomotive hood showing the layout of control actuators and control sensors with dimensions in meters.

of the eight control speakers by using the total sound pressure in each tonal peak to characterize the source strength of the exhaust.

Figure 7 shows the control speaker volume velocity output versus frequency estimated using a commercially available computer code for predicting the performance of high-fidelity speakers in enclosures. The calculations were performed for two 12-in. high-fidelity speakers mounted in a ported enclosure having the geometry of the enclosures shown in Fig. 8. Also shown in Fig. 7 are the required volume velocities at idle and throttles 4, 6, and 8 versus frequency estimated using Eq. (3) along with the measured sound-pressure levels from the locomotive for the two speakers on the center line of the locomotive (see Fig. 6), the locations requiring the highest volume velocities. Figure 7 shows that the design provides no margin for throttle 8 and very little margin for throttle 6. For this reason the control speaker enclosure layout of Fig. 6 was modified by substituting two speaker enclosures on each side of the centerline of the locomotive for the single speaker enclosure, thereby increasing the number of control speaker enclosures from eight to ten. The two speakers in each pair of enclosures are driven with the same controller signal, effectively doubling the volume velocity output.

Figure 8 shows the speaker enclosures ready for installation in the locomotive. Note also that the enclosures have been mounted in a ''cool box,'' an enclosure surrounding the speaker enclosures to route cooling air from the traction motor blowers to protect the speakers from the engine compartment heat.

C. Control microphones

The active control system uses eight identical microphone assemblies to sense the acoustic signals. The assembly, shown in Fig. 9, consists of the microphone element, signal conditioning electronics, and the housing. The microphone element is a commercially available electret microphone that is mounted to a custom board along with power and signal conditioning. The microphone and electronics are packaged in a small, sealed aluminum box with an access hole for the microphone element to fit flush with the surface of the box. This box is encased in open-celled acoustic foam and placed within a protective aluminum housing. The housing is designed to attach to the locomotive hood on the angled transition piece between the top flat roof and vertical sides. The signal from the microphone is passed through a preamplifier that provides some amplification, and frequency shaping to attenuate noise below 25 Hz. The microphone is powered from the electronics cabinet in the locomotive cab.

The package was assembled and tested by mounting it on the roof of a car. The car was then driven at a variety of speeds and the test microphone output, the output of a monitor microphone, and an accelerometer were recorded. The purpose of the test was to determine if the background noise from the microphone, exposed to wind and vibration, was sufficiently low to allow the measurement of locomotive noise on the roof of the locomotive. The data were analyzed using a narrow-band spectrum analyzer and compared with the sound-pressure level measured on the roof of the loco-

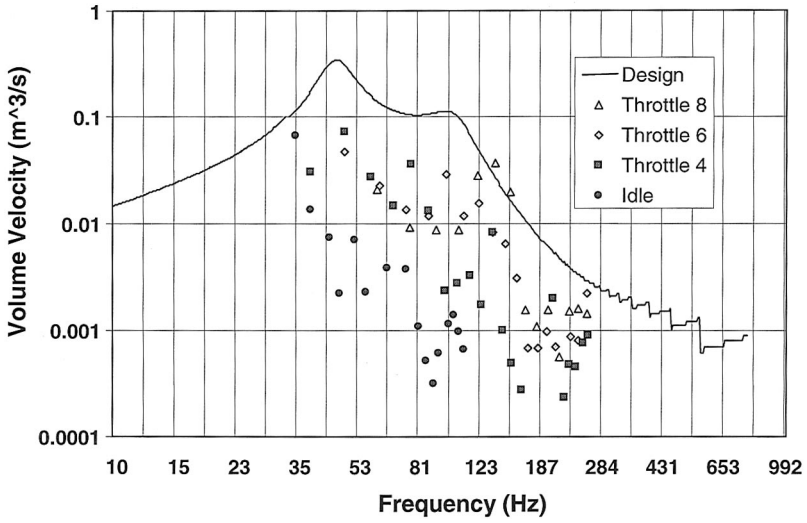


FIG. 7. Predicted speaker volume velocity versus frequency for two McCauley 12-in. speakers in a ported enclosure, each driven at 300 watts compared to the estimated volume velocity needed at idle and throttles 4, 6, and 8.

motive at throttle 4, 6, and 8. The data showed that the microphone/housing design would provide uncontaminated measurements of exhaust tones on the locomotive roof at speeds up to 80 mph for throttles 4, 6, and 8.[12]

D. Controller algorithm and system identification

Figure 10(a) illustrates the structure of the feedforward controller including system identification. The in-line control is above the dashed line and the system identification components are below the dashed line. The controller uses a tachometer on the engine as the reference signal. The output of the controller drives eight speaker channels. The signals from the eight control microphones mounted on the roof of the locomotive are brought back to the controller where, along with the tachometer signal, filtered by a copy of the plant, P, they provide inputs to the LMS algorithm to modify the coefficients of the (1×8) control filter W. The plant P is an (8×8) transfer function matrix relating the

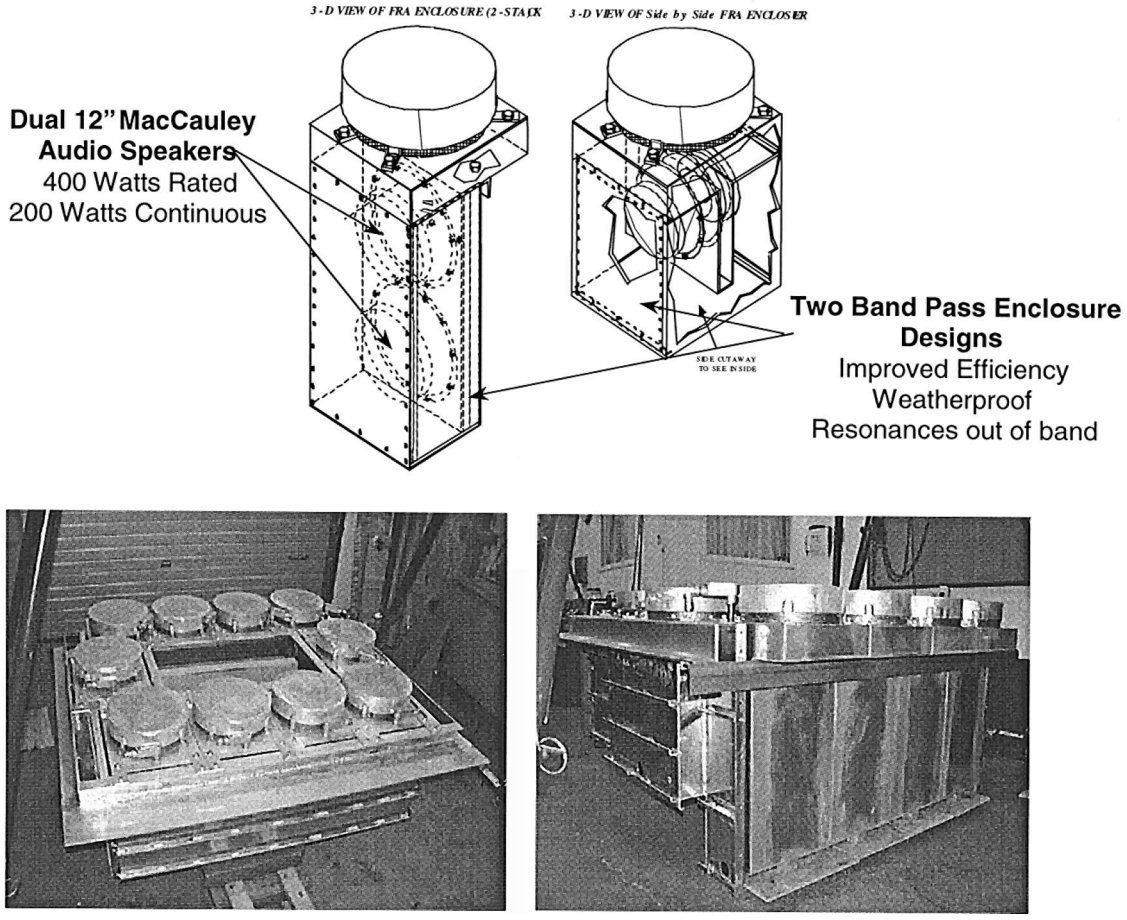


FIG. 8. Speaker enclosure designs with two different geometries.

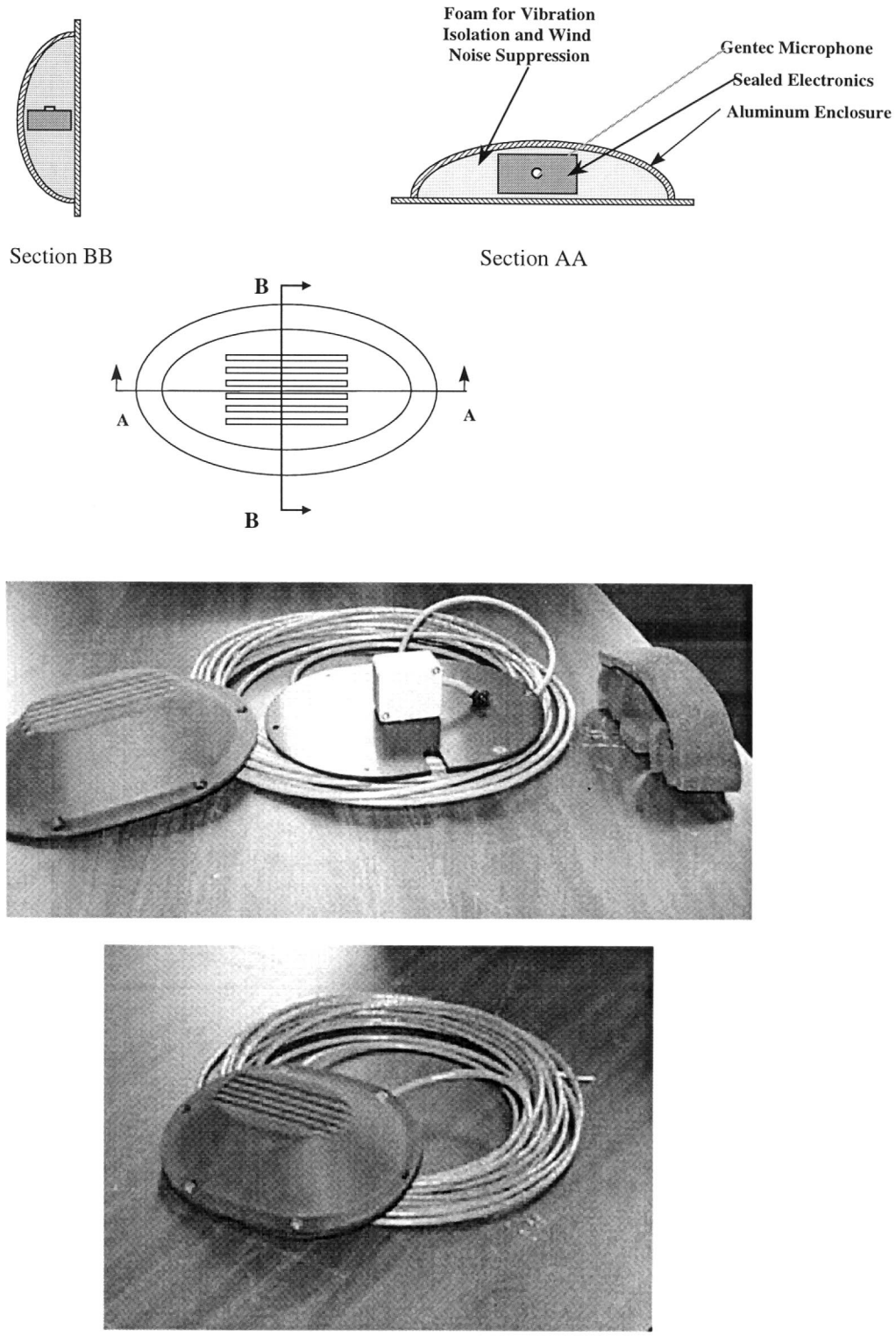


FIG. 9. The control microphone showing the protective enclosure.

eight control filter outputs to the control speakers to the control microphone signals at the controller. A copy of the plant in the controller is required to ensure convergence of the algorithm. Since P will change, it must be measured periodically and its representation in the controller as a FIR filter updated. In the system identification module (below the dashed line) a probe signal is injected into the control loop. This is a broadband signal whose average level is 6 dB below the existing noise. To prevent the probe signal interfering with the LMS algorithm, the probe signal is applied to a filter, which is a copy of the plant transfer function, P, and subtracted from the control microphone signal before that signal is used in the LMS algorithm that adapts the control filter W. The reference signal, the probe signal, and the control microphone signals are then used in the two LMS algorithms at the bottom of the block diagram. The LMS algo-

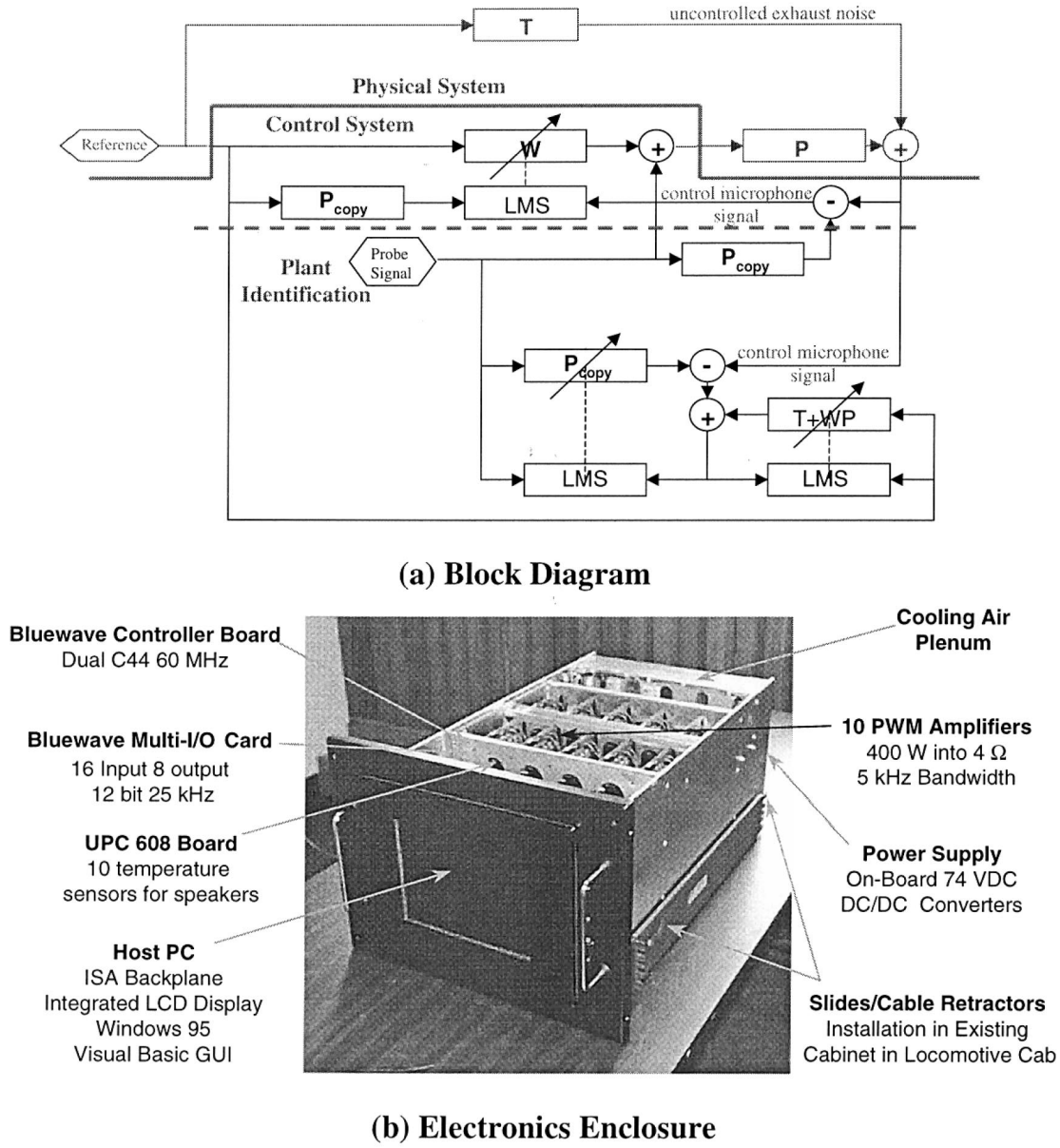


FIG. 10. The controller block diagram and electronics enclosure.

rithm on the left is the core algorithm designed to converge on the plant transfer function, P. The LMS algorithm on the right is designed to converge on $T+WP$ and is designed to further reduce the residual tones in the signal that is applied to the plant identification LMS algorithm. The system identification algorithm runs continuously but updates the plant transfer functions at discrete intervals of time after the LMS algorithm has converged. The probe signals are applied to one control speaker at a time. When the algorithm converges, those eight transfer functions are updated and the next control speaker is driven with the probe. The process continues until all 64 transfer functions have been updated and then starts again.

Because antialiasing filters can be a very expensive component in a controller, it was decided to use the low-order (3-pole) Butterworth filters available on the chosen A/D–D/A I/O boards, described below. To do so meant that it was necessary to sample the input signals in the A/D converters at a rate considerably higher than needed to achieve the 250-Hz control bandwidth. Consequently, a physical sampling rate of 2 kHz was chosen, but the samples were decimated by a factor of 4 in the controller when doing the computations. Using the higher sampling rate allowed for the use of the low-order filters on the I/O boards while still obtaining good rejection of aliased components of the signals.

In designing the controller, it was estimated that a minimum of 32 million floating point operations per second (Mflops) would be required. In sizing processors one ordinarily doubles this estimate and selects a processor with a computational capacity considerably greater than the resulting number. To be on the safe side, it was decided to use two Texas Instruments C44 DSPs, each of which is capable of 60 Mflops. Two were chosen so that the system identification calculations could be performed on one and the in-line control filter calculations could be performed on the second, separating the two processes. This arrangement clearly pro-

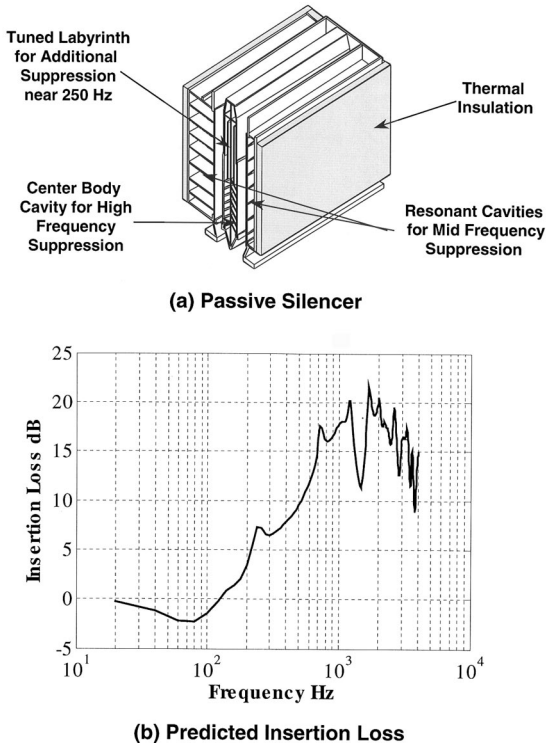


FIG. 11. The geometry and predicted insertion loss of the passive silencer.

vides more than enough computational capacity with sufficient excess capacity to deal with any unmodeled overhead in the computations.

Figure 10(b) shows the electronic enclosure and the components that it contains. In addition to the components already mentioned, the enclosure contains a host computer, boards for accepting inputs from temperature sensors on the speaker coils, and ten pulse-width modulation (PWM) power amplifiers for the speakers.

E. Passive silencer

With the limited space available in the locomotive, the performance requirements for the passive silencer identified in Sec. II presented a significant challenge. In addition, the turbo-charged diesel engine has very stringent backpressure requirements. To meet the requirement for very low backpressure it was decided to design a straight-through resistive silencer. While somewhat better performance might be obtained from a reactive silencer, the backpressure from such a design might be excessive.

The resulting muffler design is sketched in Fig. 11(a). The design consists of an exhaust duct, both surfaces of which communicate with cavities covered with flow-resistive material. In the middle of the duct is a center body with cavities on both surfaces, each covered with the same flow-resistive material. The center body also contains a tuned labyrinth. Both the large cavities in the walls of the duct and the small cavities in the center body have baffles perpendicular to the duct axis. The baffles are designed to prevent the propagation of sound energy in the cavities in the direction parallel to the duct axis. Sound propagating in this way would compromise the insertion loss (IL) of the muffler.

The deep cavities (6.5 in. and 2.56 in.) in the outer walls of the exhaust duct are designed to provide low-frequency noise suppression. These cavities would ideally have been the same depth but space constraints prevented making them so. The shallower cavities in the center body ($\sim$1.5 in.) control high-frequency noise. The tuned labyrinth is designed to control a 250-Hz tone at throttle 8 that preprototype testing showed the active system could not control. It is covered with 25% open perforated metal to provide some broadening in frequency of the IL peak at the cavity resonance. The other flow-resistive material covering the other cavities is five layers of 60 wires/in. screening with 0.011-in.-diameter wire. This material was found to provide the best performance in the numerical simulations.

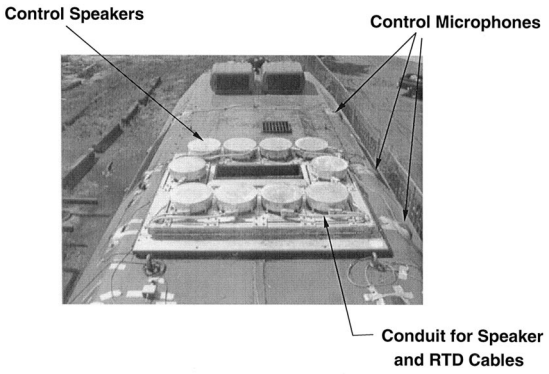
The insertion loss of the silencer was estimated using a general-purpose two-dimensional finite-element code developed for structural acoustic analysis. The air in the exhaust duct and in the muffler cavities and labyrinth was modeled using rectangular quadratic acoustic elements. The coupling of those elements to the air outside the exhaust outlet was modeled using infinite acoustic elements. The inlet to the silencer was treated as having the impedance of an infinitely long duct filled with air at 650 °F. The walls of the silencer and baffles in the cavities were all treated as rigid.

The flow resistances of the material used to cover the cavities and the tuned labyrinth were estimated using analytical procedures developed by NASA for jet engine nacelle liners.[13] The materials considered for providing the flow resistance were perforated metal sheets, multilayers of fine screening, and sintered metal sheet. The procedures include nonlinear flow resistance effects, the effects of mean flow and pressure, and the mass reactance of the air in the holes in the sheeting or screening. Predictions of liner impedance using these predictive formulas generally agreed well with measurements.[12]

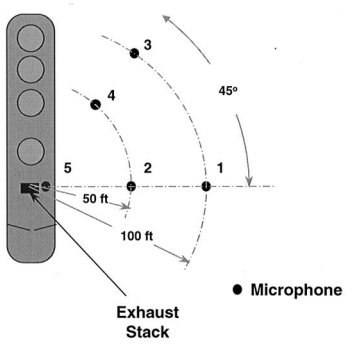
The IL of the prototype silencer in Fig. 11(a) was estimated by utilizing the FEA modeling approach described previously. In the calculation the flow resistance was estimated for the five layers of stainless-steel screening using the procedures in Ref. 13. The IL predictions are given in Fig. 11(b). The figure shows that the low-frequency requirements of 5-dB IL from 250 to 500 Hz is satisfied and the high-frequency requirement of 15 dB above 500 Hz is also satisfied except for a few narrow frequency bands up to 3 kHz. Up to 4 kHz the IL lies for the most part between 10 and 15 dB. Finally, at 250 Hz there is a slight peak in the IL due to the tuned labyrinth. All in all, the design comes very close to satisfying the requirements for the passive silencer.

IV. SYSTEM PERFORMANCE

The locomotive, provided by the Chicago Metra Commuter Rail System, with the exhaust noise control system installed is shown in Fig. 12(a). The ten control speakers and most of the eight control microphones (temporarily installed) are clearly visible. Figure 12(b) shows the position of $\frac{1}{2}$-in. condenser microphones around the locomotive used for the acoustic evaluation of the system. The numbers in the figure refer to the microphone numbers. Tests were performed before installation of the system and after installation of the



(a) The Test Locomotive with the system installed



(b) Evaluation Microphone Locations

FIG. 12. The test locomotive with the system installed and a plan view of the microphone locations.

system with the active system off and on. All data were recorded on a multichannel DAT recorder for later analysis. Here, the focus is on the reduction in noise measured on the roof of the locomotive near the exhaust stack and at the four far-field microphones. The cooling fans were turned off during all testing.

A. Noise reduction near the exhaust stack

The sound-pressure level at microphone 5 on the roof of the locomotive 4 ft. from the exhaust stack is shown in Figs. 13(a) through (d) for four representative operating conditions: idle, high idle, throttle 4 and throttle 8 with the active system on and off. For throttle 4 and 8 the locomotive was in self-load and so the dynamic brake fan was operating during those tests. The noise from this fan presented no problem during the evaluation of the tonal noise reduction performance of the active system, but did present some problems in the evaluation of the new passive silencer, as will be discussed below.

Figures 13(a) through (d) show the intense, low-frequency tonal energy characteristic of locomotive exhaust noise. The tones become significant just below 40 Hz, peak in magnitude between 40 to 150 Hz, and then decline in magnitude. The figures show a dramatic reduction in the magnitude of these tones due to the active system. In some instances the reduction of individual peaks is as much as 30 dB. At idle, fairly significant out-of-band amplification can be seen in the vicinity of 200 Hz. Out-of-band amplification is less noticeable at the other throttle settings.

Overall reductions at microphone 5 on the roof of the locomotive in the control band (25–200Hz) due to the active system are shown in the second column of Table II versus throttle setting. The reductions are 10 dB or more except for throttle 6 (loaded), and reflect the fact that the reduction in the low-frequency noise could be easily distinguished by human observers when the system was turned on and off.

The passive silencer insertion loss (IL) was also determined by measuring the noise on the roof of the locomotive at microphone 5 (where exhaust noise is likely to dominate). These measurements were performed when the OEM silencer was still installed, and were then repeated after replacing the OEM silencer with the new passive silencer. The results of those measurements are shown in the third column of Table II. The table shows the change in the A-weighted sound level above 200 Hz where the silencer was designed to function. The performance, while respectable, is less than desired. It is unclear at this time why the silencer underperformed and program resources were insufficient at the close of the program to tear down the muffler and diagnose the source of the problem. It is possible that the reduced IL for the loaded throttle settings compared to the unloaded was caused by contamination from the dynamic brake fan, which was operating for the loaded cases but not for the unloaded.

Column 4 of Table II shows the overall A-weighted noise reduction for the passive and active components of the system combined over the full frequency range of interest. The reductions under all conditions are very encouraging. For many of the operating conditions the performance is very close to the goal of 10 dBA of noise reduction.

B. Noise reduction in the far field

Expectations were that the noise reductions due to the active system in the far field would be less than those measured on the roof. Table III summarizes the overall active system noise reduction below 200 Hz at the far-field microphone locations, and shows that those expectations were fulfilled. The active system performance at the far-field microphones was less than observed at the roof microphone. Controlling the noise at eight locations on the roof does not guarantee that the control will be as good at remote locations in the far field. While the modeling carried out during the design process suggested that the number and location of the control actuators and sensors should have achieved the desired noise reduction, the modeling process was very approximate and did not include such refinements as shielding due to the locomotive hood, ground reflections, etc.

It was not possible to measure directly the overall noise reduction due to the combined active and passive systems at the far-field microphone locations because of contamination from background noise and other noise sources, such as the dynamic brake fan. This contamination was of primary concern for the noise reduction due to the new passive silencer. The change in the tones at the far-field microphones due to the active system could be easily measured. Therefore, the change in the exhaust noise spectrum at each far-field microphone as the active system was turned on and off was used to determine the noise reduction as a function of frequency due to the active system alone. The exhaust noise spectrum at

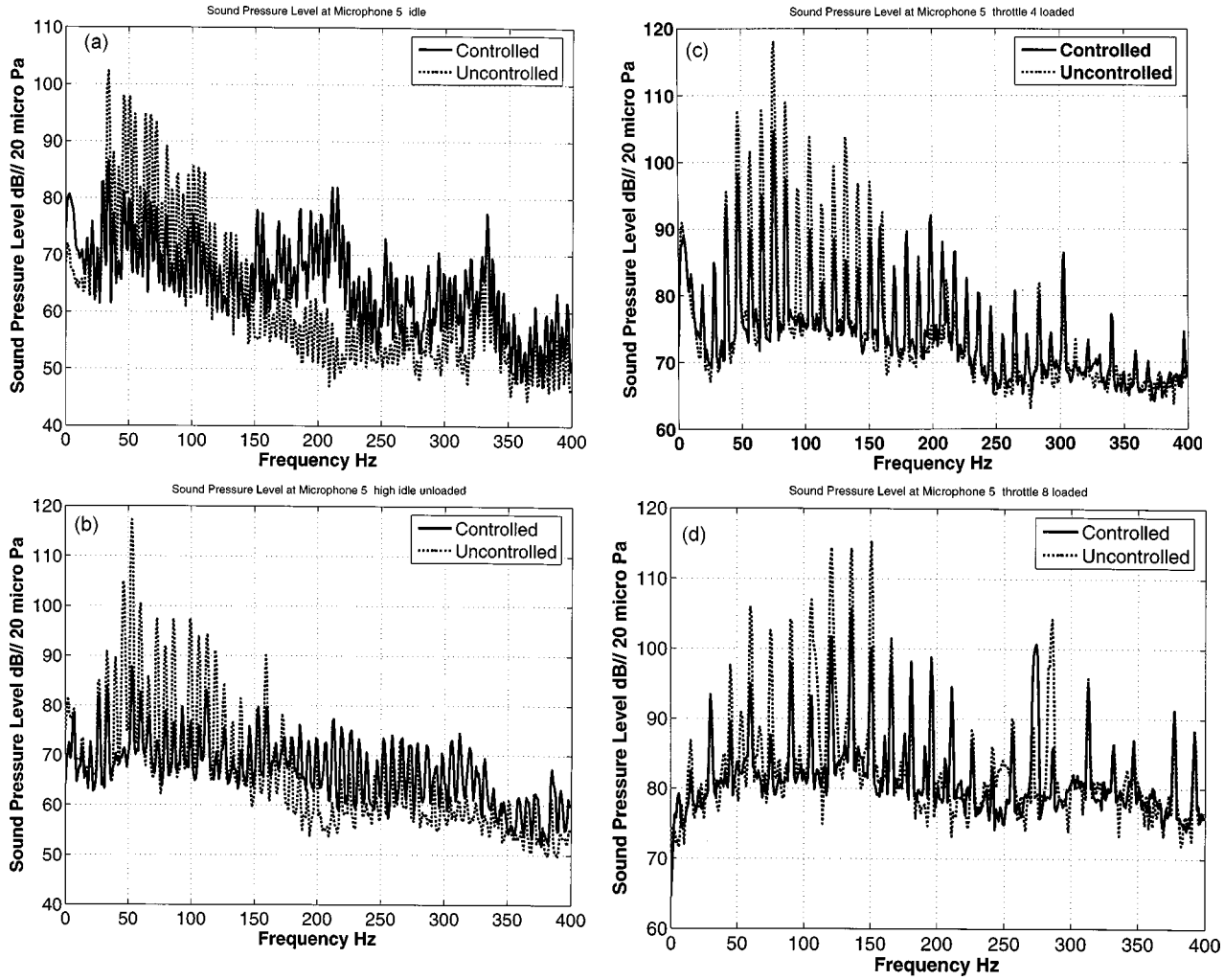


FIG. 13. (a) Narrow-band performance of the active system with the locomotive operating at idle. (b) Narrow-band performance of the active system with the locomotive operating at high idle. (c) Narrow-band performance of the active system with the locomotive operating at throttle 4 loaded. (d) Narrow-band performance of the active system with the locomotive operating at throttle 8 loaded.

microphone 5 (due to both the new passive silencer and the active system) was then computed using the active system noise reduction corresponding to that observed at the selected far-field microphone and the noise reduction as a function of frequency due to the new passive silencer measured at microphone 5. The spectra with and without treatment were then A-weighted and the overall level computed for each. The difference between the overall A-weighted levels was then taken as the noise reduction of the total system at the selected microphone. The implicit assumption in this approach is that the IL of the new passive silencer and the distribution of energy with frequency in the exhaust noise spectrum is the same near the exhaust as at the far-field locations. Table IV presents the results. Except for idle, the noise reductions at each microphone position for each operating condition are less than the near-field results. However, except for throttle 6 (loaded) and throttle 8 (unloaded) at microphone 1, where there was no decrease or an increase in noise, the noise reductions are still quite respectable.

TABLE II. Overall noise reduction of the active and passive systems as measured on the roof of the locomotive at microphone 5.

Throttle Setting	Overall noise reduction due to active system in control band below 200 Hz dB	Overall A-weighted noise reduction due to new passive silencer above 200 Hz dBA	Overall A-weighted noise reduction due to new passive silencer and active system dBA
Idle	13	8	5
High idle Unloaded	23	9	9
Throttle 4 Unloaded	14	8	7
Throttle 6 Unloaded	13	8	8
Throttle 8 Unloaded	11	6	6
Throttle 4 Loaded	12	5	6
Throttle 6 Loaded	9	4	4
Throttle 8 Loaded	10	5	7

V. CONCLUSIONS

A hybrid active passive exhaust noise control system for a locomotive has been designed, fabricated, installed on a

TABLE III. Overall noise reduction of the active system below 200 Hz at all microphone locations [dB].

Throttle Setting	Mic. #1 100 ft. 90°	Mic. #2 50 ft. 90°	Mic #3 100 ft. 45°	Mic. #4 50 ft. 45°	Mic. #5 4 ft. Roof
Idle	4	7	4	4	13
High idle	12	16	10	16	23
Throttle 4 Unloaded	6	4	7	7	14
Throttle 6 Unloaded	2	1	4	2	13
Throttle 8 Unloaded	−6	−1	4	2	11
Throttle 4 Loaded	6	6	8	7	12
Throttle 6 Loaded	2	4	5	4	9
Throttle 8 Loaded	2	2	4	4	10

locomotive, and tested with the locomotive stationary. Overall A-weighted noise reduction measured near the exhaust stack outlet (near field) ranged from 4 to 9 dBA depending on operating condition. This was less than the 10-dBA goal, due mostly to the fact that the passive silencer did not achieve its design insertion loss, although at the higher throttle settings there was some evidence that the active system performance was limited by the control speaker output. In the far field (50 to 100 ft. from the exhaust stack) the estimated overall A-weighted noise reduction was less than in the near field, ranging from a high of 9 dBA to a low of −2 dBA depending on microphone location and operating condition. Active system noise reduction was also less than in the near field. Unfortunately, program resources and a restricted testing schedule did not allow us to examine varying the number and location of control actuators or control microphones, which may have yielded some improvements in active system performance in the far field.

TABLE IV. Estimated overall A-weighted noise reduction at all microphones.

Throttle setting	Load	Mic 5	Mic 1	Mic 2	Mic 3	Mic 4
Idle	unloaded	5	5	6	6	6
High idle	unloaded	9	6	9	6	7
Throttle 4	unloaded	7	5	4	6	5
Throttle 6	unloaded	8	7	6	6	6
Throttle 8	unloaded	6	0	4	6	5
Throttle 4	loaded	6	5	4	5	5
Throttle 6	loaded	4	−2	1	3	2
Throttle 8	loaded	7	3	6	7	7

ACKNOWLEDGMENTS

The work was supported by the US Department of Transportation Federal Railroad Administration under Contract DTFR53-95-000072. We are grateful for the support and cooperation of Chicago Metra, which provided us with the test locomotive, test site, and personnel and equipment to help in the installation of the system. We also wish to acknowledge the support of the Electromotive Division of General Motors, which was responsible for all locomotive interface issues and provided personnel on site during the installation and test of the system.

[1] P. J. Remington, M. J. Rudd, and R. Mason, ''Measurement and diagnosis of diesel electric locomotive noise,'' Noise Control Eng. **14**(2), 66–73 (1980).

[2] G. B. B. Chaplin and R. A. Smith, ''The sound of silence—The silencing of diesel exhausts by out-of-phase cancellation using a microprocessor has now been achieved,'' Engineering (London) **218**, No. 7, 672–673 (1978).

[3] M. C. J. Trinder, G. B. B. Chaplin, and P. A., Nelson, ''Active control of commercial vehicle exhaust noise,'' Internoise 86, International Congress on Noise Control Engineering, Cambridge, MA, 21–23 July, 1986. Proceedings, pp. 611–616.

[4] R. Boonen and P. Sas, ''Development of an active exhaust silencer for combustion engines,'' J. Acoust. Soc. Am. **105**, 1089 (1999). Paper presented at the Joint ASA/EAA/DAGA Meeting, Berlin, DE, 14–19 March, 1999. Paper 2pSAb1.

[5] R. Boonen and P. Sas, ''Design of an active exhaust attenuating valve for internal combustion engines,'' in: Proceedings of ACTIVE 2002, The 2002 International Symposium on Active Control of Sound and Vibration, edited by P. Gardonio and B. Rafaely, Southampton, UK, 15–17 July, 2002, pp. 345–356.

[6] I. L Vér and M. Dignan, ''Active Liner System for Jet Engine Exhaust Silencers,'' Internoise 95, International Congress on Noise Control Engineering, Newport Beach, CA, 10–12 July, 1995, Proceedings, pp. 481–484.

[7] J. Clay Shipps, ''The Plymouth Prowler, active noise control exhaust system. I. Noise reduction,'' J. Acoust. Soc. Am. **105**, 1301 (1999) (Abstract) Paper 4pSAb3, presented at the Joint ASA/EAA/DAGA Meeting, Berlin, DE, 14–19 March, 1999.

[8] T. E. Meeks and J. Clay Shipps, ''The Plymouth Prowler, active noise control exhaust system. II. Noise shaping,'' J. Acoust. Soc. Am. **105**, 1301 (1999). Paper 4pSAb4, presented at the Joint ASA/EAA/DAGA Meeting, Berlin, DE, 14–19 March, 1999.

[9] S. P. Kahn, ''Design considerations for a low-impedance loudspeaker in active noise control exhaust systems,'' J. Acoust. Soc. Am. **105**, 1301 (1999), Paper 4pSAb5, presented at the Joint ASA/EAA/DAGA Meeting, Berlin, DE, 14–19 March, 1999.

[10] F. Cotana and F. Rossi, ''Active noise control technique for diesel train locomotor exhaust noise abatement,'' J. Acoust. Soc. Am. **112**, 2427–2428 (2002). Paper 5pAAa11, presented at the 1st Pan-American/Iberian Meeting on Acoustics and 144th ASA Meeting, Cancun, Mexico, 2–6 Dec., 2002.

[11] A. G. Galaitsis and I. L. Ver, ''Passive Silencers and Lined Ducts,'' Chap 10 in *Noise and Vibration Control Engineering*, edited by L. L. Beranek and I. L. Ver (Wiley, New York, 1992).

[12] P. J. Remington, S. Knight, D. Hanna, and C. Rowley, ''A hybrid active/passive exhaust noise control system (APECS) for locomotives,'' BBN Report 8302, BBN Technologies, Cambridge, MA, March, 2001.

[13] R. E. Motsinger and R. E. Kraft, ''Design and Performance of Duct Acoustic Treatment,'' Chap. 14 in *Aeroacoustics of Flight Vehicles*, Vol. 2, Noise Control, edited by Harvey H. Hubbard (Acoustical Society of America, Woodbury, NY, 1995).

Ambient noise cross correlation in free space: Theoretical approach

Philippe Roux,[a] Karim G. Sabra, and W. A. Kuperman
Marine Physical Laboratory of the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

Andre Roux
Laboratoire MMAS, CNRS FRE 2344, Université de Metz, France

(Received 19 July 2004; revised 18 October 2004; accepted 18 October 2004)

It has been experimentally demonstrated that the Green's function between two points could be recovered using the cross-correlation function of the ambient noise measured at these two points. This paper investigates the theory behind this result in the simple case of a homogeneous medium with attenuation. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1830673]

PACS numbers: 43.20.Fn, 43.60.Ac [RLW] Pages: 79–84

I. INTRODUCTION

The goal of this work is to theoretically investigate the following problem: can we retrieve the time-domain Green's function (TDGF) between two points by performing a cross correlation of the ambient noise field received on those two points? Experimental demonstration of this process has been performed in ultrasonic,[1] underwater acoustics,[2] or geophysics.[3,4] In most cases, only an estimate of the Green's function was retrieved, with quality strongly dependent on the medium complexity as well as the spatio-temporal distribution of the ambient noise sources. For example, it has been shown that an amplitude-shaded TDGF could be obtained with surface ambient noise in underwater acoustics, the shading being due to the surface location of noise sources in the ocean.[2,5] Similarly, Lobkis and Weaver[6] have shown the emergence of the exact TDGF from the diffuse field due to thermal fluctuations in a reverberant ultrasonic cavity. Finally, earlier works by Rickett and Claerbout conjectured that this process could also be used to retrieve the sound-speed structure of the upper crust of the earth in geophysics, the experimental demonstration being done in helioseismology from data describing the random vibration of the sun's surface.[3] Because noise sources are difficult to work with, some related results have also been obtained with noise-like events, where signals recorded from randomly distributed sources were used to perform the cross correlation.[7,8] For example, in Ref. 7, the cross correlation of the coda waves from a distribution of seismic events provided the Rayleigh wave between two seismometers.

From a theoretical point of view, earlier works have investigated the problem of spatial correlation with noise fields[9,1] or with wave fields obtained from a distribution of random sources.[10,11] In the case of a 3D free-space medium with a spatially uniform noise source distribution, the field at each receiver can be decomposed as a superposition of uncorrelated plane waves from various directions. It has been established[9] that the normalized cross-spectral density $C_{1,2}(\omega)$ at frequency ω between two receivers 1 and 2 separated by a distance r is $C_{1,2}(\omega)=[\sin(\omega r/c)]/(\omega r/c)$. In the time domain, the normalized correlation function is

$$C_{1,2}(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} C_{1,2}(\omega)\exp(i\omega t)\,d\omega,$$

which can be written as

$$C_{1,2}(t)=\frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{\exp[i\omega(t+r/c)]}{i\omega r/c}\,d\omega - \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{\exp[i\omega(t-r/c)]}{i\omega r/c}\,d\omega. \quad (1)$$

The time derivative of the correlation function is then

$$\frac{d}{dt}C_{1,2}(t)=\frac{1}{4\pi r/c}[\delta(t+r/c)-\delta(t-r/c)]. \quad (2)$$

The two terms in Eq. (2) correspond to the backward and forward Green's function between the receivers, which demonstrates the connection between the correlation function and the Green's function. However, the drawback of this elegant result is to start from a *normalized* correlation function, normalization that is required because the overall spatial contribution from noise sources in a lossless infinite medium is infinite. Experimentally though, ambient noise signals are always finite, as is the noise correlation function. The contradiction comes from the fact that the theory is developed in lossless environments while experiments are always performed in the presence of attenuation. Thus, normalization acts as a subterfuge for avoiding inclusion of the required attenuation in the theory. The goal of our work is to show how the result in Eq. (2) could be derived rigorously without the need of normalization when attenuation is present in the medium.

In this paper, we work directly with noise in the time domain. Thus, the mathematical developments in this paper start from an infinite-bandwidth formulation of the Green's function. To be as general as possible, we deal here with two receivers simultaneously recording ambient noise in a 3D homogeneous medium. The choice of free-space propagation has been made in regards to the complex developments

[a]Electronic mail: proux@ucsd.edu

needed to achieve the demonstration in a waveguide.[5] The incident field on the two receivers comes from a homogeneous spatial-temporal distribution of uncorrelated broadband noise sources. This model is reasonable in our case because we evaluate the ensemble average of the noise correlation function. It also presents the advantage of considerably simplifying the mathematical developments performed in Secs. II and III.

This paper is structured as follows. In Sec. II, we examine the case of a homogeneous space without attenuation. We use a geometrical interpretation to investigate the relationship between the noise correlation function and the Green's function. In Sec. III, we extend the results to a medium with attenuation. Section IV is a discussion that links this theoretical approach with earlier experimental works.

II. FREE SPACE WITHOUT ATTENUATION

In a 3D homogeneous medium without attenuation, the Green's function between points A (in $\mathbf{r}_1$) and B (in $\mathbf{r}_2$) is

$$G(\mathbf{r}_2,t;\mathbf{r}_1,0)=\frac{1}{|\mathbf{r}_2-\mathbf{r}_1|}\,\delta\left(t-\frac{|\mathbf{r}_2-\mathbf{r}_1|}{c}\right), \tag{3}$$

where c is the constant sound speed in the medium. Assuming a random spatial-temporal distribution of noise sources amplitudes $S(\mathbf{r}_s,t_s)$, the total field received in A is

$$P(\mathbf{r}_1;t)=\int_{-\infty}^{\infty}d\mathbf{r}_s\int_{-\infty}^{t}dt_s S(\mathbf{r}_s,t_s)G(\mathbf{r}_1,t;\mathbf{r}_s,t_s)$$

$$=\int_{-\infty}^{\infty}\frac{d\mathbf{r}_s}{|\mathbf{r}_1-\mathbf{r}_s|}S\left(\mathbf{r}_s,t-\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c}\right). \tag{4}$$

Here, the causality requires that the noise sources in $(\mathbf{r}_s;t_s)$ that contribute to the pressure field in A at a given time t satisfy the condition $t=t_s+|\mathbf{r}_1-\mathbf{r}_s|/c$. Then, the cross correlation of the two signals recorded in A and B is defined as

$$C(\mathbf{r}_1,\mathbf{r}_2;t)=C_{1,2}(t)=\int_{-\infty}^{+\infty}P(\mathbf{r}_1;\tau)P(\mathbf{r}_2;t+\tau)d\tau, \tag{5}$$

which leads to

$$C_{1,2}(t)=\int_{-\infty}^{+\infty}d\mathbf{r}_s\int_{-\infty}^{+\infty}d\mathbf{r}_{s'}\int_{-\infty}^{+\infty}d\tau\frac{1}{|\mathbf{r}_1-\mathbf{r}_s||\mathbf{r}_2-\mathbf{r}_{s'}|}$$

$$\times S\left(\mathbf{r}_s,\tau-\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c}\right)S\left(\mathbf{r}_{s'},\tau+t-\frac{|\mathbf{r}_2-\mathbf{r}_{s'}|}{c}\right). \tag{6}$$

$C_{1,2}(t)$ corresponds to one realization of the ambient noise cross-correlation function. To evaluate the average noise correlation function over an ensemble of realization, we use the fact that noise sources are spatially and temporally uncorrelated

$$\langle S(\mathbf{r}_s,t_s)S(\mathbf{r}_{s'},t_{s'})\rangle=Q^2\delta(t_s-t_{s'})\,\delta(\mathbf{r}_s-\mathbf{r}_{s'}), \tag{7}$$

where the notation $\langle X\rangle$ corresponds to the ensemble average of X. Q^2 represents the acoustic power of the noise sources and is taken constant over time and space. Then, it follows, after integration over $d\mathbf{r}_{s'}$

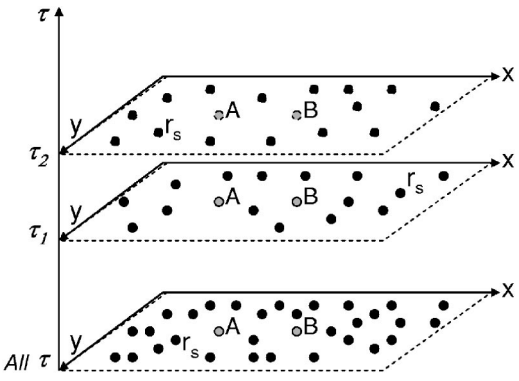


FIG. 1. Representation of the receivers in A and B and the noise sources (in $\mathbf{r}_s$) in the xy plane. Each plane corresponds to a new distribution of noise sources at time τ. The integral over τ in Eq. (6) could be done by accumulating all noise sources over time as shown in the lower plane.

$$\langle C_{1,2}(t)\rangle=Q^2\int_{-\infty}^{+\infty}d\mathbf{r}_s\int_{-\infty}^{+\infty}d\tau\frac{1}{|\mathbf{r}_2-\mathbf{r}_s||\mathbf{r}_1-\mathbf{r}_s|}$$

$$\times\delta\left(t+\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c}-\frac{|\mathbf{r}_2-\mathbf{r}_s|}{c}\right). \tag{8}$$

Note that, as written in Eq. (5), τ corresponds to the running time of the signals received in A and B. However, the cross-correlation function $C_{1,2}(t)$ is a comparison between those two signals, meaning that $C_{1,2}(t)$ will extract the relative propagation times between the noise source in $\mathbf{r}_s$, and the receivers in $\mathbf{r}_1$ and $\mathbf{r}_2$. As a consequence, the integration over τ corresponds just to an accumulation of noise sources over time. As shown in Fig. 1, all those noise events could be excited at one time only without changing the final result in the average correlation function. Actually, the integration over τ in Eq. (8) leads to the usual divergence of a correlation function over an infinite time. However, in a practical experimental case, the pressure fields in A and B first have to be recorded over a finite interval time T before the cross correlation is performed. Assuming that the random noise sources have a creation rate ν ($\mathrm{m}^{-3}\,\mathrm{s}^{-1}$) per unit time per volume, the integral over τ is changed into the product $T\nu$. The accumulation of noise events described in Fig. 1 corresponds then to an increase of ν over a smaller interval time T such that $T\nu$ remains constant. Finally, using this approach, Eq. (8) becomes

$$\langle C_{1,2}(t)\rangle=Q^2T\nu\int_{-\infty}^{+\infty}d\mathbf{r}_s\frac{1}{|\mathbf{r}_2-\mathbf{r}_s||\mathbf{r}_1-\mathbf{r}_s|}$$

$$\times\delta\left(t+\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c}-\frac{|\mathbf{r}_2-\mathbf{r}_s|}{c}\right). \tag{9}$$

Equation (9) shows that the noise correlation function in free space reduces to the calculation of a spatial integral over the noise source's locations. In the following, we show that a geometrical argument allows us to obtain an analytical solution for $\langle C_{1,2}(t)\rangle$. We first define a Cartesian coordinate system for the 3D space in which A is $(a,0,0)$, B is $(-a,0,0)$, and a point $\mathbf{r}_s$ is (x,y,z). The argument of the delta function in Eq. (9) gives a contribution to the correlation function at time t if $\mathbf{r}_s$ is such that $|\mathbf{r}_2-\mathbf{r}_s|-|\mathbf{r}_1-\mathbf{r}_s|$

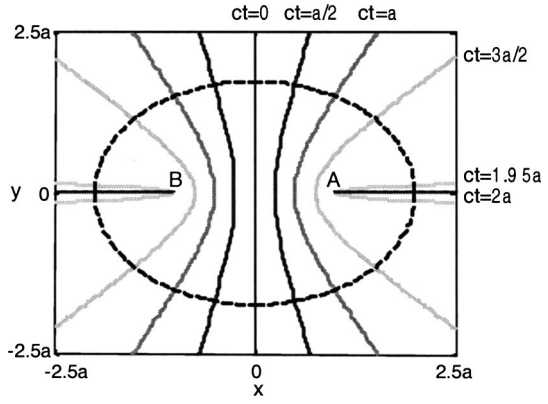


FIG. 2. Representation in the xy plane of the hyperbola that contributes to a given time t in the noise correlation function. On each hyperbola, the noise sources satisfy $|\mathbf{r}_2-\mathbf{r}_s|-|\mathbf{r}_1-\mathbf{r}_s|=ct$. The line $x=0$ corresponds to $ct=0$. The line $y=0$ corresponds to $ct=2a$ for $x\geqslant a$, and $ct=-2a$ for $x\leqslant -a$. The ellipse (dashed line) represents the noise sources for which $|\mathbf{r}_2-\mathbf{r}_s|+|\mathbf{r}_1-\mathbf{r}_s|=ct$, for $ct=4a$. The receivers A and B are at $(a,0)$ and $(-a,0)$.

$=ct$. For a time t satisfying the condition $-2a\leqslant ct\leqslant 2a$, the noise sources must lie on a hyperboloid defined by the equation (Fig. 2)

$$y^2+z^2-x^2\left(\frac{4a^2}{c^2t^2}-1\right)=a^2\left(1-\frac{c^2t^2}{4a^2}\right), \quad \text{if } ct\neq 0, \tag{10}$$

and $x=0$, if $ct=0$.

The hyperbola in Fig. 2 correspond to the noise sources in a 2D section at constant depth z of the 3D space that will contribute to $C_{1,2}(t)$ at a given time t. We will see in Sec. III the importance of the ellipse in Fig. 2 that is made of the noise sources in $\mathbf{r}_s$ such that $|\mathbf{r}_2-\mathbf{r}_s|+|\mathbf{r}_1-\mathbf{r}_s|=ct$. Note that these conical shapes are invariant by rotation around the axis of the two receivers (x axis in Fig. 2). This means that, in 3D, the hyperbola and the ellipse turn out to be a hyperboloid and an ellipsoid whose symmetry axis is the lines between the two receivers. As can be seen in Fig. 2, every noise source in space belongs to a unique hyperbola. Each hyperbola is parametrized by ct, with the condition $-2a\leqslant ct\leqslant 2a$. This implies that $C_{1,2}(t)=0$ for t outside the interval $[-2a/c,2a/c]$. To perform the spatial integration in Eq. (9), we make a change of variable from the Cartesian coordinate to a coordinate system adapted to the hyperbola in Fig. 2

$$\begin{cases} x=a\sin(\theta)\cosh(\varphi) \\ y=a\cos(\theta)\sinh(\varphi)\cos(\psi) \\ z=a\cos(\theta)\sinh(\varphi)\sin(\psi) \end{cases} \text{ with } \begin{cases} \varphi\in\mathcal{R}^+ \\ \theta\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right] \\ \psi\in[0,2\pi] \end{cases}. \tag{11}$$

The Jacobian of this change of variable is

$$J(\varphi,\theta,\psi)=a^3\cos(\theta)\sinh(\varphi)[\cosh^2(\varphi)-\sin^2(\theta)]. \tag{12}$$

Then, we have for any point $\mathbf{r}_s$ defined by the coordinates (φ,θ,ψ) $|\mathbf{r}_2-\mathbf{r}_s|=a(\cosh(\varphi)+\sin(\theta))$ and $|\mathbf{r}_1-\mathbf{r}_s|=a(\cosh(\varphi)-\sin(\theta))$, from which it follows

$$\langle C_{1,2}(t)\rangle=2\pi aQ^2T\nu\int_0^{+\infty}\sinh(\varphi)d\varphi\int_{-\pi/2}^{\pi/2}\cos(\theta) \times\delta\left(t+\frac{2a\sin(\theta)}{c}\right)d\theta. \tag{13}$$

A last change of variable $a\sin(\theta)=x$ finally gives

$$\langle C_{1,2}(t)\rangle=2\pi Q^2T\nu\int_0^{+\infty}\sinh(\varphi)d\varphi\int_{-a}^{a}\delta\left(t+\frac{2x}{c}\right)dx. \tag{14}$$

Knowing that the integration of a Dirac function $\delta(t)$ yields a Heaviside step function $H(t)$, the second integral in Eq. (14) gives a rectangle function

$$\Pi(t)=H\left(t+\frac{2a}{c}\right)-H\left(t-\frac{2a}{c}\right), \tag{15}$$

whose amplitude is 1 between $-2a/c$ and $2a/c$ and 0 elsewhere. The first integral (over φ) in Eq. (14) yields the amplitude of the noise correlation function. The limits of the integral in φ have to be defined to prevent $\langle C_{12}(t)\rangle$ from diverging. Actually, φ is the curvilinear coordinate along each hyperbola. In Fig. 2, $\varphi=0$ is the intersection of the hyperbola with the x axis, while the asymptotic branch of the hyperbola corresponds to $\varphi\rightarrow+\infty$. Integrating over φ between 0 and φ_0 corresponds then to the measure of the length of each hyperbola on this interval. In Fig. 2, the points verifying $\varphi=\varphi_0$ describe an ellipse that is orthogonal to each of the hyperbola. Interestingly enough, this ellipse (or ellipsoid in 3D) is such that $|\mathbf{r}_2-\mathbf{r}_s|+|\mathbf{r}_1-\mathbf{r}_s|=ct$, i.e., the noise sources whose cumulated travel to A and B are the same. We will see in the next section that this ellipse corresponds to the points that have suffered from the same attenuation in the correlation process. Defining the $\varphi=\varphi_0$ ellipsoid as the 3D compact support on which the integration in Eq. (14) is performed leads to the final result

$$\langle C_{1,2}(t)\rangle=2\pi Q^2T\nu[\cosh(\varphi_0)-1]\Pi(t). \tag{16}$$

The time derivative of the average correlation function is then

$$\frac{d}{dt}\langle C_{1,2}(t)\rangle=4\pi aQ^2T\nu[\cosh(\varphi_0)-1] \times\left[\frac{1}{2a}\delta\left(t+\frac{2a}{c}\right)-\frac{1}{2a}\delta\left(t-\frac{2a}{c}\right)\right]. \tag{17}$$

The Dirac functions in Eq. (17) are the causal and anticausal (or time-reversed) Green's functions from A to B. There is no problem in obtaining a time-reversed expression of the Green's function because a correlation function is defined for negative and positive times. Physically speaking, the time symmetry of Eq. (17) results from our hypothesis of spatially uniform ambient noise distribution. Noise sources surrounding the receivers in A and B, the correlation function contains both propagation information from A to B and B to A. Taking the time derivative of the correlation function, the time symmetry means that both the forward and backward Green's functions between A and B are retrieved from ambient noise cross correlation. The amplitude of these Green's functions is

driven by the noise excitation power and the area of the ellipsoidal compact support on which noise sources are taken. Note again that the compact support is necessary in this analysis to prevent the noise correlation function from diverging.

III. FREE SPACE WITH ATTENUATION

One remaining question is the physical meaning of the boundary limit $\varphi=\varphi_0$ necessary to perform the integration in Eq. (14). Actually, φ_0 is strongly related to the presence of attenuation in the medium. Up to now, we have considered a homogeneous propagation medium without attenuation. In that case, the amplitude contribution from noise sources far away from the receivers A and B is not lowered, which leads to a diverging correlation function as seen in Eq. (14). The goal of this section is to show that the presence of attenuation in the medium solves this problem. Volume attenuation is included in the medium by adding an imaginary component to the sound speed $c=c_0+ic_i$, with $c_i \ll c_0$. We chose this usual way to account for attenuation because of its convenience to pursue the mathematical derivation. The frequency dependence of the Green's function that results from attenuation is discussed later. The Green's function is then modified as follows:

$$G(\mathbf{r}_2,t;\mathbf{r}_1,0)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}d\omega\,\frac{1}{|\mathbf{r}_2-\mathbf{r}_1|}\exp\left[i\omega\left(t-\frac{|\mathbf{r}_2-\mathbf{r}_1|}{c_0}\right)\right]\times\exp\left(-\omega c_i\frac{|\mathbf{r}_2-\mathbf{r}_1|}{c_0^2}\right). \tag{18}$$

Assuming again that $\langle S(\mathbf{r}_s,t_s)S(\mathbf{r}_{s'},t_{s'})\rangle=Q^2\delta(t_s-t_{s'})\delta(\mathbf{r}_s-\mathbf{r}_{s'})$ and using the same development as in Eqs. (4) to (8), we have now

$$\langle C_{1,2}(t)\rangle=\frac{Q^2}{4\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{\tau}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{d\mathbf{r}_s dt_s d\omega d\omega' d\tau}{|\mathbf{r}_2-\mathbf{r}_s||\mathbf{r}_1-\mathbf{r}_s|}\times\exp\left[i\omega\left(\tau-t_s-\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c_0}\right)+i\omega'\times\left(t+\tau-t_s-\frac{|\mathbf{r}_2-\mathbf{r}_s|}{c_0}\right)\right]\times\exp\left[-c_i\left(\omega\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c_0^2}+\omega'\frac{|\mathbf{r}_2-\mathbf{r}_s|}{c_0^2}\right)\right]. \tag{19}$$

After a change of variable $\tau-t_s=\tau'$, we perform the integration on τ' knowing that $\int_0^{+\infty}\exp[i(\omega+\omega')\tau']d\tau'=\delta(\omega+\omega')$, which leads to

$$\langle C_{1,2}(t)\rangle=\frac{Q^2}{4\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{d\tau d\mathbf{r}_s d\omega}{|\mathbf{r}_2-\mathbf{r}_s||\mathbf{r}_1-\mathbf{r}_s|}\times\exp\left[i\omega\left(t+\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c_0}-\frac{|\mathbf{r}_2-\mathbf{r}_s|}{c_0}\right)\right]\times\exp\left[-\frac{\omega c_i}{c_0^2}(|\mathbf{r}_1-\mathbf{r}_s|+|\mathbf{r}_2-\mathbf{r}_s|)\right]. \tag{20}$$

As in Sec. II, the integral over $d\tau$ is changed into the product $T\nu$. It corresponds to the accumulation of noise sources with a creation rate ν ($\mathrm{m}^{-3}\,\mathrm{s}^{-1}$) over the finite duration signals of length T recorded in A and B. Then, we have

$$\langle C_{1,2}(t)\rangle=\frac{Q^2T\nu}{4\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{d\mathbf{r}_s d\omega}{|\mathbf{r}_2-\mathbf{r}_s||\mathbf{r}_1-\mathbf{r}_s|}\times\exp\left[i\omega\left(t+\frac{|\mathbf{r}_1-\mathbf{r}_s|}{c_0}-\frac{|\mathbf{r}_2-\mathbf{r}_s|}{c_0}\right)\right]\times\exp\left[-\frac{\omega c_i}{c_0^2}(|\mathbf{r}_1-\mathbf{r}_s|+|\mathbf{r}_2-\mathbf{r}_s|)\right]. \tag{21}$$

Equation (21) is the equivalent of Eq. (9) in the presence of attenuation. We recognize in the two exponentials the hyperbola spatial dependence for the phase term and the ellipse spatial dependence for the amplitude term (Fig. 2). We apply then the change of variable done in Sec. II [Eq. (11)] to decouple the phase term from the amplitude term in Eq. (21). It follows

$$\langle C_{1,2}(t)\rangle=\frac{Q^2T\nu}{2\pi}\int_{-\infty}^{+\infty}d\omega\int_0^{+\infty}d\varphi\,\sinh(\varphi)\times\exp\left(-\frac{2a\cosh(\varphi)\omega c_i}{c_0^2}\right)\int_{-a}^{+a}dx\times\exp\left[i\omega\left(t+\frac{2x}{c_0}\right)\right]. \tag{22}$$

The presence of attenuation in the medium makes the integral over φ converge as

$$\int_0^{+\infty}d\varphi\,\sinh(\varphi)\exp\left(-\frac{2a\cosh(\varphi)\omega c_i}{c_0^2}\right)=\frac{c_0^2}{2a\omega c_i}\exp\left(-\frac{2a\omega c_i}{c_0^2}\right), \tag{23}$$

while the integral over x gives

$$\int_{-a}^{+a}dx\exp\left[i\omega\left(t+\frac{2x}{c_0}\right)\right]=\frac{c_0}{2i\omega}\left[\exp\left(i\omega\left(t+\frac{2a}{c_0}\right)\right)-\exp\left(i\omega\left(t-\frac{2a}{c_0}\right)\right)\right]. \tag{24}$$

Finally, combining Eqs. (23) and (24), it follows

$$\langle C_{1,2}(t)\rangle=\frac{Q^2T\nu c_0^3}{8\pi a c_i}\int_{-\infty}^{+\infty}d\omega\,\frac{1}{i\omega}\left[\exp\left(i\omega\left(t+\frac{2a}{c_0}\right)\right)-\exp\left(i\omega\left(t-\frac{2a}{c_0}\right)\right)\right]\frac{1}{\omega}\exp\left(-\frac{2a\omega c_i}{c_0^2}\right). \tag{25}$$

From the result derived in Sec. I, we know that the time derivative of the noise correlation function yields the Green's function. In the case of a medium with attenuation, the time derivative of the correlation function gives

$$\frac{d}{dt}\langle C_{1,2}(t)\rangle = \frac{Q^2 T \nu c_0^3}{4\pi c_i}\left\{ \int_{-\infty}^{+\infty} \frac{d\omega}{\omega}\frac{1}{2a}\exp\left(i\omega\left(t+\frac{2a}{c_0}\right)\right)\right.$$
$$\times\exp\left(-\frac{2a\omega c_i}{c_0^2}\right) - \int_{-\infty}^{+\infty}\frac{d\omega}{\omega}\frac{1}{2a}$$
$$\left.\times\exp\left(i\omega\left(t-\frac{2a}{c_0}\right)\right)\exp\left(-\frac{2a\omega c_i}{c_0^2}\right)\right\}. \tag{26}$$

The two integrals in Eq. (26) contain a frequency-dependent term $1/\omega$ that prevents us from directly identifying these integrals as the exact Green's function from A to B and B to A. Actually, this $1/\omega$ is due to the choice made to describe the attenuation in the medium. Adding an imaginary part to the sound speed ($c = c_0 + ic_i$) means that we made the assumption of a linear dependence of the attenuation with frequency, as shown in the Green's function formulation [Eq. (18)]. If we had chosen a ω^n dependence of the attenuation in the Green's function, we would have obtained a $1/\omega^n$ amplitude term in the derivative of the correlation function [Eq. (26)]. Physically speaking, this means that the attenuation in the medium will impact the estimate of the Green's function from the ambient noise correlation function. More precisely, attenuation acts as a low-pass filter whose frequency behavior follows the frequency dependence of the attenuation in the medium. Thus, the physical interpretation of Eq. (26) is still the same as Eq. (17)

$$\frac{d}{dt}\langle C_{1,2}(t)\rangle \approx 4\pi a Q^2 T \nu \Omega[G(\mathbf{r}_2,0;\mathbf{r}_1,-t)$$
$$-G(\mathbf{r}_1,t;\mathbf{r}_2,0)], \tag{27}$$

where the $\approx$ sign means that the Green's function has been low-pass filtered according to the frequency dependence of the medium attenuation. Comparing Eqs. (26) and (18), we note that the amplitude term Ω is directly related to the formulation of the attenuation term used in of the definition of the Green's function.

We see from the mathematical developments made in Sec. III that the introduction of a small attenuation in the medium makes the correlation function converge without any constraint on the noise source statistics. The final result is the same as in Sec. II, except that the frequency filtering occurs when a frequency-dependent attenuation is introduced. The derivative of the ambient noise correlation function gives birth to a causal and anticausal (or time-reversed) estimate of the Green's function between the two points at which noise has been recorded.

IV. DISCUSSION

In the literature, most experimental results have been obtained by using the noise correlation function (and not its derivative) as a close estimation of the Green's function. Only Weaver's results in ultrasonic reverberant cavities[1] have clearly demonstrated, both theoretically and experimentally, that the Green's function would be retrieved from the diffuse noise correlation function derivative. In similar works

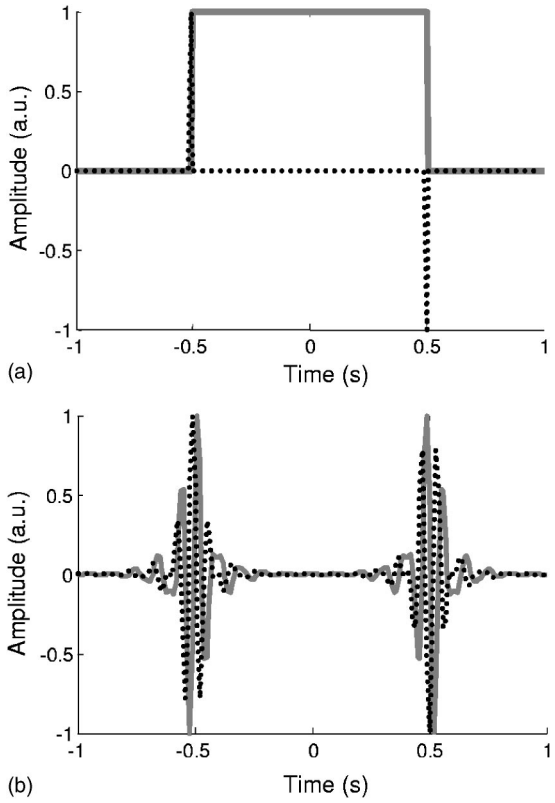


FIG. 3. Representation of the noise correlation function (gray line) and its time derivative (black dotted line): (a) in an infinite bandwidth case, with $a = 0.25$ m and $c = 1$ m/s; (b) with the same parameters, in a limited bandwidth case [10–20 Hz].

in ultrasonics,[8] underwater acoustics,[2] and geophysics,[4] the correlation function is preferred to its derivative to approximate for the Green's function. Indeed, performing a time derivative on experimental data is usually avoided because it could be the source of strong undesirable noise. However, the mathematical demonstration above clearly shows that it is the derivative of the ambient noise correlation function that converges to the Green's function. What do we lose when we don't perform the time derivative?

An element of response is given in Fig. 3, where ambient noise correlation functions $\langle C_{1,2}(t)\rangle$ are plotted versus their derivatives for a infinite bandwidth [Fig. 3(a)] and a limited bandwidth signal [Fig. 3(b)], respectively. The two functions look very different in the infinite bandwidth case, mostly because the zero-frequency component creates the plateau of the correlation function. However, this dc component will usually not be available in realistic experiments. In the case of a finite bandwidth problem, we see that the two functions resemble each other. Their principal difference is a $\pi/2$ phase shift that does not affect the overall shape of the waveform but that could be of importance in the case of tomography applications where exact arrival times need to be estimated. However, if undesired noise becomes an experimental issue when performing the time derivative of the correlation function, it is not a bad approximation to estimate the Green's function as the noise correlation function itself.

Finally, how could the theoretical demonstration done in Sec. III be adapted to the case of a heterogeneous medium? For example, if we assume a spatial dependence of the sound

speed $c_0 = c_0(x,y,z)$, is the final result in Eq. (27) still correct? It is possible to show that, for uncorrelated noise sources, Eq. (21) is still valid if the arrival times $|\mathbf{r}_1 - \mathbf{r}_s|/c_0$ and $|\mathbf{r}_2 - \mathbf{r}_s|/c_0$ are changed into a more general formulation $\int_{\mathbf{r}_s}^{\mathbf{r}_1} ds/c(s)$ and $\int_{\mathbf{r}_s}^{\mathbf{r}_2} ds'/c(s')$, where the paths S and S' between the noise source in $\mathbf{r}_s$ and the receivers in A and B are given by the Fermat principle. In this case, the noise sources that contribute to the noise correlation function at a given time t have to satisfy the following equation:

$$\int_{\mathbf{r}_s}^{\mathbf{r}_1} \frac{ds}{c(s)} - \int_{\mathbf{r}_s}^{\mathbf{r}_2} \frac{ds'}{c(s')} = t. \tag{28}$$

In general, the noise sources that satisfy Eq. (28) are no longer located on a hyperbola as in Fig. 2 and the change of variable done in Eq. (11) is now irrelevant. However, the physical insight derived from the geometrical interpretation in Fig. 2 is still correct. Actually, two conditions are required for this to be true. First, there must exist, for a given t, a 3D surface made of the noise sources that satisfy Eq. (28), and this set of 3D surfaces must cover the whole 3D space when $-2a \leqslant ct \leqslant 2a$. Second, one point in space must belong to one and only one of these 3D surfaces. Those two conditions mean that there exists an isomorphism between the hyperboloids in the homogeneous space case and the 3D sheets in the heterogeneous medium. If this is so, a conformal transformation could be used to shift from the heterogeneous space to the homogeneous one, in which the integration from Eq. (21) to Eq. (27) is performed before the inverse conformal transformation is made to go back to the heterogeneous space. The existence of such a conformal transformation ensures that we have again in a heterogeneous space

$$\frac{d}{dt}\langle C_{1,2}(t)\rangle \propto G(\mathbf{r}_2,0;\mathbf{r}_1,-t) - G(\mathbf{r}_1,t;\mathbf{r}_2,0). \tag{29}$$

However, such a conformal transformation is not always likely to exist. For example, it is well known in underwater acoustics that refraction could generate multiple arrival times between two points. If a noise source creates multiple echoes in A and B, it means that it belongs to different t-invariant 3D surfaces of the correlation function. There is then no isomorphism between the hyperbola in the homogeneous medium and the t-invariant sheets in the refractive medium. In this case, further analysis is necessary to understand the relationship between the noise correlation function and the Green's function.[5,12]

[1] R. L. Weaver and O. J. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).

[2] P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wavefronts fom acoustic ambient noise in the ocean," J. Acoust. Soc. Am. (submitted).

[3] J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge **18**(8), 957–960 (1999).

[4] N. M. Shapiro and M. Campillo, "Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise," Geophys. Res. Lett. **31**, L07614 (2004).

[5] K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival times structure of the long-time ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. (in press).

[6] O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).

[7] M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).

[8] A. Derode, E. Larose, M. Tanter, J. De Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from the field-field correlations in an open scattering medium (L)," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).

[9] H. Cox, "Spatial correlation in arbitrary noise fields with application to ambient sea noise," J. Acoust. Soc. Am. **54**, 1289–1301 (1973).

[10] B. A. Van Tiggelen, "Green function retrieval and time reversal in a disordered world," Phys. Rev. Lett. **91**, 243904 (2003).

[11] R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).

[12] K. Wapenaar, "Synthesis of an inhomogeneous medium from its acoustic transmission response," Geophysics **68**, 1756–1759 (2003).

On the reciprocity relationship between direct field radiation and diffuse reverberant loading

P. J. Shorter[a] and R. S. Langley[b]
ESI US R&D Inc., 12555 High Bluff Drive, Suite 250, San Diego, California 92130

(Received 23 January 2004; revised 19 July 2004; accepted 7 September 2004)

This analysis is concerned with the derivation of a ''diffuse field'' reciprocity relationship between the diffuse field excitation of a connection to a structural or acoustic subsystem and the radiation impedance of the connection. Such a relationship has been derived previously for connections described by a single degree of freedom. In the present work it is shown that the diffuse–field reciprocity relationship also arises when describing the ensemble average response of connections to structural or acoustic subsystems with uncertain boundaries. Furthermore, it is shown that the existing diffuse–field reciprocity relationship can be extended to encompass connections that possess an arbitrary number of degrees of freedom. The present work has application to (i) the calculation of the diffuse field response of structural–acoustic systems modeled by Finite Elements, Boundary Elements, and Infinite Elements; (ii) the general calculation of the Coupling Loss Factors employed in Statistical Energy Analysis (SEA); and (iii) the derivation of an alternative analysis method for describing the dynamic interactions of coupled subsystems with uncertain boundaries (a generalized ''boundary'' approach to SEA).
[DOI: 10.1121/1.1810271]

PACS numbers: 43.20.Rz, 43.55.Br, 43.30.Jx, 43.28.Lv [RLW] Pages: 85–95

I. INTRODUCTION

The Maxwell–Betti or Rayleigh reciprocity relationship for a linear system is well known in physics and engineering.[1,2] The reciprocity relationship states that for two generalized coordinates *a* and *b* (whose product can be related to work), the work done by *a* on *b* per unit displacement of *a* equals the work done by *b* on *a* per unit displacement of *b*. A separate reciprocity relationship also exists for structural–acoustic systems and relates the diffuse field excitation of a structure to the radiation impedance of the structure. This reciprocity relationship is often referred to as the ''diffuse–field'' reciprocity relationship and was introduced (in different contexts) by Smith[3] and Diestel[4] (the relationship has also been discussed by Shaw[5]). The diffuse field reciprocity relationship states that the magnitude of the blocked force on a resonant structural mode in a diffuse acoustic field is proportional to the resistive radiation impedance of the mode (radiating into an anechoic space). The constant of proportionality is independent of the mode shape and only depends on the overall pressure level in the diffuse field. The Maxwell–Betti reciprocity relationship is true for any system, while the diffuse–field reciprocity relationship is only true in an ensemble average sense (in the form presented by Smith[3] and Diestel[4] the average is taken over an ensemble of uncorrelated point sources located in the farfield).

The diffuse-field reciprocity relationship has found numerous applications including, for example, the diffuse field calibration of microphones,[4] and the reciprocal calculation of structural–acoustic coupling loss factors in Statistical Energy Analysis.[6] However, the diffuse–field reciprocity relationship has only been derived for a structure whose response can be described by a single degree of freedom (Smith's original derivation is restricted to a structure whose response is dominated by a single resonant mode). Further work by Smith[7] has demonstrated that the diffuse–field reciprocity relationship can also be applied to the transverse response of a rigid disk embedded in a thin plate. However, this analysis is also restricted to a single degree of freedom.

In this paper, it is shown that the diffuse–field reciprocity relationship arises naturally when describing the ensemble average response of connections to structural or acoustic subsystems with uncertain boundaries. Furthermore, it is shown that the diffuse–field reciprocity relationship can be extended to encompass connections that possess an arbitrary number of degrees of freedom. This work has application to (i) the calculation of the diffuse field response and transmission loss of structural–acoustic systems modeled by Finite Elements, Boundary Elements, and Infinite Elements, (ii) the calculation of the coupling loss factors employed in SEA, and (iii) the derivation of an alternative analysis method for describing the dynamic interactions of coupled subsystems with uncertain boundaries (a ''boundary'' approach to SEA). The latter application is discussed in more detail in a separate publication.[8]

II. THE RESPONSE OF A SUBSYSTEM

A. Direct and reverberant fields

Consider the subsystem illustrated in Fig. 1. The subsystem consists of a homogeneous domain Ω with a boundary surface Γ and may, for example, represent a structural subcomponent or acoustic cavity that forms part of a larger

[a]Electronic-mail: pj.shorter@esi-group-na.com
[b]Visiting Research Scientist. Permanent address: Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom.

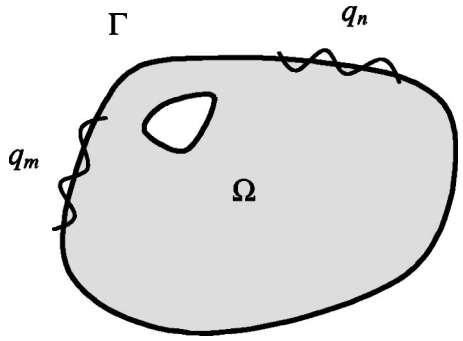


FIG. 1. A subsystem with domain Ω (shaded) and boundary Γ; the boundary response of the subsystem is described by a set of generalized coordinates q.

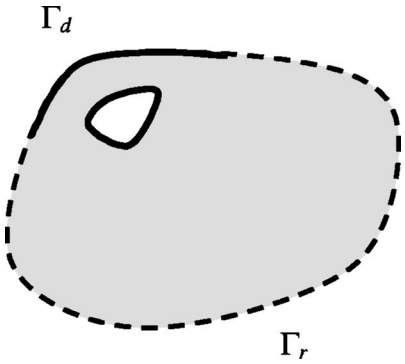


FIG. 2. The boundary of the subsystem is partitioned into a part that is known precisely (termed the deterministic boundary Γ_d), and a part that is known imprecisely (termed the random boundary Γ_r).

system. Steady-state harmonic vibration of frequency ω is considered in what follows and all response quantities are represented by complex amplitudes, so that $\mathbf{u}$, for example, corresponds to the time-dependent response $\mathbf{u}(t) = \mathrm{Re}\{\mathbf{u}\exp(i\omega t)\}$. It is assumed that any excitation of the subsystem occurs through the excitation of certain regions of the boundary (in what follows, no distinction is made between excitation of the boundary due to external loading and excitation due to coupling to adjacent subsystems). If the free-field Green's function for the subsystem is known, then the subsystem response can be described using a boundary integral equation. It is convenient to describe the boundary response in terms of a discrete set of generalized coordinates so that

$$\mathbf{u}(x \in \Gamma) = \sum_k \boldsymbol{\phi}_k(x) q_k , \tag{1}$$

where $\mathbf{u}(x \in \Gamma)$ is the displacement response at a location x on the boundary, $\boldsymbol{\phi}_k$ is a basis function defined across the boundary, and q_k is the amplitude of the kth boundary degree of freedom. A set of equations can then be obtained that relate the generalized force and generalized displacement of each boundary degree of freedom at a given frequency of interest. These equations can be written in the form

$$\mathbf{Hq} = \mathbf{Gf}, \tag{2}$$

where $\mathbf{q}$ is a vector that contains the generalized boundary displacements and $\mathbf{f}$ is a vector that contains the generalized boundary forces. The entries of $\mathbf{H}$ and $\mathbf{G}$ can be obtained in numerous ways; in the current discussion it is assumed that a direct Boundary Element approach is used, as described by Brebbia.[9]

Suppose now that the boundary of the subsystem is partitioned into a part that is known precisely (termed the deterministic boundary) and a part that is known imprecisely (termed the random boundary), as illustrated in Fig. 2. Any regions of the boundary that are subjected to external excitation or connected to other subsystems are assumed to form part of the deterministic boundary (such regions allow energy to be transported into or out of the subsystem). It is noted in passing that the deterministic boundary need not be contiguous or simply connected. The boundary degrees of freedom can then be partitioned into those that describe the response over the deterministic boundary $\mathbf{q}_d$, and those that describe the response across the random boundary $\mathbf{q}_r$ (the basis function associated with a given degree of freedom is assumed to be nonzero only across its specified boundary). The boundary equations of motion for the subsystem can then be partitioned so that

$$\begin{bmatrix} \mathbf{H}_{dd} & \mathbf{H}_{dr} \\ \mathbf{H}_{rd} & \mathbf{H}_{rr} \end{bmatrix} \begin{bmatrix} \mathbf{q}_d \\ \mathbf{q}_r \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{dd} & \mathbf{G}_{dr} \\ \mathbf{G}_{rd} & \mathbf{G}_{rr} \end{bmatrix} \begin{bmatrix} \mathbf{f}_d \\ \mathbf{f}_r \end{bmatrix}. \tag{3}$$

Since the properties of the random boundary are uncertain, it is useful to reduce the random boundary degrees of freedom from the previous equation. The top line of Eq. (3) can then be written as

$$\mathbf{D}_{\mathrm{dir}}\mathbf{q}_d = \mathbf{f}_d + \mathbf{f}_{\mathrm{rev}}, \tag{4}$$

where $\mathbf{D}_{\mathrm{dir}}$ is termed the ''direct field dynamic stiffness matrix'' for the deterministic degrees of freedom and is defined by

$$\mathbf{D}_{\mathrm{dir}} = \mathbf{G}_{dd}^{-1}\mathbf{H}_{dd}. \tag{5}$$

For real-valued basis functions, this matrix is complex and symmetric. The direct field dynamic stiffness matrix gives the dynamic stiffness of the deterministic degrees of freedom in the absence of the random boundary (and can be thought of as being the dynamic stiffness associated with direct field radiation into an infinitely extended subsystem). An examination of Eq. (4) shows that the influence of the random boundary can be viewed as a ''blocked reverberant'' force, $\mathbf{f}_{\mathrm{rev}}$, applied to the deterministic boundary that accounts for reflections of the direct field into a ''reverberant'' field. This blocked reverberant force depends on the boundary conditions applied to the random boundary. For example, for a clamped random boundary ($\mathbf{q}_r = 0$), the blocked reverberant force is given by

$$\mathbf{f}_{\mathrm{rev}} = (-\mathbf{G}_{dd}^{-1}\mathbf{G}_{dr}\mathbf{G}_{rr}^{-1}\mathbf{G}_{rd})\mathbf{f}_d + (\mathbf{G}_{dd}^{-1}\mathbf{G}_{dr}\mathbf{G}_{rr}^{-1}\mathbf{H}_{rd})\mathbf{q}_d , \tag{6}$$

while for a free random boundary ($\mathbf{f}_r = 0$), the blocked reverberant force is given by[10]

$$\mathbf{f}_{\mathrm{rev}} = (-\mathbf{G}_{dd}^{-1}\mathbf{H}_{dr}\mathbf{H}_{rr}^{-1}\mathbf{G}_{rd})\mathbf{f}_d + (\mathbf{G}_{dd}^{-1}\mathbf{H}_{dr}\mathbf{H}_{rr}^{-1}\mathbf{H}_{rd})\mathbf{q}_d . \tag{7}$$

For a given random boundary, the blocked reverberant force is completely coherent[11] with the direct field [for example,

Eqs. (6) and (7) express the coherent relationships that arise for clamped and free random boundaries, respectively].

B. The statistics of the blocked reverberant force

The cross-spectral response of the deterministic boundary degrees of freedom can be found by inverting Eq. (4) to give

$$\mathbf{q}_d\mathbf{q}_d^H = \mathbf{D}_{\mathrm{dir}}^{-1}(\mathbf{f}_d\mathbf{f}_d^H + \mathbf{f}_d\mathbf{f}_{\mathrm{rev}}^H + \mathbf{f}_{\mathrm{rev}}\mathbf{f}_d^H + \mathbf{f}_{\mathrm{rev}}\mathbf{f}_{\mathrm{rev}}^H)\mathbf{D}_{\mathrm{dir}}^{-H}. \quad (8)$$

The first term in the parentheses represents the cross-spectrum of the external loading applied to the deterministic boundary degrees of freedom. The fourth term represents the cross-spectrum of the reverberant loading applied to the deterministic boundary degrees of freedom. The second and third terms are contributions to the cross-spectral response arising from coherence between the external loading and the reverberant loading.

In what follows, it is of interest to investigate the statistics of the previous expression across an ensemble of random boundaries. The expected value of the previous expression is given by

$$\langle \mathbf{q}_d\mathbf{q}_d^H \rangle = \mathbf{D}_{\mathrm{dir}}^{-1}(\mathbf{f}_d\mathbf{f}_d^H + \mathbf{f}_d\langle \mathbf{f}_{\mathrm{rev}}^H \rangle + \langle \mathbf{f}_{\mathrm{rev}} \rangle \mathbf{f}_d^H + \langle \mathbf{f}_{\mathrm{rev}}\mathbf{f}_{\mathrm{rev}}^H \rangle)\mathbf{D}_{\mathrm{dir}}^{-H}, \quad (9)$$

where $\langle \cdot \rangle$ denotes an ensemble average. It can be seen that the expected value of the cross-spectral response of the deterministic boundary degrees of freedom depends on certain ensemble statistics of the reverberant loading. If the reverberant loading has a zero mean when averaged over a large enough ensemble of random boundaries, then the (ensemble average) reverberant loading and the external excitation will be incoherent.

In the following sections, it will be shown that if there is sufficient uncertainty regarding the random boundary, the statistics of the blocked reverberant force tend to the following limits:

$$\langle \mathbf{f}_{\mathrm{rev}} \rangle = 0; \quad \langle \mathbf{f}_{\mathrm{rev}}\mathbf{f}_{\mathrm{rev}}^H \rangle = \alpha\, \mathrm{Im}\{\mathbf{D}_{\mathrm{dir}}\}, \quad (10)$$

where α is a constant that is related to the amplitude of the reverberant field.

III. THE NATURAL COORDINATES FOR A CONNECTION

A. Physical interpretation of the direct and reverberant fields

In principle, one could write explicit expressions for the entries of the matrices **H** and **G**, substitute these expressions into Eq. (8), and then search for the mathematical conditions that give rise to the statistics in Eq. (10). However, the search for the conditions that lead to Eq. (10) can be difficult in the absence of a physical context. The following sections therefore attempt to provide a physical interpretation of the direct and reverberant fields.

The direct field defined in the previous section describes the outgoing displacement field associated with a prescribed displacement of the deterministic boundary, in the absence of the random boundary. This displacement field satisfies the displacement boundary condition across the deterministic boundary but does not (necessarily) satisfy the boundary condition prescribed across the random boundary. The direct field response across the random boundary is, in general, complex and may be denoted as $\boldsymbol{\phi}$, as shown in Fig. 3. In order to satisfy the random boundary condition it is necessary add a second field (termed the reverberant field), which (i) satisfies a blocked (i.e., zero displacement) boundary condition across the deterministic boundary, and (ii) which, when added to the direct field, satisfies a prescribed boundary condition across the random boundary. For example, if the random boundary is clamped then the reverberant field must satisfy a displacement boundary condition of $-\boldsymbol{\phi}$ across the random boundary and a blocked displacement boundary condition across the deterministic boundary, as illustrated in Fig. 4. The blocked reverberant force defined in Eq. (4) can therefore be interpreted as the force that needs to be applied to the deterministic boundary degrees of freedom in order to keep them blocked, when subjected to a prescribed displacement boundary condition across the random boundary.

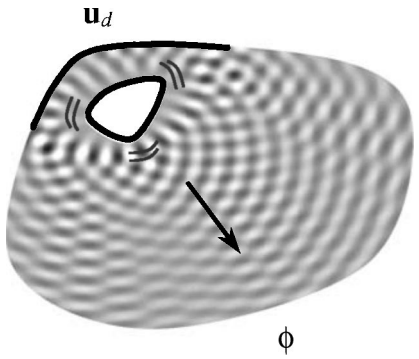


FIG. 3. The direct field due to a displacement of the deterministic boundary.

B. The natural coordinates for the direct field

Thus far, no mention has been made of the specific choice of basis function used to describe the boundary response. For the deterministic boundary a natural set of coordinates exists that is particularly useful for describing the flow of energy in the direct field. These coordinates are obtained by diagonalizing the imaginary[12] (or resistive) part of the direct field dynamic stiffness matrix and are here termed the ''direct field radiation components'' (in analogy with the radiation modes[13,14] defined previously for acoustic radiation

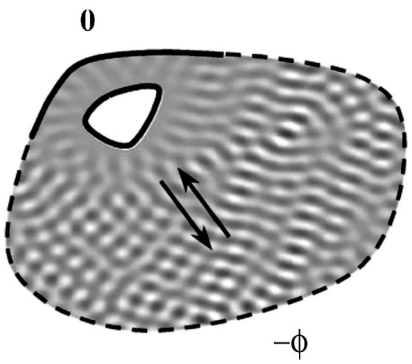


FIG. 4. The reverberant field (associated with scattering of the direct field at the random boundary when the deterministic boundary is blocked).

problems[15]). The direct field radiation components can be found by solving the following eigenvalue problem:

$$\mathrm{Im}\{\mathbf{D}_{\mathrm{dir}}\}\mathbf{P}=\mathbf{P}\boldsymbol{\lambda}, \tag{11}$$

where $\boldsymbol{\lambda}$ is a real diagonal matrix of eigenvalues and $\mathbf{P}$ is a real matrix of eigenvectors. It is convenient to normalize the eigenvectors so that $\mathbf{P}^T\mathbf{P}=\mathbf{I}$. The amplitudes of the various direct field radiation components $\mathbf{a}$ may be determined from the amplitudes of the deterministic boundary degrees of freedom by

$$\mathbf{a}=\mathbf{P}^T\mathbf{q}. \tag{12}$$

The displacement response at a location x in the direct field can then be written as

$$\mathbf{u}_{\mathrm{dir}}(x)=\sum_k \boldsymbol{\phi}_k^{\mathrm{dir}}(x)a_k, \tag{13}$$

where $\boldsymbol{\phi}_k^{\mathrm{dir}}$ represents the displacement field associated with the kth direct field radiation component. The displacement field $\boldsymbol{\phi}_k^{\mathrm{dir}}$ is real-valued across the deterministic boundary and, in general, complex within the domain. The time average energy flow into the direct field (in the absence of the random boundary) is then given by

$$P_{\mathrm{dir}}=\frac{1}{2}\mathrm{Re}\{(i\omega\mathbf{q})^H\mathbf{D}_{\mathrm{dir}}\mathbf{q}\}=\frac{\omega}{2}\sum_k \lambda_k|a_k|^2. \tag{14}$$

For a passive subsystem, the resistive component of the direct field dynamic stiffness matrix is positive semidefinite so that the eigenvalues $\boldsymbol{\lambda}$ are greater than or equal to zero.

Radiation components with zero eigenvalues indicate motions of the deterministic boundary that generate a local evanescent field with no net transport of energy (a localized ''sloshing'' of energy in the vicinity of the connection). Radiation components with nonzero eigenvalues indicate motions of the deterministic boundary that radiate or transport energy into the direct field of the subsystem (in general, such motions also involve a local reactive component). A finite-sized connection typically has a small number of dominant direct field radiation components (regardless of the number of degrees of freedom used to describe the connection). The remaining radiation components tend to be weak radiators (consider, for example, the radiation impedance of the spherical harmonics of a finite sphere[13,17]).

C. The natural coordinates for the reverberant field

Consider now the choice of coordinates for describing the reverberant field. Since interest lies in describing the way in which the direct field is scattered (and reflected back to the blocked connection), it would seem natural to choose a set of coordinates that can be easily related to the direct field radiation components. Since the direct field radiation components provide a good basis for describing outgoing energy flows in the farfield of a subsystem, one might postulate that an additional set of shape functions are required that describe energy flowing from the random boundary back towards the deterministic boundary. One way to obtain such shape functions is to perform a time reversal of the outgoing direct field radiation components. In the current context, this equates to a set of displacement fields obtained from the complex conjugates of the outgoing direct field radiation components. These displacement fields can be thought of as incoming boundary waves that are power orthogonal to the outgoing direct field radiation components. As discussed in what follows, a combination of the incoming boundary waves and the outgoing direct field radiation components provides a particularly useful basis for describing the reverberant field.

Consider a given incoming ''boundary wave'' incident upon the deterministic boundary. When the wave reaches the blocked deterministic boundary it is reflected and scattered into a set of outgoing waves. However, due to the orthogonality of the direct field radiation components, each incoming wave is scattered into a single outgoing direct field radiation component. The superposition of the kth incoming wave and the kth outgoing direct field component therefore results in a combined wave that may be termed the kth ''reverberant field component.''[18] The reverberant field may therefore be described in terms of a set of such reverberant field components so that

$$\mathbf{u}_{\mathrm{rev}}(x)=\sum_k \boldsymbol{\phi}_k^{\mathrm{rev}}(x)b_k, \tag{15}$$

where b_k is the amplitude of the kth reverberant field component and where the displacement field associated with the kth reverberant field component is given by

$$\boldsymbol{\phi}_k^{\mathrm{rev}}(x)=(\boldsymbol{\phi}_k^{\mathrm{dir},*}(x)-\boldsymbol{\phi}_k^{\mathrm{dir}}(x))=-2\,\mathrm{Im}\{\boldsymbol{\phi}_k^{\mathrm{dir}}(x)\}i. \tag{16}$$

It is noted in passing that the displacement field $\boldsymbol{\phi}_k^{\mathrm{rev}}$ is zero across the deterministic boundary and therefore naturally satisfies the blocked boundary condition across the deterministic boundary.

Consider now the blocked force on the deterministic boundary degrees of freedom due to the kth reverberant field component. From Eq. (12), the force on the deterministic boundary degrees of freedom due to the generation of a set of outgoing radiation components $\mathbf{a}$ is given by

$$\mathbf{f}=\mathbf{D}_{\mathrm{dir}}\mathbf{P}\mathbf{a}. \tag{17}$$

From the previous discussion, the blocked reverberant force can be viewed as being the superposition of the force generated by a set of outgoing and incoming radiation components so that

$$\begin{aligned}\mathbf{f}_{\mathrm{rev}}&=((\mathbf{D}_{\mathrm{dir}}\mathbf{P})^*-\mathbf{D}_{\mathrm{dir}}\mathbf{P})\mathbf{b}=(\mathbf{D}_{\mathrm{dir}}^*-\mathbf{D}_{\mathrm{dir}})\mathbf{P}\mathbf{b}\\&=-2i\,\mathrm{Im}\{\mathbf{D}_{\mathrm{dir}}\}\mathbf{P}\mathbf{b}.\end{aligned} \tag{18}$$

Inserting Eq. (11) into the previous expression gives

$$\mathbf{f}_{\mathrm{rev}}=-2i\mathbf{P}\boldsymbol{\lambda}\mathbf{b}. \tag{19}$$

The cross-spectrum of the blocked force is therefore given by

$$\mathbf{f}_{\mathrm{rev}}\mathbf{f}_{\mathrm{rev}}^H=4\mathbf{P}\boldsymbol{\lambda}\mathbf{b}\mathbf{b}^H\boldsymbol{\lambda}\mathbf{P}^T. \tag{20}$$

The statistics of the blocked reverberant force can therefore be calculated if the statistics of the reverberant field components can be found.

IV. BOUNDARY CONDITIONS AT THE RANDOM BOUNDARY

A. The constraint between the direct and reverberant fields

The amplitudes of the direct and reverberant field components are not entirely independent but are instead constrained to satisfy the boundary condition applied at the random boundary. In general, the constraint at the random boundary can be phrased as

$$\mathbf{C}_1\mathbf{a}+\mathbf{C}_2\mathbf{b}=0. \tag{21}$$

The entries of the constraint matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ can be obtained in a number of ways using, for example, collocation or a Galerkin Weighted Residual approach. For a passive random boundary, the response of the reverberant field components can be fully determined from the direct field components, so that

$$\mathbf{b}=\mathbf{T}_0\mathbf{a}; \quad \mathbf{T}_0=-\mathbf{C}_2^{-1}\mathbf{C}_1. \tag{22}$$

The matrix $\mathbf{T}_0$ is complex and can be viewed as a scattering matrix that describes the scattering of the direct field at the random boundary. If the location of the random boundary is uncertain, then (for two- and three-dimensional subsystems), there exists a large number of possible reverberant field components into which the direct field can be scattered. Consider, for example, acoustic radiation from the spherical harmonics of a finite sphere. While only a small number of such harmonics are efficient radiators, the outgoing energy in a single direct field radiation component may be scattered into a large number of incoming harmonics at the random boundary.

It is convenient in what follows to adopt an alternative set of coordinates that are normalized to have unit incident power. A set of coordinates **c** and **d** can therefore be defined so that

$$c_k=a_k\sqrt{\lambda_k}; \quad d_k=b_k\sqrt{\lambda_k}, \tag{23}$$

where λ_k is the eigenvalue associated with the kth direct field component. The constraint in Eq. (22) can then be written as

$$\mathbf{d}=\mathbf{T}\mathbf{c}; \quad \mathbf{T}=-\boldsymbol{\lambda}^{1/2}\mathbf{C}_2^{-1}\mathbf{C}_1\boldsymbol{\lambda}^{-1/2}. \tag{24}$$

For a single random boundary curve, the amplitudes of the (normalized) direct field radiation components and reverberant field components will be entirely coherent. If there is uncertainty in the precise details of the random boundary (both in terms of location and applied boundary condition), then the entries of the matrix **T** become random variables. As the boundary curve is perturbed, the coherent relationships between the various entries of **T** are also perturbed. It is therefore of interest to consider the statistics of **T** across a large ensemble of possible boundary curves.

B. Maximum entropy and the statistics of a diffuse reverberant field

At first sight it might appear that no meaningful statements can be made about the statistics of the scattering matrix in Eq. (24) without a more precise definition of the ensemble of random boundaries under consideration. However, defining an ensemble (by specifying the statistics of **T**), implicitly defines some information about the random boundary. Of all possible ensemble definitions, there exists a unique ensemble that provides the minimum amount of information about the random boundary. This ensemble has the maximum uncertainty, or maximum entropy[19] of any possible ensemble of random boundaries.

One way to establish the statistics of the scattering matrix for such an ensemble is to consider the way in which the statistics of **T** change when the basis defining **c** and **d** is changed. Suppose, for example, that an alternative power orthogonal basis is adopted for describing the direct and reverberant fields so that

$$\hat{\mathbf{c}}=\mathbf{R}_1\mathbf{c}; \quad \hat{\mathbf{d}}=\mathbf{R}_2\mathbf{d}, \tag{25}$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ represent transformation matrices. In what follows, transformations that conserve the overall power in the direct field and overall energy (or an equivalent quadratic quantity) in the reverberant field are considered so that

$$\mathbf{R}_1^H\mathbf{R}_1=\mathbf{I}; \quad \mathbf{R}_2^H\mathbf{R}_2=\mathbf{I}. \tag{26}$$

It follows from Eqs. (24) and (25) that the scattering matrix in the new basis is given by

$$\hat{\mathbf{T}}=\mathbf{R}_2\mathbf{T}\mathbf{R}_1^H. \tag{27}$$

The mean value of the scattering matrix in the new basis is then given by

$$\langle\hat{\mathbf{T}}\rangle=\mathbf{R}_2\langle\mathbf{T}\rangle\mathbf{R}_1^H. \tag{28}$$

If the scattering matrix has a nonzero mean then the statistics of the scattering that occurs at the random boundary will be basis dependent; adopting an alternative basis will therefore provide a different set of statistics. In principle, an investigation of different sets of bases would then provide some information about the scattering process at the random boundary. This therefore leads to the premise that *the minimum information about the random boundary will be provided by an ensemble in which the statistical properties of the scattering matrix* **T** *are invariant to a change in the basis used to define either* **c** *or* **d**. Equally, this premise can be viewed as an existence statement for a set of universal statistical properties for the matrix **T**. For this premise to apply, Eq. (28) requires that the ensemble mean of the scattering matrix must equal zero so that

$$\langle T_{mn}\rangle=0. \tag{29}$$

Consider now the second-order statistics of **T**. It follows from Eq. (27) that

$$\langle\hat{T}_{jk}\hat{T}_{rs}^*\rangle=\sum_{mnpq}\langle T_{mn}T_{pq}^*\rangle R_{2,jm}R_{2,rp}^*R_{1,sq}R_{1,kn}^*. \tag{30}$$

Given the properties of the transformation matrices represented by Eq. (26), it follows that the second-order statistics will be independent of the chosen basis if

$$\langle T_{mn}T_{pq}^*\rangle=C'\,\delta_{mp}\delta_{nq}, \tag{31}$$

where C' is a constant.

In summary, the condition for maximum entropy has been interpreted as a situation in which the statistics of the scattering matrix are invariant to a change of basis. The sta-

tistical properties of **T** that give rise to such a result have been obtained and are represented by Eqs. (29) and (31). The implications of these statistics on the nature of the reverberant field are discussed in what follows.

Inserting Eqs. (29) and (31) into Eq. (24) gives

$$\langle d_m d_n^* \rangle = \delta_{mn} C; \quad \langle d_m c_n^* \rangle = 0; \quad \langle d_m \rangle = 0, \tag{32}$$

where C is a constant related to the incident power in the direct field. The state of maximum entropy therefore corresponds to a situation in which the power amplitudes **d** of the reverberant field components are incoherent (with each other and with the direct field radiation components), and have equal magnitude and random phase. Inserting Eq. (23) into Eq. (32) gives

$$\langle b_m b_n^* \rangle = \frac{C}{\lambda_m} \delta_{mn}; \quad \langle a_m b_n^* \rangle = 0; \quad \langle b_m \rangle = 0. \tag{33}$$

Reverberant fields that possess the (ensemble) statistics defined in Eq. (33) are here termed ''diffuse.'' Each reverberant field component in a diffuse field is incoherent and has the same incident power on the deterministic boundary. Such a definition of a diffuse field can be viewed as being the boundary equivalent of Lyon's ''maximum randomness'' modal description of a diffuse field.[20] It should be noted that the diffuse reverberant field discussed here is a statistical property of an ensemble of subsystems; the response of an individual subsystem need not be ''diffuse'' (however, it is noted in passing that similar statistics may arise for certain individual subsystems when an average is performed over frequency rather than across an ensemble). Conceptually, one may view the imposition of the random boundary condition as being equivalent to applying spatially incoherent (rain-on-the-roof) excitation to the reverberant field components (arising from the use of point collocation to enforce the random boundary condition across the ensemble).

The previous result is based on an assumption that the direct field can be scattered into a large number of possible reverberant field components. A diffuse reverberant field may not arise for ensembles of subsystems that possess a limited number of reverberant field components. For example, an ensemble of one-dimensional waveguides with a single propagating wave type may not possess a sufficient amount of uncertainty[21] to attain a diffuse reverberant field (as discussed by Mace[22] and Scharton[23]).

C. The statistics of a diffuse blocked reverberant force

Inserting Eq. (33) into Eq. (19) shows that

$$\langle \mathbf{f}_{\text{rev}} \rangle = -2i\mathbf{P}\boldsymbol{\lambda}\langle \mathbf{b} \rangle = 0. \tag{34}$$

Inserting Eq. (33) into Eq. (20) shows that

$$\langle \mathbf{f}_{\text{rev}} \mathbf{f}_{\text{rev}}^H \rangle = 4\mathbf{P}\boldsymbol{\lambda}\langle \mathbf{b}\mathbf{b}^H \rangle \boldsymbol{\lambda}\mathbf{P}^T = 4C\mathbf{P}\boldsymbol{\lambda}\mathbf{P}^T = \alpha\, \text{Im}\{\mathbf{D}_{\text{dir}}\}, \tag{35}$$

where the constant of proportionality α describes the amplitude of the diffuse reverberant field (the constant α is described in more detail in the following section). The previous expression can be viewed as the extension of the diffuse–field reciprocity principle to connections described by an arbitrary number of degrees of freedom; this concludes the derivation of Eq. (10).

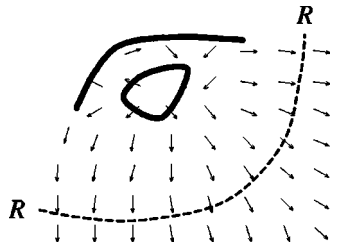


FIG. 5. The normal of the enclosing surface R (dashed line) is aligned with the local intensity field (arrows).

D. The incident power in a diffuse reverberant field

It was noted in the previous section that the constant α describes the amplitude of the diffuse reverberant field. There are numerous ways in which one might define this amplitude. For example, Smith[3] discusses the use of a reference position in the farfield (and notes that ''*It is to be expected that a reasonable analyst will avoid choosing locations for which ... [the reference pressure] ... is zero.*''). In the following analysis, the amplitude is related to the energetics of the reverberant field. Expressions are obtained for the incident power, energy density, and effective group velocity, by considering the energy flows that occur across a region in the farfield of the deterministic boundary (a similar approach was presented previously by the authors for the specific case of diffuse cylindrical waves[25]). The analysis in the following section enables the amplitude of the reverberant field to be defined in terms of quantities commonly employed in SEA.

Consider the ensemble average energy density and intensity in a diffuse reverberant field. Since the reverberant field components in a diffuse reverberant field are energy and power orthogonal, the energetics of each reverberant field component can be considered separately. The kth reverberant field component can be thought of as the superposition of two waves derived from the kth direct field radiation component. Since the displacement field associated with each direct field component is known, it is possible (in principle) to compute the local energy density and intensity fields for each direct field component. Suppose that such information has been computed; an enclosing surface R can then be defined through which all energy in the farfield flows. Furthermore, suppose that the surface R is chosen so that the surface normal is aligned with the local intensity field, as illustrated in Fig. 5.

The two direct field components associated with the kth reverberant field component transport an equal amount of energy toward and away from the deterministic boundary (so that the net energy flow across R is zero). It is possible to define an effective group velocity $c_{g,k}$ for each component by considering the ratio of the power incident on the surface to the total surface energy density. The total surface energy density $e_{\text{tot},k}$ gives the total energy contained within a strip of unit width in a direction normal to the surface R. The effec-

tive group velocity for the kth reverberant field component is then given by

$$c_{g,k}=\frac{P_{\mathrm{inc},k}}{\frac{1}{2}e_{\mathrm{tot},k}}. \tag{36}$$

The factor of 1/2 in the previous expression arises because half of the total surface energy density is transported by each direct field component. The net incident power on the surface R is the same for each direct field component (from the derivation of Sec. IV B). Reverberant field components with a small energy density will therefore have a large group velocity (and *vice versa*). In principle, the group velocity for each reverberant field component can be calculated explicitly using the previous expression. However, as discussed in what follows, such a calculation is seldom necessary for commonly encountered structural and acoustic subsystems.

It follows from Eqs. (14) and (33) that the incident power associated with each reverberant field component is given by

$$P_{\mathrm{inc},k}=\frac{\omega}{2}C. \tag{37}$$

Inserting the previous expression into Eq. (36) and rearranging gives

$$C=\frac{e_{\mathrm{tot},k}c_{g,k}}{\omega}. \tag{38}$$

Now the total diffuse field energy contained in a strip of unit width centered around R can be written as

$$E_{\mathrm{tot,unit}}=\sum_k e_{\mathrm{tot},k}=\sum_k \frac{\omega C}{c_{g,k}}. \tag{39}$$

The modal density[24] of this strip relating to modes associated with the kth reverberant field component can be written as

$$n_{k,\mathrm{unit}}=\frac{1}{\pi c_{g,k}}. \tag{40}$$

Equation (38) then yields

$$E_{\mathrm{tot,unit}}=\pi\omega C\sum_k n_{k,\mathrm{unit}}=\pi\omega C n_{\mathrm{tot,unit}}, \tag{41}$$

where $n_{\mathrm{tot,unit}}$ is the total modal density of the strip. It then follows that

$$C=\frac{E_{\mathrm{tot,unit}}}{\pi\omega n_{\mathrm{tot,unit}}}. \tag{42}$$

Since Eq. (42) is valid for any unit strip within the subsystem, and both the energy and the modal density are additive over any number of strips, it then follows that Eq. (42) must also hold if the total energy and total modal density are employed, so that

$$C=\frac{E_{\mathrm{tot}}}{\pi\omega n_{\mathrm{tot}}}. \tag{43}$$

The asymptotic modal densities of many commonly encountered structural and acoustic subsystems can be found analytically; in such instances, it is not therefore necessary to compute the group velocities of the individual reverberant field components explicitly using Eq. (36). Inserting Eq. (43) into Eq. (35) gives the general diffuse field reciprocity relationship,

$$\langle \mathbf{f}_{\mathrm{rev}}\mathbf{f}_{\mathrm{rev}}^{H}\rangle=\frac{4E_{\mathrm{tot}}}{\pi\omega n_{\mathrm{tot}}}\mathrm{Im}\{\mathbf{D}_{\mathrm{dir}}\}. \tag{44}$$

In many instances the reverberant field components can be grouped into a number of distinct sets that are assumed to be uncoupled at the random boundary (for example, one might choose to describe the response of a flat plate subsystem in terms of a group of ''flexural'' reverberant field components and a group of ''in-plane'' reverberant field components). In such instances, Eq. (44) may be used to find the reverberant loading associated with each group separately [the direct field dynamic stiffness in Eq. (44) then represents the contribution to the direct field dynamic stiffness arising from a given set of direct field radiation components].

Equations (10) and (44) represent a general reciprocity relationship between the diffuse field excitation of a connection to a subsystem and the radiation impedance (or, in the form presented, the direct field dynamic stiffness) of the connection radiating into an anechoic space. This relationship is the main contribution of this paper. The following section provides a number of examples that demonstrate the application of the reciprocity relation.

V. APPLICATIONS

A. The blocked force at the end of a planar beam

Consider the simple planar beam illustrated in Fig. 6. A diffuse field is assumed to exist in the beam and interest lies in predicting the blocked force that the beam exerts on the blocked connection (for simplicity it is assumed that the beam only supports flexural waves). For this system, a diffuse field can be described by an incident propagating wave with fixed amplitude and random phase. If a propagating wave of displacement amplitude a^+ is incident upon the blocked connection then it can be shown[26] that the force resulting from the incident, reflected and evanescent waves is given by

$$\mathbf{f}=\begin{bmatrix} f \\ m \end{bmatrix}=2Bk^2\begin{bmatrix} k(-1+i) \\ (1-i) \end{bmatrix}a^+, \tag{45}$$

where B is the bending stiffness of the beam and k is the flexural wave number. The cross-spectrum of this force is then given by

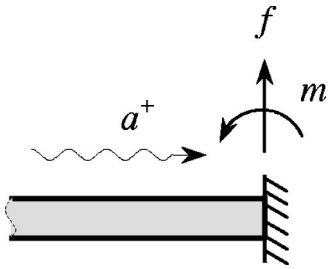


FIG. 6. Force on the end of a blocked beam subjected to a diffuse field.

$$\mathbf{ff}^H = 8B^2k^4 \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} |a^+|^2. \quad (46)$$

The following relationships hold for a beam:

$$E = m\omega^2 l |a^+|^2; \quad n = \frac{l}{\pi c_g}; \quad c_g = \frac{2\omega}{k} = 2k\sqrt{\frac{B}{m}}. \quad (47)$$

Inserting Eq. (47) into Eq. (46) and rearranging gives

$$\mathbf{ff}^H = \frac{4E}{\pi\omega n} Bk \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix}. \quad (48)$$

The previous expression therefore gives the blocked force on the connection, calculated from an incident wave analysis. For comparison, the diffuse field reciprocity relationship can also be used to calculate the blocked force on the connection. The dynamic stiffness matrix at the right-hand end of a semi-infinite beam is given by[26]

$$\mathbf{D}_{\text{dir}} = Bk \begin{bmatrix} (-1+i)k^2 & -ik \\ -ik & (1+i) \end{bmatrix}. \quad (49)$$

Inserting the previous expression into the diffuse field reciprocity relation in Eq. (44) gives

$$\langle \mathbf{f}_{\text{rev}} \mathbf{f}_{\text{rev}}^H \rangle = \frac{4E_{\text{tot}}}{\pi\omega n_{\text{tot}}} \text{Im}\{\mathbf{D}_{\text{dir}}\} = \frac{4E_{\text{tot}}}{\pi\omega n_{\text{tot}}} Bk \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix}. \quad (50)$$

A comparison of Eqs. (50) and (48) shows that the blocked force calculated with the reciprocity result agrees with the blocked force calculated using a traditional wave approach.

B. The SEA coupling loss factors between coupled subsystems

In Statistical Energy Analysis (SEA), the flow of energy between two coupled subsystems is described in terms of a parameter termed a coupling loss factor. The coupling loss factor η_{ij} is a statistical parameter that describes the input power to the direct field of the *j*th subsystem, due to the presence of a diffuse reverberant field with unit energy in the *i*th subsystem, so that

$$\eta_{ij} = \frac{P_{\text{dir},j}}{\omega E_{\text{tot},i}}. \quad (51)$$

The use of the diffuse field reciprocity relationship in Statistical Energy Analysis is discussed in more detail in a separate publication by the authors.[8] However, by applying the diffuse field reciprocity relationship to Eq. (51), it can be shown that the SEA coupling loss factor between two subsystems *i* and *j* can be written as

$$\eta_{ij} = \frac{2}{\pi\omega n_i} \sum_{rs} \text{Im}\{\mathbf{D}_{\text{dir}}^{(j)}\}_{rs} (\mathbf{D}_{\text{tot}}^{-1} \text{Im}\{\mathbf{D}_{\text{dir}}^{(i)}\} \mathbf{D}_{\text{tot}}^{-H})_{rs}, \quad (52)$$

where $\mathbf{D}_{\text{dir}}^{(i)}$ is the dynamic stiffness of the direct field of the *i*th subsystem and where $\mathbf{D}_{\text{tot}}$ is the total dynamic stiffness of the junction (the dynamic stiffness of the junction and the direct fields of all connected subsystems). For a connection with a single degree of freedom, the previous expression becomes

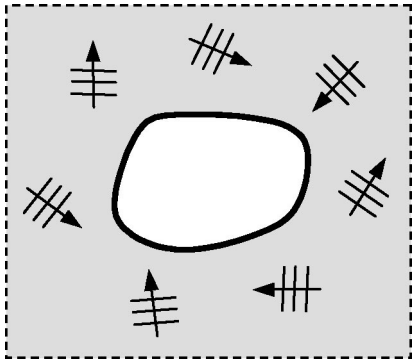


FIG. 7. General structure immersed in a diffuse acoustic field.

$$\eta_{ij} = \frac{2}{\pi\omega n_i} \frac{\text{Im}\{D_{\text{dir}}^{(j)}\} \text{Im}\{D_{\text{dir}}^{(i)}\}}{|D_{\text{tot}}|^2}, \quad (53)$$

which is in agreement with existing expressions for the coupling loss factor of point connected subsystems.[26,27] Equation (52) is, however, quite general and can be used to calculate the coupling loss factors of complex junctions described by multiple degrees of freedom. The diffuse field reciprocity approach therefore provides an efficient alternative to the traditional wave and mobility approaches for calculating SEA coupling loss factors.

C. Structural response to a diffuse sound field

Consider the problem of determining the response of a finite structure subjected to a diffuse acoustic field (with a farfield mean square pressure $\langle \overline{p^2} \rangle$) as illustrated in Fig. 7. The fluid may be bounded or unbounded (in the former instance, the dimensions of the fluid volume are assumed to be large or uncertain in comparison with an acoustic wavelength). This problem is encountered in numerous applications ranging from the qualification of launch vehicle payloads to the response of underwater structures. The traditional (low-frequency) approach to this problem is to couple a finite element model of the structure with a boundary element model of the fluid. The diffuse field excitation of the structure is then typically approximated by either (i) a set of incoherent point sources distributed across a sphere in the farfield or (ii) a set of incoherent plane waves applied at various discrete headings. The accuracy with which this approach describes a diffuse field is dependent on the number of uncorrelated sources that are used. The diffuse field reciprocity result derived in this paper provides an alternative approach to this problem.

The *in-vacuo* dynamic stiffness matrix for the structure (at a given frequency of interest) can be denoted as $\mathbf{D}_s$. This matrix can be found in numerous ways using, for example, a finite element model or an experimentally measured modal model (the structure need not be conservative and the matrix $\mathbf{D}_s$ is, in general, complex). The choice of degrees of freedom used in the analysis is arbitrary (the degrees of freedom may, for example, be nodal degrees of freedom or *in-vacuo* modal degrees of freedom); a sufficient number of degrees of freedom are, however, required for convergence. The dynamic stiffness that the fluid presents to the structure at a

given frequency of interest may be obtained in a variety of ways using, for example, a boundary element model or a combined finite and infinite element model of the fluid. The fluid dynamic stiffness can then be projected onto the structural degrees of freedom and the resulting dynamic stiffness matrix denoted as $\mathbf{D}_a$. The cross-spectral response of the structural degrees of freedom is then given by

$$\langle \mathbf{q}_d \mathbf{q}_d^H \rangle = [\mathbf{D}_a + \mathbf{D}_s]^{-1} \mathbf{S}_{ff} [\mathbf{D}_a + \mathbf{D}_s]^{-H}, \tag{54}$$

where the matrix $\mathbf{S}_{ff}$ gives the blocked force on the structural degrees of freedom due to the diffuse field. Using the diffuse field reciprocity result of Eq. (44), the blocked force associated with a diffuse reverberant field in the fluid can be written as

$$\mathbf{S}_{ff} = \frac{4E_{\text{tot}}}{\pi \omega n_{\text{tot}}} \text{Im}\{\mathbf{D}_a\}. \tag{55}$$

For this particular application, it is usually more convenient to describe the amplitude of the diffuse field in terms of the mean square pressure in the farfield. The total energy and modal density of a large three-dimensional acoustic cavity can be written as[24]

$$E = V \frac{\langle \overline{p^2} \rangle}{\rho c^2}; \quad n = \frac{V \omega^2}{2\pi^2 c^3}, \tag{56}$$

where $\langle \overline{p^2} \rangle$ is the mean square reverberant pressure in the farfield (nb: the mean square pressure is related to the magnitude of the pressure by $\overline{p^2} = 1/2|p|^2$). Inserting Eq. (56) into Eq. (55) gives

$$\mathbf{S}_{ff} = \langle \overline{p^2} \rangle \frac{8\pi c}{\rho \omega^3} \text{Im}\{\mathbf{D}_a\}. \tag{57}$$

Inserting the previous expression into Eq. (54) gives an expression for the cross-spectral response of the structural degrees of freedom in a diffuse acoustic field,

$$\langle \mathbf{q}_d \mathbf{q}_d^H \rangle = \langle \overline{p^2} \rangle \frac{8\pi c}{\rho \omega^3} [\mathbf{D}_a + \mathbf{D}_s]^{-1} \text{Im}\{\mathbf{D}_a\} [\mathbf{D}_a + \mathbf{D}_s]^{-H}. \tag{58}$$

Equation (58) provides a general way to compute the response of a structure to a diffuse acoustic field and is the main result of this section (it is noted in passing that a similar approach may be used to calculate the diffuse field transmission loss of a system).

It is possible to demonstrate the validity of the previous result analytically for structures that are planar and contained within an infinite rigid baffle. Consider a planar structure whose displacement response is described by a finite number of degrees of freedom with real-valued basis functions $\boldsymbol{\phi}_n$. The expected value of the cross-spectral force matrix due to a distributed random pressure loading is given by

$$S_{ff,mn} = \int_s \int_{s'} \langle p(s) p^*(s') \rangle \boldsymbol{\phi}_m(s) \boldsymbol{\phi}_n(s') ds\, ds'. \tag{59}$$

The expected pressure cross-spectrum on the surface of a rigid infinite plane in a diffuse acoustic field (with farfield mean square pressure $\langle \overline{p^2} \rangle$) is given by[28]

$$\langle p(s) p^*(s') \rangle = 4 \langle \overline{p^2} \rangle \frac{\sin(kr)}{kr}. \tag{60}$$

The factor of 4 in the previous expression arises from the pressure doubling that occurs across a rigid surface in a diffuse acoustic field (and from the relationship between the pressure magnitude and mean square value). Inserting Eq. (60) into Eq. (59) gives

$$S_{ff,mn} = \frac{4\langle \overline{p} \rangle^2}{k} \int_s \int_{s'} \frac{\sin(kr)}{r} \boldsymbol{\phi}_m(s) \boldsymbol{\phi}_n(s') ds\, ds'. \tag{61}$$

Consider now the calculation of the cross-spectral force matrix using the diffuse-field reciprocity approach. For a planar radiator, the pressure response associated with a unit displacement of the mth basis function can be found using the Rayleigh integral[24] and is given by

$$p(x) = \frac{-\omega^2 \rho}{2\pi} \int_s \frac{\boldsymbol{\phi}_n(x') \exp(-ikr)}{r} dx', \tag{62}$$

where $r = |x - x'|$. The mnth entry of the direct field dynamic stiffness matrix gives the blocked force on the nth basis function due to a unit displacement of the mth basis function so that

$$\begin{aligned} D_{a,mn} &= \int_{s'} \boldsymbol{\phi}_m(x) p(x) dx \\ &= \frac{-\omega^2 \rho}{2\pi} \int_{s'} \int_s \frac{\exp(-ikr)}{r} \boldsymbol{\phi}_m(x) \boldsymbol{\phi}_n(x') dx\, dx'. \end{aligned} \tag{63}$$

From the diffuse field reciprocity theorem expressed in Eq. (57), the blocked force is given by

$$S_{ff,mn} = \langle \overline{p^2} \rangle \frac{8\pi c}{\rho \omega^3} \text{Im}\{D_{a,mn}\}, \tag{64}$$

so that

$$S_{ff,mn} = \frac{4\langle \overline{p^2} \rangle}{k} \int_{s'} \int_s \frac{\sin(kr)}{r} \boldsymbol{\phi}_m(x) \boldsymbol{\phi}_n(x') dx\, dx'. \tag{65}$$

The previous expression is in agreement with Eq. (61) and demonstrates the use of the diffuse field reciprocity theorem for calculating the response of planar structures to a diffuse acoustic field. Equations (57) and (58) are, however, quite general and can be applied to structures with arbitrary geometry.

VI. CONCLUDING REMARKS

In this paper a diffuse field reciprocity relationship has been discussed which relates the direct field dynamic stiffness of a connection to a subsystem and the ensemble average reverberant loading on the connection degrees of freedom. This diffuse field reciprocity relation has been discussed previously by a number of authors for connections that can be described by a single generalized coordinate. The current analysis extends the reciprocity relation to connections described by an arbitrary number of generalized coordinates [as shown by Eq. (10)]. The constant of proportion-

ality in the diffuse field reciprocity relation is related to the expected amplitude of the reverberant field across the ensemble. Various approaches for defining this amplitude were discussed. Equation (44) defines the amplitude of the reverberant field in terms of the power incident upon the connection (expressed in terms of a modal energy density).

The main assumption on which the reciprocity relation is based is that there is a significant amount of uncertainty regarding the properties and location of the random boundary of a subsystem. Two- and three-dimensional subsystems (with a potentially infinite number of direct and reverberant field components) are likely to satisfy this criteria; one-dimensional subsystems with a limited number of reverberant field components may not satisfy this criteria. The work has three main applications: (i) the efficient calculation of the diffuse field response of a structural–acoustic system modeled using finite elements, boundary elements and/or infinite elements, (ii) the general calculation of the coupling loss factors used in Statistical Energy Analysis, and (iii) the derivation of an alternative method for describing the dynamic interactions of coupled subsystems with uncertain boundaries (a boundary approach to Statistical Energy Analysis). The latter application is discussed in a separate publication.[8]

ACKNOWLEDGMENT

The authors would like to thank Paul Bremner for making this work possible.

APPENDIX: LIST OF SYMBOLS

$\mathbf{a}$	*amplitude of direct field radiation components*
$\mathbf{b}$	*amplitude of reverberant field components*
B	*bending stiffness*
$\mathbf{C}$	*constraint matrix*
C	*constant related to incident power in diffuse field*
$\mathbf{c}$	*amplitudes of normalized direct field radiation components*
c	*fluid speed of sound*
c_g	*group velocity*
$\mathbf{D}$	*dynamic stiffness matrix*
$\mathbf{d}$	*amplitudes of normalized reverberant field radiation components*
e	*energy density*
E	*total time average energy*
$\mathbf{f}$	*generalized force*
$\mathbf{G}$, $\mathbf{H}$	*boundary element matrices*
k	*wave number*
l	*length*
m	*moment (45) and mass per unit length (47)*
n	*asymptotic modal density (s/rad)*
p	*pressure*
$\mathbf{P}$	*column matrix of eigenvectors*
P_{inc}	*incident power*
q	*generalized displacement*
$\mathbf{R}$	*transformation matrix*
R	*enclosing surface in farfield*
$\mathbf{S}_{ff}$	*cross-spectral force matrix*
t	*time*
$\mathbf{T}$	*scattering matrix*
$\mathbf{u}$	*displacement field*
V	*volume*
x	*location*
Γ	*boundary*
Ω	*domain*
δ	*Kronecker delta*
ϕ	*basis function*
η_{ij}	*coupling loss factor from subsystem i to subsystem j*
λ	*diagonal matrix of eigenvalues*
ρ	*fluid density*
ω	*radian frequency*
$\cdot_{r\cdot d}$	*random and deterministic portions of boundary*
$\cdot_{\text{dir}\cdot\text{rev}}$	*direct field, reverberant field*
$.^H$	*complex conjugate or Hermitian transpose*
$\langle\cdot\rangle$	*ensemble average*

[1] A. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (The Acoustical Society of America, Woodbury, New York, 1989).
[2] R. Cook, ''Lord Rayleigh and reciprocity in physics,'' J. Acoust. Soc. Am. **99**, 24–29 (1996).
[3] P. Smith, Jr., ''Response and radiation of structural modes excited by sound,'' J. Acoust. Soc. Am. **34**, 640–647 (1962).
[4] H. Diestel, ''Reciprocity calibration of microphones in a diffuse sound field,'' J. Acoust. Soc. Am. **33**, 514–518 (1961).
[5] E. Shaw, ''Diffuse field response, receiver impedance, and the acoustical reciprocity principle,'' J. Acoust. Soc. Am. **84**, 2284–2287 (1988).
[6] R. Lyon, *Statistical Energy Analysis of Dynamical Systems* (MIT Press, Cambridge, MA, 1975).
[7] P. Smith, Jr., ''Normal input admittance of a disk in a thin flat plate,'' J. Acoust. Soc. Am. **69**, 155–157 (1981).
[8] P. Shorter and R. Langley, ''Vibro-acoustic analysis of complex systems,'' accepted for publication in J. Sound Vib. (in press).
[9] C. Brebbia and J. Dominguez, *Boundary Elements: An Introductory Course* (Computational Mechanics Publications, McGraw-Hill, New York, 1992), p. 133 and p. 283.
[10] Similar expressions can be derived for mixed boundary conditions and general reactive impedance boundary conditions applied to the random boundary; Eqs. (6) and (7) are not used explicitly in subsequent derivation but are presented for illustration.
[11] Two quantities are said to be ''coherent'' if they maintain a fixed phase relationship at a given frequency of interest.
[12] The subsystem is initially assumed to be conservative. An approximate damping model can be introduced when describing the amplitude of the reverberant field.
[13] A. Sarkissian, ''Acoustic radiation from finite structures,'' J. Acoust. Soc. Am. **90**, 574–578 (1991).
[14] K. Cunefare, ''The minimum multimodal radiation efficiency of baffled finite beams,'' J. Acoust. Soc. Am. **90**, 2521–2529 (1991).
[15] It is noted in passing that alternative radiation components can be obtained by performing a simultaneous diagonalisation of the reactive and resistive components of the direct field dynamic stiffness [see for example, Chen and Ginsberg (Ref. 17)]. The use of such modes (and/or the simultaneous diagonalisation of the matrices H and G) may provide an interesting alternative to the current derivation. The term ''components'' is preferred to ''modes'' here to avoid confusion with the system modes of vibration.
[16] P. Chen and J. Ginsberg, ''Complex power, reciprocity, and radiation modes for submerged bodies,'' J. Acoust. Soc. Am. **98**, 3343–3351 (1995).
[17] P. Morse and J. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).
[18] In this context, the term ''component'' is used to refer to a basis function that is ''natural'' for describing a given response. The reverberant field components are very different from the subsystem modes shapes. For example, at a specified excitation frequency a rod has a single reverberant field component, consisting of the sum of a left and a right traveling wave. This contrasts with the modes of vibration of the rod, which are (theoretically) infinite in number and defined independently of the excitation frequency.
[19] C. Shannon and W. Weaver, *Mathematical Theory of Communication* (University of Illinois, 1963).

[20] R. Lyon, ''Needed: a new definition of diffusion,'' J. Acoust. Soc. Am. **56**, 1300–1302 (1974).
[21] The amount of ''uncertainty'' is related to the number of possible configurations that the random boundary can possess (or to the amount of ''information'' that is required to describe the random boundary). The random boundary of a waveguide with a single wave type can be described by a single degree of freedom and can therefore be viewed as possessing a relatively small amount of uncertainty.
[22] B. Mace, ''Wave coherence, coupling power and statistical energy analysis,'' J. Sound Vib. **199**, 369–380 (1997).
[23] T. Scharton, ''Frequency averaged power flow into a one-dimensional acoustic system,'' J. Acoust. Soc. Am. **50**, 373–381 (1971).
[24] F. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic, New York, 1985).
[25] R. Langley and P. Shorter, ''Diffuse wavefields in cylindrical coordinates,'' J. Acoust. Soc. Am. **112**, 1465–1470 (2002).
[26] R. Langley and P. Shorter, ''The wave transmission coefficients and coupling loss factors of point connected structures,'' J. Acoust. Soc. Am. **113**, 1947–1964 (2003).
[27] J. Manning, ''Formulation of SEA parameters using mobility functions,'' Philos. Trans. R. Soc. London, Ser. A **346**, 477–488 (1994).
[28] R. Cook, R. Waterhouse, R. Berendt, S. Edelman, and M. Thompson, ''Measurement of correlation coefficients in reverberant sound fields,'' J. Acoust. Soc. Am. **27**, 1072–1077 (1955).

Effect of resonator dimensions on nonlinear standing waves

C. Luo, X. Y. Huang, and N. T. Nguyen
School of Mechanical and Production Engineering, Nanyang Technological University, Singapore 639798

(Received 13 May 2004; revised 7 October 2004; accepted 9 October 2004)

An investigation of the effect of resonator dimensions on nonlinear standing waves in shaped resonators is conducted. Simple forms of the shear viscosity term in the momentum equations are developed for an axisymmetric (2D) resonator and a low aspect ratio rectangular (3D) resonator. The cross sections of the resonators are exponentially expanded and the one-dimensional wave equations are solved by using the Galerkin's method. The quality factors, pressure waveforms, compression ratios, and resonance frequencies are calculated for different dimensionless cross sections and lengths of the resonators. The results show that, apart from the resonator length, the ratio of the cross-section dimension to the length of the resonator is an important parameter. If the ratio is greater than 0.04, the characteristics of the shaped resonator are not affected significantly. However, when the ratio is less than 0.01, the resonance becomes weak, the compression ratio drops substantially, and the frequency response changes as well.
[DOI: 10.1121/1.1828611]

PACS numbers: 43.25.Gf, 43.20.Hq [MFH] Pages: 96–103

I. INTRODUCTION

The generation of high-amplitude pressure oscillations has recently been achieved, both in experiment[1] and in theoretical modelings,[2–5] in shaped resonators. In contrast to a normal cylindrical resonator, the higher-order modal frequencies in the shaped resonators are no longer multiples of the fundamental frequency, so that the energy transfer from the fundamental frequency to the higher harmonics is less efficient and the large-amplitude pressures are therefore generated. Lawrenson *et al.*[1] conducted experiments by shaking shaped resonators, and standing wave overpressures in excess of 340% ambient pressure were recorded. Ilinskii *et al.*[2] developed a one-dimensional model to analyze the nonlinear standing waves in shaped resonators. The large-amplitude pressures, waveform distortion, and resonance frequency shift were calculated and compared with the experimental results. The one-dimensional model was later improved by including the shear viscosity term in the momentum equation,[3] and the energy losses and quality factors were evaluated. An analytical study was conducted by Hamilton *et al.*[4] to predict and explain the dependence of the nonlinear frequency response on resonator shapes. Most recently, Erickson and Zinn[5] proposed a procedure to solve the one-dimensional wave equation for an exponentially shaped resonator by the Galerkin method. The compression ratio of the maximum to minimum pressures was calculated for different flare constants.

The large-amplitude pressure waves may have applications in miniature- or microscale fluid mechanics, known as microfluidics. This is currently an active research field[6,7] driven by fast development in biomedical engineering and life science. In the microfluidics, conventional fluid devices are shrunk in size to control and deliver fluid in miniature- or microscales. The applications of nonlinear acoustics in microfluidics have been seen, for example, in acoustic streaming[8] for liquid fluid manipulation in microchannels and in miniature synthetic jets[9] for aerodynamics control. The generation of high-pressure gas in micro- or miniature scales is probably more difficult than pumping liquids in microchannels, and it would be interesting to see if this can be achieved by the shaped acoustic resonators. The results may be used to develop miniature- or microcompressors, which are important in microscale cooling technologies and other bioengineering applications. This motivates the present study. The purpose of this study is to investigate how the nonlinear pressure waves change in a shaped acoustic resonator when the overall dimension of the resonator is shrunk, especially when the cross section is reduced. It is known that the resonance frequencies are determined primarily by lengths of the resonators, also by the shape variations in the axial direction. The cross sections of the resonators, on the other hand, affect the acoustical field through the shear viscosity. With a full recognition of the comprehensive work on energy losses conducted by Ilinskii *et al.*,[3] we develop in the present article a simple form for the shear viscosity term, so that the effect of the size reduction may be evaluated easily. The one-dimensional wave equation is then solved by the Galerkin method[5] for two kinds of shaped resonators. One is an exponentially expanded horn with axisymmetric cross sections, in which the acoustic field is two-dimensional (2D). The length of the resonator is l and radius at the small end is r_0. The other is an exponentially expanded horn with low aspect ratio rectangular cross sections, which would be easily fabricated with micromachine technologies. The acoustic field in this low aspect ratio rectangular resonator is essentially three-dimensional (3D). The height of the rectangular cross section is fixed at h and is much smaller than the width b for the most part, and the dynamics of the gas inside the horn can be considered as quasi-two-dimensional. The low aspect ratio rectangular resonator is referred as the 3D resonator in this paper to distinguish it from the axisymmetric resonator, which is referred as the 2D resonator. The quality factors, dynamic pressure waveforms, compression ratios, and resonance frequencies are calculated for different values

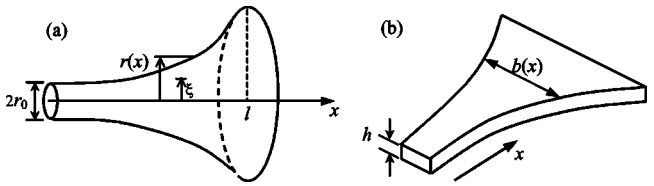


FIG. 1. Geometry and coordinates of the horn-shaped resonators. (a) 2D resonator, and (b) 3D resonator.

of r_0/l and h/l ranging from 0.01 to 0.1, in which the shear viscosity may play an important role in affecting the characteristics of shaped resonators.

II. SHEAR VISCOSITY TERMS AND WAVE EQUATIONS

A. Basic equations

We consider the one-dimensional acoustic wave field in a 2D resonator driven by an acceleration $a(t)$ along the axial direction, as shown in Fig. 1(a). The resonator is closed at both ends and the cross section is expanded in the x direction. The axial velocity along the x direction is $u=u(x,\xi,t)$, depending on x, t, and the radial location ξ. The velocity is not uniform over the cross section of the resonator, especially $u(x,\xi,t)=0$ on the wall of the resonator. The average velocity of u over the cross section can be calculated by

$$\overline{u}(x,t)=\frac{1}{\pi r^2}\int_0^r 2\pi u(x,\xi,t)\xi\, d\xi, \tag{1}$$

and the mass flux through the cross section is therefore

$$F=\int_0^r 2\pi\rho u(x,\xi,t)\xi\, d\xi=\rho\overline{u}\pi r^2. \tag{2}$$

It can be deduced, according to Eq. (2), that the velocity u in the continuity equation and momentum equation developed by Ilinskii *et al.*[2] was actually the average velocity $\overline{u}(x,t)$, and these equations can be rewritten as

$$\frac{\partial\rho}{\partial t}+\frac{1}{r^2}\frac{\partial}{\partial x}(r^2\rho\overline{u})=0, \tag{3}$$

$$\frac{\partial\overline{u}}{\partial t}+\underline{u}\frac{\partial\overline{u}}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}-a(t)+\frac{(\zeta+4\eta/3)}{\rho}\frac{\partial}{\partial x}\left(\frac{1}{r^2}\frac{\partial}{\partial x}(r^2\overline{u})\right), \tag{4}$$

where p is the pressure, $a(t)$ is the acceleration of the resonator, and ζ and η are coefficients of bulk and shear viscosities, respectively. We show in the following that the dissipative term in Eq. (4) can be derived from the viscous stress, together with an additional shear viscosity term. The general dissipative term used by Ilinskii *et al.*[2] can be expressed as[10]

$$\eta\frac{\partial^2 u_i}{\partial x_k\partial x_k}+\left(\zeta+\frac{1}{3}\eta\right)\frac{\partial^2 u_k}{\partial x_i\partial x_k}=\eta(\nabla^2\mathbf{u})_i+\left(\zeta+\frac{1}{3}\eta\right)[\nabla(\nabla\cdot\mathbf{u})]_i=\eta\{(\nabla^2\mathbf{u})_i-[\nabla(\nabla\cdot\mathbf{u})]_i\}+\left(\zeta+\frac{4}{3}\eta\right)[\nabla(\nabla\cdot\mathbf{u})]_i. \tag{5}$$

Let i in Eq. (5) be the x component of the velocity vector $\mathbf{u}$; the second term on the right-hand side becomes

$$\left(\zeta+\frac{4}{3}\eta\right)[\nabla(\nabla\cdot\mathbf{u})]_i=\left(\zeta+\frac{4}{3}\eta\right)\frac{\partial}{\partial x}(\nabla\cdot\mathbf{u}). \tag{6}$$

By taking average velocity over the cross section and using the continuity equation (3) for $\nabla\cdot\mathbf{u}$, one obtains the dissipative term in Eq. (4), which is the viscosity term associated with the volume expansion of the gas and was derived by Ilinskii *et al.*[2] The first term on the right-hand side of Eq. (5) is expressed in the cylindrical coordinates (x,ξ,θ) for the 2D resonator, and for the x component of the velocity it becomes

$$\eta\{(\nabla^2\mathbf{u})_i-[\nabla(\nabla\cdot\mathbf{u})]_i\}=\eta\frac{1}{\xi}\frac{\partial}{\partial\xi}\left(\xi\frac{\partial u}{\partial\xi}\right). \tag{7}$$

Equation (7) is a term induced by the shear motion of the velocity and is referred as the shear viscosity term. The evaluation of this term and its average depends on the velocity profile over the cross section, i.e., $u(x,\xi,t)$. By including the average of Eq. (7) in Eq. (4), one obtains

$$\frac{\partial\overline{u}}{\partial t}+\underline{u}\frac{\partial\overline{u}}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}-a(t)+\frac{(\zeta+4\eta/3)}{\rho}\frac{\partial}{\partial x}\times\left(\frac{1}{r^2}\frac{\partial}{\partial x}(r^2\overline{u})\right)+\frac{\eta}{\rho}\overline{\frac{1}{\xi}\frac{\partial}{\partial\xi}\left(\xi\frac{\partial u}{\partial\xi}\right)}. \tag{8}$$

A similar procedure has been applied to the 3D resonator, shown in Fig. 1(b), in a Cartesian coordinate system, and Eqs. (3) and (4) are modified as

$$\frac{\partial\rho}{\partial t}+\frac{1}{hb}\frac{\partial}{\partial x}(hb\rho\overline{u})=0, \tag{9}$$

$$\frac{\partial\overline{u}}{\partial t}+\underline{u}\frac{\partial\overline{u}}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}-a(t)+\frac{(\zeta+4\eta/3)}{\rho}\frac{\partial}{\partial x}\times\left(\frac{1}{hb}\frac{\partial}{\partial x}(hb\overline{u})\right)+\frac{\eta}{\rho}\overline{\frac{\partial^2 u}{\partial\xi^2}}. \tag{10}$$

In Eq. (10), $\eta\overline{\partial^2 u/\partial\xi^2}$ is the shear viscosity term, and the evaluation of this term depends also on the velocity profile over the duct cross section. The shear viscosity terms in both Eqs. (8) and (10) are derived in the next section.

B. Velocity profiles and shear viscosity terms

For the 2D resonator, we consider a tube with a circular cross section of radius r; the linear equation associated with the shear viscosity is

$$\rho\frac{\partial u}{\partial t}=-\frac{\partial p}{\partial x}+\eta\frac{1}{\xi}\frac{\partial}{\partial\xi}\left(\xi\frac{\partial u}{\partial\xi}\right). \tag{11}$$

By setting $u=v(x,\xi)e^{i\omega t}$, $\partial p/\partial x=p_x e^{i\omega t}$ (the real part will be taken from the results), and substituting them into Eq. (11), one obtains

$$\frac{\partial^2 v}{\partial \xi^2}+\frac{1}{\xi}\frac{\partial v}{\partial \xi}+\beta^2 v=\frac{p_x}{\eta}, \tag{12}$$

where $\beta^2=-(i\rho\omega/\eta)$, and $\beta=(1-i)\sqrt{\rho\omega/2\eta}$. The solution of Eq. (12), subject to boundary condition $v|_{\xi=r}=0$ and v being finite at $\xi=0$, is

$$v=\frac{p_x}{\eta\beta^2}\left[1-\frac{J_0(\beta\xi)}{J_0(\beta r)}\right].\ \text{Thus,}$$

$$u=ve^{i\omega t}=U(x,t)\left[1-\frac{J_0(\beta\xi)}{J_0(\beta r)}\right], \tag{13}$$

where $U(x,t)=p_x e^{i\omega t}/\eta\beta^2$. The average of u over the cross section of the tube is

$$\overline{u}=\frac{U(x,t)}{\pi r^2}\int_0^r 2\pi\xi\left[1-\frac{J_0(\beta\xi)}{J_0(\beta r)}\right]d\xi$$

$$=U(x,t)\left[1-\frac{2J_1(\beta r)}{\beta r J_0(\beta r)}\right]. \tag{14}$$

Using the velocity profile (13), the shear viscosity term in the momentum equation can be evaluated as

$$\eta\frac{1}{\xi}\frac{\partial}{\partial\xi}\left(\xi\frac{\partial u}{\partial\xi}\right)=U(x,t)\frac{\eta\beta^2}{J_0(\beta r)}J_0(\beta\xi). \tag{15}$$

The average of the shear viscosity term (15) over the tube cross section is

$$\overline{\eta\frac{1}{\xi}\frac{\partial}{\partial r}\left(\xi\frac{\partial u}{\partial\xi}\right)}=U(x,t)\frac{\eta\beta^2}{J_0(\beta r)}\frac{1}{\pi r^2}\int_0^r 2\pi\xi J_0(\beta\xi)d\xi$$

$$=U(x,t)\frac{2\eta\beta}{r}\frac{J_1(\beta r)}{J_0(\beta r)}$$

$$=\frac{2\eta\beta}{r}\frac{J_1(\beta r)}{J_0(\beta r)}\left[1-\frac{2J_1(\beta r)}{(\beta r)J_0(\beta r)}\right]^{-1}\overline{u}, \tag{16}$$

in which Eq. (14) has been used to eliminate $U(x,t)$. It can be shown[11] that for $|\beta r|\gg 1$, $J_1(\beta r)/J_0(\beta r)\to -i$; therefore

$$\overline{\eta\frac{1}{\xi}\frac{\partial}{\partial r}\left(\xi\frac{\partial u}{\partial\xi}\right)}\xrightarrow{|\beta r|\gg 1}\frac{2\eta\beta}{r}(-i)\left[1+\frac{2i}{(\beta r)}\right]^{-1}\overline{u}$$

$$\simeq(-1-i)\frac{\sqrt{2\eta\rho\omega}}{r}\overline{u}. \tag{17}$$

By taking the real part of Eq. (17), we have

$$\overline{\eta\frac{1}{\xi}\frac{\partial}{\partial r}\left(\xi\frac{\partial u}{\partial\xi}\right)}\xrightarrow{|\beta r|\gg 1}-\frac{\sqrt{2\eta\rho\omega}}{r}\overline{u}. \tag{18}$$

For the 3D resonator, we consider a two-dimensional duct; Eqs. (11) and (12) in this case become

$$\rho\frac{\partial u}{\partial t}=-\frac{\partial p}{\partial x}+\eta\frac{\partial^2 u}{\partial\xi^2}, \tag{19}$$

$$\frac{\partial^2 v}{\partial\xi^2}-\beta^2 v=\frac{p_x}{\eta}, \tag{20}$$

where $\beta^2=i\rho\omega/\eta$, and $\beta=(1+i)\sqrt{\rho\omega/2\eta}$. The velocity u obtained from the solution of Eq. (20), subject to boundary conditions $v|_{\xi=h}=v|_{\xi=0}=0$, is

$$u=ve^{i\omega t}=U(x,t)\left\{\cosh(\beta y)-1-\frac{\cosh(\beta h)-1}{\sinh(\beta h)}\sinh(\beta y)\right\}, \tag{21}$$

where $U(x,t)=p_x e^{i\omega t}/\eta\beta^2$. The average of u over the duct height is

$$\overline{u}=\frac{U(x,t)}{\beta h}\left\{\sinh(\beta h)-\beta h-\frac{[\cosh(\beta h)-1]^2}{\sinh(\beta h)}\right\}. \tag{22}$$

The average of shear stress $\eta\overline{\partial^2 u/\partial\xi^2}$ can be worked out based on Eq. (21). By following the same procedure for the tube, we have

$$\eta\overline{\frac{\partial^2 u}{\partial\xi^2}}\xrightarrow{|\beta h|\gg 1}-\frac{\sqrt{2\eta\rho\omega}}{h}\overline{u}. \tag{23}$$

The condition $|\beta r|\gg 1$ used in Eqs. (18) and (23) can be justified by taking η in order of 10^{-5}, ω in order of 10^3 (assuming the oscillation frequency to be 500 Hz), so that $|\beta r|\sim r\sqrt{\rho\omega/\eta}\sim r\times 10^4$. The condition $|\beta r|\gg 1$ is therefore equivalent to $|\beta r|\sim r\times 10^4\gg 1$, or, $r\gg 10^{-4}$ m$=0.1$ mm, which can always be met in the present study. Equations (18) and (23) are the shear viscosity terms to be included in the momentum equation for $\overline{u}$. They are equivalent to the term introduced by Ilinskii *et al.*[3] but in simpler forms. It should be pointed out, however, that the dissipations in the work by Ilinskii *et al.*[3] include the turbulence induced energy loss and therefore are more comprehensive when the pressure amplitudes are large.

C. Dimensionless wave equations

With the shear viscosity term expressed in Eq. (18), the momentum equation (8) for the 2D resonator can now be written as

$$\frac{\partial\overline{u}}{\partial t}+\overline{u}\frac{\partial\overline{u}}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}-a(t)$$

$$+\frac{(\zeta+4\eta/3)}{\rho}\frac{\partial}{\partial x}\left(\frac{1}{r^2}\frac{\partial}{\partial x}(r^2\overline{u})\right)$$

$$-\frac{\sqrt{2\rho_0\eta\omega}}{r_\rho}\overline{u}. \tag{24}$$

It can be seen from Eq. (24) that the shear viscosity term takes into account the effect of frequency ω and dimension r. The dissipation increases as ω increases and r decreases. By introducing the velocity potential

$$\overline{u}=\frac{\partial\varphi}{\partial x}, \tag{25}$$

and following the same procedure developed by Ilinskii *et al.*[2] and Erickson and Zinn,[4] the dimensionless wave equation becomes

$$\frac{1}{\pi^2 S}\frac{\partial}{\partial X}\left(S\frac{\partial \Phi}{\partial X}\right)-\Omega^2\frac{\partial^2\Phi}{\partial T^2}+\frac{G_B\Omega}{\pi^3 S}\frac{\partial^2}{\partial T\partial X}\times\left(S\frac{\partial\Phi}{\partial X}\right)-\frac{G_S\Omega^{3/2}}{R}\frac{\partial\Phi}{\partial T}=\Omega\frac{\partial A}{\partial T}X+A(T)\frac{\partial\Phi}{\partial X}+\frac{\gamma-1}{S}A(T)X\frac{\partial}{\partial X}\left(S\frac{\partial\Phi}{\partial X}\right)+2\Omega\frac{\partial^2\Phi}{\partial X\partial T}\frac{\partial\Phi}{\partial X}+\frac{(\gamma-1)\Omega}{S}\frac{\partial\Phi}{\partial T}\frac{\partial}{\partial X}\left(S\frac{\partial\Phi}{\partial X}\right)+\frac{\gamma-1}{2S}\left(\frac{\partial\Phi}{\partial X}\right)^2\frac{\partial}{\partial X}\left(S\frac{\partial\Phi}{\partial X}\right). \tag{26}$$

The dimensionless variables are

$$X=\frac{x}{l},\quad T=\omega t,\quad S=\frac{\pi r^2}{l^2},\quad R=\frac{r}{l},\quad A=\frac{a}{l\omega_0^2},$$

$$\Phi=\frac{\varphi}{l^2\omega_0},\quad \Omega=\frac{\omega}{\omega_0}. \tag{27}$$

$$G_B=\frac{\pi\delta\omega_0}{c_0^2},\quad \delta=\frac{\zeta+4\eta/3}{\rho_0},\quad G_S=\sqrt{\frac{2\eta}{\pi\rho_0 c_0 l}},$$

$$\omega_0=\frac{\pi c_0}{l}. \tag{28}$$

Here, G_B is the same as G and D introduced by Ilinskii *et al.*[2] and Erickson and Zinn,[4] respectively, G_S is a parameter associated with the shear viscosity, l is length of the resonator, and c_0 is the speed of sound. Equation (26) shows that the coefficient of the shear viscosity term is inversely proportional to the dimensionless cross-section size, R, and the square root of the resonator length, $\sqrt{l}$.

Similarly, by substituting Eq. (23) into Eq. (10), we obtain the momentum equation for the 3D resonator

$$\frac{\partial\overline{u}}{\partial t}+\overline{u}\frac{\partial\overline{u}}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}-a(t)+\frac{(\zeta+4\eta/3)}{\rho}\frac{\partial}{\partial x}\left(\frac{1}{hb}\frac{\partial}{\partial x}(hb\overline{u})\right)-\frac{\sqrt{2\rho_0\eta\omega}}{h}\overline{u}, \tag{29}$$

and the dimensionless wave equation

$$\frac{1}{\pi^2 B}\frac{\partial}{\partial X}\left(B\frac{\partial \Phi}{\partial X}\right)-\Omega^2\frac{\partial^2\Phi}{\partial T^2}+\frac{G_B\Omega}{\pi^3 B}\frac{\partial^2}{\partial T\partial X}\times\left(B\frac{\partial\Phi}{\partial X}\right)-\frac{G_S\Omega^{3/2}}{H}\frac{\partial\Phi}{\partial T}=\Omega\frac{\partial A}{\partial T}X+A(T)\frac{\partial\Phi}{\partial X}+\frac{\gamma-1}{B}A(T)X\frac{\partial}{\partial X}\left(B\frac{\partial\Phi}{\partial X}\right)+2\Omega\frac{\partial^2\Phi}{\partial X\partial T}\frac{\partial\Phi}{\partial X}+\frac{(\gamma-1)\Omega}{B}\frac{\partial\Phi}{\partial T}\frac{\partial}{\partial X}\left(B\frac{\partial\Phi}{\partial X}\right)+\frac{\gamma-1}{2B}\left(\frac{\partial\Phi}{\partial X}\right)^2\frac{\partial}{\partial X}\left(B\frac{\partial\Phi}{\partial X}\right), \tag{30}$$

where $H=h/l$ and $B=[b(x)]/l$, and other parameters are the same as those in Eqs. (27) and (28). It has been assumed that the shear viscosity is important only on upper and lower walls in the 3D resonator.

III. EFFECT OF RESONATOR DIMENSIONS

A. Quality factors

The energy loss in a dynamic system may be measured by quality factors, which is the ratio of total energy stored in the system to the energy dissipated in one cycle. There are many ways to calculate the quality factor and one of them is to use the decay factor, κ, from the time factor $e^{-\kappa t}$, and the quality factor is obtained by[12]

$$Q=\frac{\pi f_0}{\kappa}, \tag{31}$$

where f_0 is the fundamental frequency of the system. Equation (26) can be reduced to a linear equation by dropping all quadratic terms and the driving force A

$$\frac{\partial^2\Phi}{\partial T^2}-\frac{1}{\Omega^2\pi^2}\frac{\partial^2\Phi}{\partial X^2}=\frac{G_B}{\Omega\pi^3}\frac{\partial^3\Phi}{\partial T\partial X^2}-\frac{G_S}{\Omega^{1/2}R}\frac{\partial\Phi}{\partial T}. \tag{32}$$

By substituting the fundamental mode $\Phi=A\cos(\pi X)e^{(i-\sigma)T}=A\cos(\pi X)e^{(i-\sigma)\omega t}$ into Eq. (32), where $\sigma=\kappa/\omega_0$ is the dimensionless decay factor, taking $\Omega=1$ ($\omega=\omega_0$), at which the energy losses are the maximum,[3] and keeping terms of the first order of σ, one obtains

$$\sigma=\frac{1}{2}\left(\frac{G_B}{\pi}+\frac{G_S}{R}\right). \tag{33}$$

In Eq. (33), R follows the resonator expansion profile $R(X)=(r_0/l)f(X)$. By denoting

$$K=\int_0^1\frac{dX}{f(X)}, \tag{34}$$

the average of σ over the resonator length can be calculated as

$$\overline{\sigma}=\frac{1}{2}\left(\frac{G_B}{\pi}+\frac{G_S}{r_0/l}\int_0^1\frac{dX}{f(X)}\right)=\frac{1}{2}\left(\frac{G_B}{\pi}+\frac{KG_S}{r_0/l}\right). \tag{35}$$

The quality factor is calculated by

$$Q=\frac{\pi f_0}{k}=\frac{\pi f_0}{\overline{\sigma}\omega_0}=\left(\frac{G_B}{\pi}+\frac{KG_S}{r_0/l}\right)^{-1}=\left(\frac{1}{Q_B}+\frac{1}{Q_S}\right)^{-1}, \tag{36}$$

where

$$Q_B=\frac{\pi}{G_B}=\frac{c_0\rho_0 l}{\pi(\zeta+4\eta/3)}, \tag{37}$$

being the quality factor due to the dissipation associated with the volume change[2]

$$Q_S=\frac{r_0/l}{KG_S}=\frac{r_0}{l}\frac{1}{K}\sqrt{\frac{\pi l c_0\rho_0}{2\eta}}, \tag{38}$$

which is the quality factor due to the dissipation associated with the shear motion in tubes.[13] Since the viscosity coeffi-

cients ζ and η are of the order of 10^{-5} and Q_S/Q_B is of the order of $\sqrt{\eta}<10^{-2}$, the overall quality factor can be approximated by Q_S, i.e.,

$$Q \simeq Q_S = \frac{r_0}{l}\frac{1}{K}\sqrt{\frac{\pi l c_0 \rho_0}{2\eta}}. \tag{39}$$

Equation (39) shows that the quality factor is proportional to ratio r_0/l and $\sqrt{l}$. The quality factors are calculated, based on Eq. (39), for four resonators used in the experiment by Lawrenson *et al.*[1] The parameters of these resonators are listed below.

Cylinder ($l=0.10$ m), $r(x)=0.0222$ m.
Cone ($l=0.17$ m), $r(x)=0.0056+0.2680x$ m.
Horn-cone ($l=0.24$ m),

$$r(x)=\begin{cases} 0.0068\cosh(23.86x), & 0\leqslant x\leqslant 0.06 \text{ m}; \\ 0.015+0.134x, & 0.06\leqslant x\leqslant 0.24 \text{ m}. \end{cases}$$

Bulb ($l=0.28$ m),

$$r(x)=\begin{cases} 0.007-0.15x+4.1x^2-9.9\times10^4x^3-9\times10^6x^4, & -0.005\,08 \text{ m}\leqslant x\leqslant 0; \\ l[0.025-0.15(x/l)+1.15(x/l)^2-0.9(x/l)^4], & 0\leqslant x\leqslant 0.28 \text{ m}. \end{cases}$$

The results are given in Table I for the resonators filled with refrigerant (R-134a).[14] The results obtained by Ilinskii *et al.*[3] are also listed in Table I, which are in a range for each resonator depending on the excitation level. It can be seen that most of quality factors are of order of 500, which was the value measured in the experiment.[1] The quality factors predicted by the present study are generally higher than the values obtained by Ilinskii *et al.*,[3] especially for the straight cylinder. This is probably due to the additional energy loss associated with turbulence generated by the large-amplitude pressures in the shaped resonators or the shock waves in the straight cylinder.

For the 2D resonator shown in Fig. 1(a), the radius expansion function is

$$f(X)=e^{\alpha X/2}, \tag{40}$$

where α is the flare constant indicating the expansion rate. The quality factors are computed for different α and the results are presented in Fig. 2. The quality factor for the 3D resonator, shown in Fig. 1(b), can be obtained by simply replacing r_0/l with h/l and setting $K=1$ in Eq. (39), which is the same for a straight cylinder. The quality factors for the 3D resonator and cylinder are also plotted in Fig. 2. It is seen, from Fig. 2, that the 2D resonators have higher quality factors than that of the 3D resonator and cylinder. This is because the energy dissipation due to the shear viscosity is reduced along the 2D resonators as the cross sections are expanded, whereas the energy dissipation in the 3D resonator is increased from the small end to the big end, as the height h between the upper/lower walls is fixed and wall areas are increased. Figure 2 also shows that the 2D resonators with larger flare numbers have higher quality factors. The difference of the quality factors between the 2D and 3D resonators affects the pressure waves in the resonators.

TABLE I. Quality factors of resonators filled with R-134a.

	Results of Ilinskii *et al.* (Ref. 3)	Results of present study	Measured value (Ref. 1)
Cylinder	350–450	1047	~500
Horn-	200–600	740	
Bulb	400–900	574	
Cone	···	745	

B. Waveforms and compression ratios

The one-dimensional wave equations (26) and (30) are solved by the Galerkin's method developed by Erickson and Zinn.[5] For comparison purposes, all parameters of the gas inside the resonators are assumed to have the same values as those used by Erickson and Zinn.[5] The coefficients of shear viscosity and bulk viscosity are of the same order, and for simplicity, we set them to be equal to $1.7E-005$ Pa·s. The resonators are oscillated at

$$A=A_0\cos(T), \tag{41}$$

and the amplitude is fixed at $A_0=5\times10^{-4}$ throughout the calculations. The cross sections of both resonators are expanded exponentially as

$$S=S_0e^{\alpha X}, \tag{42}$$

where $S_0=\pi r_0^2/l^2$ for the 2D resonator and $S_0=b_0h/l^2$ for the 3D resonator. The flare constant, α, is fixed at $\alpha=5.75$ in the following calculations because the compression ratio is the maximum at this value.[5] The dimensionless velocity potential, Φ, is expressed as

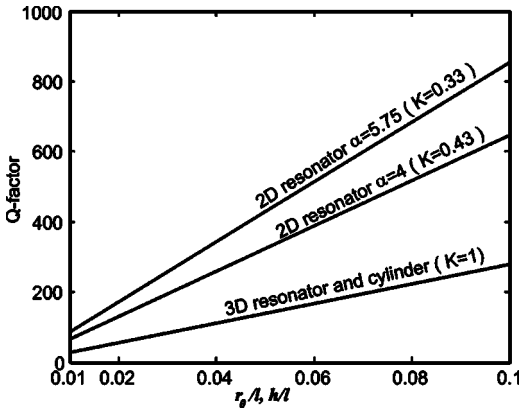


FIG. 2. The quality factor versus r_0/l and h/l for the 2D-resonator, 3D resonator, and a straight cylinder.

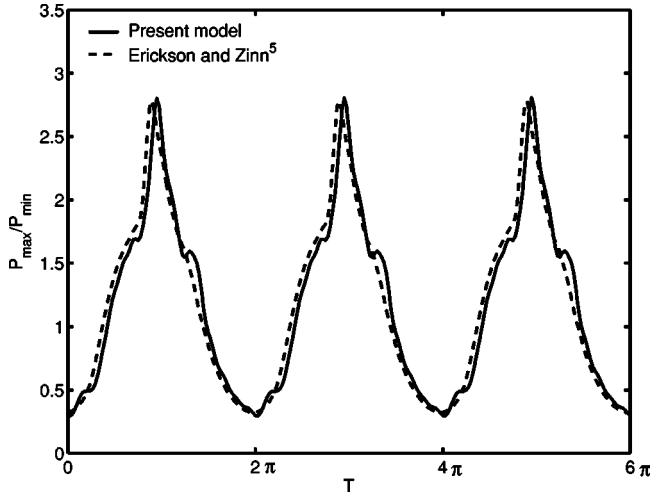


FIG. 3. Pressure waveforms at the small end of the 2D resonator. $S=\pi(r_0/l)^2e^{\alpha X}$, α=5.75. Dashed line: reproduced for the results by Erickson and Zinn (Ref. 5). Solid line: calculated from Eqs. (26) and (44), $l=0.2$ m, $r_0/l=0.023$.

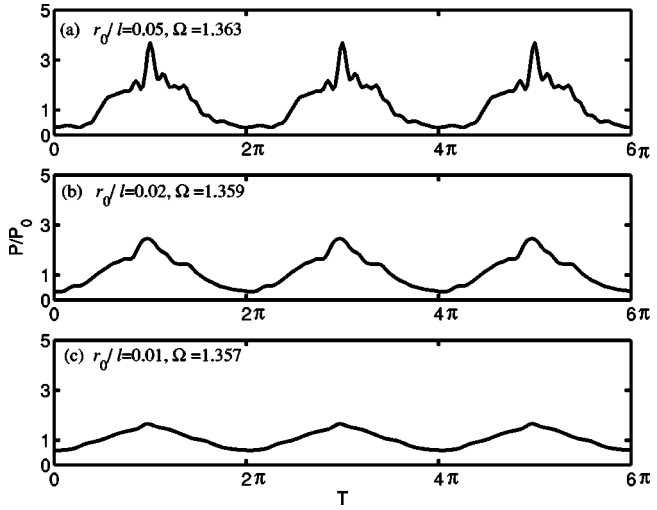


FIG. 4. Pressure waveforms at the small end of the 2D resonator. $S=\pi(r_0/l)^2e^{\alpha X}$, α=5.75, $l=0.2$ m. (a) $r_0/l=0.05$; (b) $r_0/l=0.02$; and (c) $r_0/l=0.01$.

$$\Phi(X,T)=\sum_{n=1}^{\infty}\eta_n(T)\Psi_n(X). \tag{43}$$

The trial functions, $\Psi_n(X)$, have the same forms as those used by Erickson and Zinn[5] for the resonators sealed at both ends. The time-dependent amplitudes, $\eta_n(T)$, are solved by the techniques suggested by Erickson and Zinn,[5] in which the series of Eq. (43) is truncated at $n=20$. Finally, the pressures are calculated for the 2D resonator by

$$\frac{P}{P_0}=\left[1-(\gamma-1)\pi^2\left(\Omega\frac{\partial\Phi}{\partial T}+\frac{1}{2}\left(\frac{\partial\Phi}{\partial X}\right)^2+AX-\frac{G_B}{\pi^3}\frac{1}{S}\frac{\partial}{\partial X}\left(S\frac{\partial\Phi}{\partial X}\right)+\frac{G_S\Omega^{1/2}\Phi}{R}\right)\right]^{\gamma/(\gamma-1)}, \tag{44}$$

and for the 3D resonator by

$$\frac{P}{P_0}=\left[1-(\gamma-1)\pi^2\left(\Omega\frac{\partial\Phi}{\partial T}+\frac{1}{2}\left(\frac{\partial\Phi}{\partial X}\right)^2+AX-\frac{G_B}{\pi^3}\frac{1}{S}\frac{\partial}{\partial X}\left(S\frac{\partial\Phi}{\partial X}\right)+\frac{G_S\Omega^{1/2}\Phi}{H}\right)\right]^{\gamma/(\gamma-1)}. \tag{45}$$

In Eqs. (44) and (45), P_0 is the static pressure in the resonators. As a reference, the pressure waves obtained by Erickson and Zinn[5] at the small end in the 2D resonator are reproduced here with the same condition by setting $G_S=0$ and assigning $G_B=0.01$ in Eq. (26), shown in Fig. 3 by the dashed lines. In their results (the dashed line), the pressure waveforms were not dependent on the resonator dimensions, such as the resonator length l and ratio r_0/l, because there was no shear viscosity term in the wave equation. The solid lines in Fig. 3 are the pressure waves calculated in the present study using Eqs. (26) and (44), in which G_B and G_S are evaluated according to their definitions and gas parameters, the resonator length is $l=0.2$ m, $R=r_0/l=0.023$. It is seen that two waveforms are very similar. In fact, the pressure waveforms in the resonator are dependent on the cross-section dimensions, r_0/l, due to the shear viscosity term. The resonance pressure waveforms at the small end of the 2D resonator are plotted for different values of r_0/l in Fig. 4. When r_0/l is 0.05, the pressure amplitudes are large with sharp peaks [Fig. 4(a)], indicating a strong resonance. When r_0/l is reduced to 0.02, the pressure amplitudes are reduced, as well as the sharp peaks [Fig. 4(b)]. When r_0/l is further reduced to 0.01 [Fig. 4(c)], the pressure amplitudes are reduced further and there are no sharp peaks, indicating that the resonance has been weakened due to the energy dissipation. The resonance frequency changes slightly when the ratio r_0/l varies, as denoted in the figure for each case. The similar situation is also observed for the pressure waves in the 3D resonator, which are plotted in Fig. 5. In this case, the waveforms have been smoothed down at $r_0/l=0.02$, because the 3D resonator is more dissipative than the 2D resonator, as shown by the quality factors in Fig. 3.

The compression ratios are calculated based on the

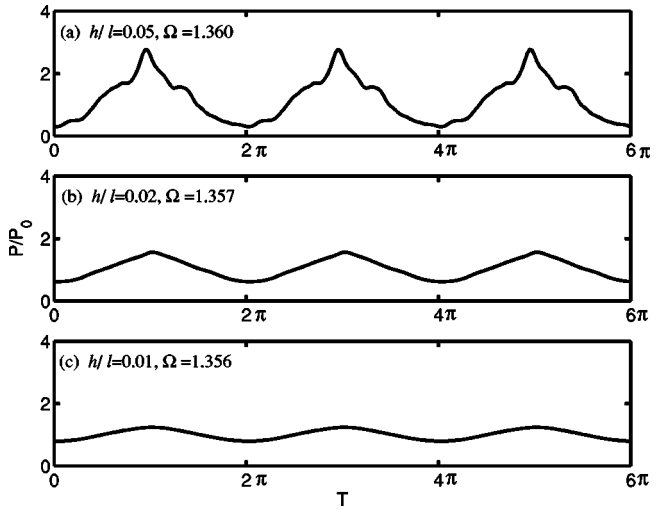


FIG. 5. Pressure waveforms at the small end of the 3D resonator. $S=b_0h/l^2e^{\alpha X}$, α=5.75, $l=0.2$ m. (a) $h/l=0.05$; (b) $h/l=0.02$; and (c) $h/l=0.01$.

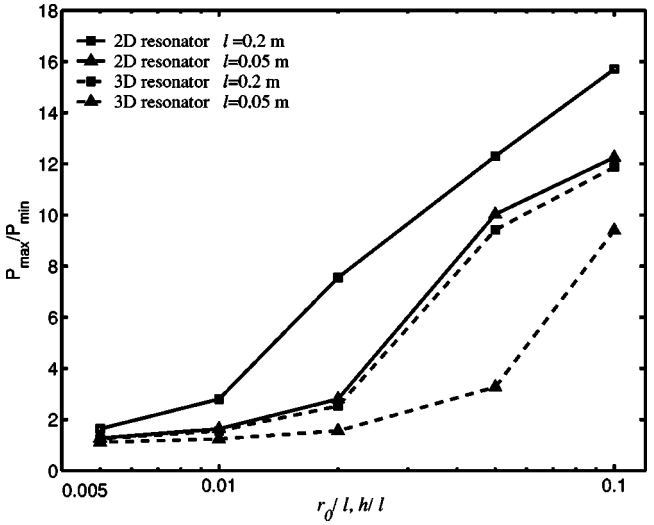


FIG. 6. Compression ratios versus r_0/l and h/l for the 2D and 3D resonators. The cross-section area is $S=\pi(r_0/l)^2e^{\alpha X}$ for the 2D resonator, and $S=hb_0/l^2e^{\alpha X}$ for the 3D resonator, $\alpha=5.75$.

maximum and minimum pressures from the waveforms, and the results for both 2D resonator and 3D resonator are presented in Fig. 6, versus r_0/l for $l=0.2$ and 0.05 m, respectively. It is seen that the compression ratios, for $l=0.2$ m, can be greater than 10 when r_0/l is greater than 0.04 for the 2D resonator and h/l is greater than 0.05 for the 3D resonator, but drop to 2 or less when r_0/l and h/l are less than 0.01, showing that the cross-section dimension of the resonator has tremendous effect on the compression ratio. On the other hand, when the overall length of the resonator is reduced, the compression ratio is also reduced, as shown in Fig. 6 by the results of $l=0.05$ m for both resonators. This is because the shorter length will lead to a higher resonance frequency, and therefore higher dissipation by the shear viscosity. It is seen from Fig. 6 that the compression ratio drops substantially at $r_0/l=0.02$ for the 2D resonator when l is reduced from 0.2 to 0.05 m, and increases quickly afterward. A similar situation is also seen for the 3D resonator but at $h/l=0.05$. The mechanism associated with these inconsistencies is not clear from the present modeling, and further study is needed. The results presented in Fig. 6 suggest that r_0/l, h/l, and l are probably the control parameters to the resonator dimensions for the given driving strength and required compression ratio. The comparison between the 2D resonator and 3D resonator in Fig. 6 shows that the compression ratios of the 3D resonator are always lower than that of 2D resonator. If one takes $l=0.05$ m and $h/l=0.05$, i.e., $l=50$ mm and $h=2.5$ mm, which is a small resonator, one may expect to have pressure waves with a compression ratio of 3, according to the results in Fig. 6. On the other hand, the resonance may be easier to be excited in a smaller resonator than in a bigger one for the same driving power. When the excitation is doubled by setting $A_0=1\times10^{-3}$ in the above-mentioned small resonator, the calculation results show that the compression ratio is increased from 3 to 10.

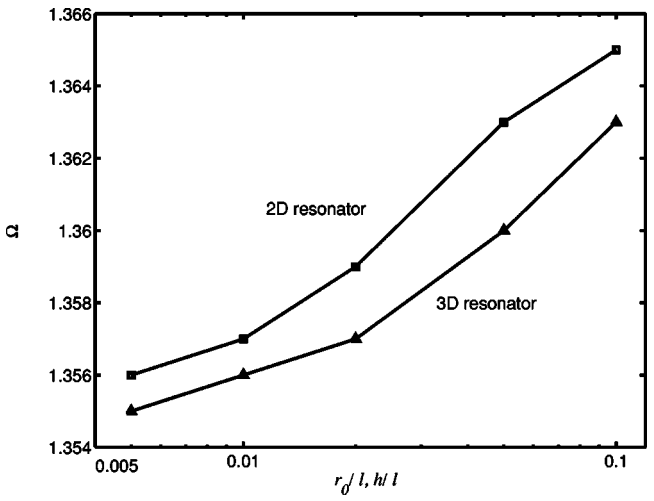


FIG. 7. Dimensionless resonance frequency versus cross-section dimensions for both 2D and 3D resonators with length $l=0.2$ m.

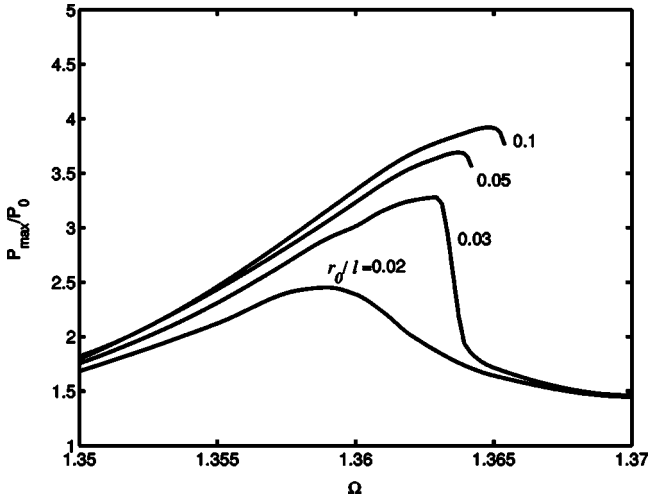


FIG. 8. Frequency response curves of the 2D resonator with different dimension ratios r_0/l. $S=\pi(r_0/l)^2e^{\alpha X}$, $\alpha=5.75$, $l=0.2$ m.

The cross-section dimensions of the resonators are found to have minor effect on the resonance frequencies, which are primarily determined by the resonator lengths and the axial-expansion shape. The calculated results for both 2D and 3D resonators, plotted out in Fig. 7, show that the variation of the dimensionless frequency (Ω) is less than 1% when the cross-section dimensions (r_0/l and h/l) change from 0.1 to 0.005. However, the changes in the resonance frequency and dimension parameters produce tremendous difference in the frequency response curves, which are illustrated in Fig. 8 by plotting the maximum pressure versus the frequency Ω for the 2D resonator. The frequency response curves at different values of parameter r_0/l in this case are very similar to the hardening behaviors of a conical resonator under different excitation levels. This resemblance indicates that the resonator dimensions affect the dynamic energy of the system through the shear viscosities.

IV. CONCLUSIONS

The effect of resonator dimensions on the nonlinear standing waves inside the shaped resonators has been studied. A shear viscosity term has been added to the one-dimensional momentum equation for the average axial veloc-

ity, and the wave equation has been solved by the Galerkin method for a 2D (axisymmetric) resonator and a 3D (low aspect ratio rectangular) resonator, whose cross sections are exponentially expanded. By calculations of quality factors, the pressure waveforms, and compression ratios, it is found that the shear viscosity dissipation plays a crucial role when the resonator sizes are reduced. The resonator length (l) and the ratio of the cross-section dimension to the length (r_0/l and h/l in the present study) are two controlling parameters. If r_0/l and h/l are greater than 0.04 for the present resonators, the typical characteristics of shaped resonators, such as high-amplitude pressures and hardening behaviors, are observed. When r_0/l and h/l are less than 0.01, the resonance becomes weak and the compression ratios drop to 2 and below, indicating that the resonators are not functioning in these cases to generate large-amplitude pressures, unless the excitation levels are increased. The results also show that the dimension ratios control the frequency response curves in the same way as the excitation level does, following the hardening behavior of the shaped resonators. Although the study is conducted based on only two specially shaped resonators, the results may be similar to other shaped resonators.

ACKNOWLEDGMENT

Luo Cheng wishes to gratefully acknowledge the scholarship from Nanyang Technological University, Singapore.

[1] C. C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, ''Measurement of macrosonic standing wave in oscillating closed cavities,'' J. Acoust. Soc. Am. **104**, 623–636 (1998).

[2] Y. A. Ilinskii, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, ''Nonlinear standing waves in an acoustical resonator,'' J. Acoust. Soc. Am. **104**, 2664–2674 (1998).

[3] Y. A. Ilinskii, B. Lipkens, and E. A. Zabolotskaya, ''Energy losses in an acoustical resonator,'' J. Acoust. Soc. Am. **109**, 1859–1870 (2001).

[4] M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, ''Linear and nonlinear frequency shifts in acoustical resonator with varying cross section,'' J. Acoust. Soc. Am. **110**, 109–119 (2001).

[5] R. R. Erickson and B. T. Zinn, ''Modeling of finite amplitude acoustic wave in closed cavities using the Galerkin method,'' J. Acoust. Soc. Am. **103**, 1863–1870 (2003).

[6] N. T. Nguyen and S. T. Wereley, *Fundamentals and Applications of Microfluidics* (Artech House, 2002).

[7] M. Gad-et-Hak, ''The fluid mechanics of microdevices—The Freeman Scholar Lecture,'' J. Fluids Eng. **121**, 5–32 (1999).

[8] N. T. Nguyen and R. M. White, ''Acoustic streaming in micromachined flexural plate wave devices: numerical simulation and experimental verification,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**(6), 1463–1471 (2000).

[9] D. J. Coe, M. G. Allen, M. A. Trautman, and A. Glezer, ''Micromachined Jets for Manipulation of Macroflows,'' Solid-State Sensor and Actuator Workshop, 13–16 June, Hilton Head, SC (1994).

[10] L. D. Landau and E. M. Lishitz, *Fluid Mechanics* (Pergamon, Oxford, 1987).

[11] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (U.S. Department of Commerce, 1972).

[12] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[13] M. J. Moloney and D. L. Hatten, ''Acoustic quality factor and energy losses in cylindrical pipes,'' Am. J. Phys. **69**, 311–314 (2001).

[14] Japanese Association of Refrigeration, ''Thermophysical Properties of Environmentally Acceptable Fluorocarbons, HFC-134a and HCFC-123,'' (1991).

Cavitation bubble-driven cell and particle behavior in an ultrasound standing wave

Larisa A. Kuznetsova[a)]
School of Biosciences, Cardiff University, Park Place, Cardiff CF10 3TL, United Kingdom

Sanjay Khanna and Nazar N. Amso
Department of Obstetrics and Gynaecology, University of Wales College of Medicine, Heath Park, Cardiff CF14 4XN, United Kingdom

W. Terence Coakley
School of Biosciences, Cardiff University, Park Place, Cardiff CF10 3TL, United Kingdom

Alexander A. Doinikov
Institute of Nuclear Problems, Belarus State University, 11 Bobruiskaya Street, Minsk 220050, Belarus

(Received 25 May 2004; revised 15 October 2004; accepted 31 October 2004)

The behavior of human erythrocytes and 1-μm-diameter fluorescent latex beads in the presence of Optison® contrast agent in a single half-wavelength ($\lambda/2$) ultrasound standing wave (USSW) resonator has been studied. The particle movements were observed with an epi-fluorescent microscope and the velocity of the particles and cells was measured by particle image velocimetry (PIV). Acoustic emissions were monitored with a microphone and a spectrum analyzer. Optison® contrast agent disintegrated immediately on exposure to ultrasound of 0.98-MPa acoustic pressure amplitude or higher in a chamber driven at its resonance frequency of 1.56 MHz. A discrete cloud of active microbubbles, detected at the pressure node plane, disappeared gradually and was completely lost within 15 s. The microscopy showed three-dimensional regions of circulation of both 1-μm tracer particles and erythrocytes in planes perpendicular to the pressure node plane. A numerical simulation showed that, for parameters that conform to the experimental conditions, a bubble of a subresonance size moves towards and translates about a pressure node plane. This result is in agreement with the experimental observation that the particle and cell circulation is induced by the presence and/or translational motion of microbubbles at the pressure node plane. [DOI: 10.1121/1.1835503]

PACS numbers: 43.25.Yw, 43.25.Gf, 43.25.Nm, 43.35.Wa [AJS] Pages: 104–112

I. INTRODUCTION

It has long been appreciated that, in low-pressure amplitude acoustic fields, bubbles larger than resonant size accumulate in the node plane and smaller, active, bubbles are in the pressure antinode plane (Eller, 1968). It has been pointed out, however, that at high peak pressure amplitudes active subresonant-size bubbles may be trapped at and reciprocate about the node plane (Watanabe and Kukita, 1993; Doinikov, 2002). Miller (1977) has reported that bubbles of resonance size in 1-MHz standing wave field could form stable arrays and orbit about a point in a pressure-null surface. He hypothesized that the interplay of size change due to diffusion, and of change in the radiation forces on the bubbles due to the size change, could explain the orbit phenomenon. Khanna *et al.* (2003) provided experimental support for the presence of subresonance-size bubbles in a pressure node plane. They studied the behavior of human red blood cells in the presence of acoustic contrast agent in a $\lambda/2$-ultrasound standing wave resonator for the range of acoustic pressure amplitudes of 0.98–1.96 MPa. The cells circulated about the pressure node plane and experienced some apparent structural damage while subharmonic, ultraharmonic, and harmonic emissions were detected from the insonated volume. The cell damage was taken as an indicator of the presence of mechanically active subresonant-size bubbles in the region of the pressure node, while the cell circulation was attributed to translation of the subresonant bubbles about the node plane.

Cavitation is, as generally believed, the cause of bioeffects of ultrasound in the absence of ultrasonic heating (Carstensen *et al.*, 1993). Acoustic cavitation is the interaction of a sound field in a fluid medium with gaseous inclusion in that medium. The subsequent behavior of these cavitation bubbles in a sound field is complex and can include stable oscillation, growth and collapse of bubbles within a few cycles (Apfel, 1981), and microbubble emission from large bubbles excited in a surface wave mode (Neppiras and Coakley, 1976). Miller *et al.* (1996) reviewed the biological effects of ultrasound with the emphasis on the physical and chemical mechanisms of action of inertial cavitation.

Much recent work has concentrated on efficacy of encapsulated contrast microbubbles in causing membrane permeabilization through cavitation mechanisms both *in vitro* and *in vivo*, whereby the microbubbles themselves may be driven as stable cavities or whereby their destruction instantly provides cavitation nuclei whose subsequent behavior will stress cells (Miller *et al.*, 2001; Frenkel *et al.*, 2002; May and Allen, 2002).

a)Electronic mail: kuznetsoval@cardiff.ac.uk

Optison® employed in the current work is a second generation contrast agent consisting of hollow albumin microspheres of 2–4 μm in diameter filled with octafluoropropane (OFP). It possesses longer persistence in an ultrasound field compared with the air-filled agents like Albunex owing to the lower solubility of OFP in aqueous solutions and hence, prolonged gas retention in the microsphere core (Podell *et al.*, 1999). Recent research showed that on ultrasound application the contrast agent bubbles undergo rapid fragmentation (time scale of microseconds), resulting in microbubbles of smaller size than the initial ones (Chomas *et al.*, 2000).

It has also been reported that in a 160-kHz standing wave field, oscillating contrast agent microbubbles can produce acoustic streaming pattern near them with the streaming velocity in the range of 50–100 μm/s (Gormley and Wu, 1998). Wu *et al.* (2002) showed that ultrasonic microstreaming generated by a 21.2-kHz Mason's horn could reversibly permeabilize cells.

Optical techniques of flow visualization, particle image velocimetry (PIV) in particular, have been successfully applied for several decades to acoustic streaming studies (Campbell *et al.*, 2000). In PIV instantaneous whole field velocities are obtained by measuring the distance traveled by tracer particles in the flow within a known time interval. Owing to their noninvasive character, flow visualization techniques have also been used for studying biological fluids including blood, where introduced microparticles or red blood cells were employed as tracers for the flow. PIV is a very common and useful tool to investigate the flow phenomena in models of blood vessels, heart valves, or artificial organs (Kim *et al.*, 2004; Balducci *et al.*, 2004; Nakano *et al.*, 2003).

The present paper reports the detection by microscopy and the evolution of microbubble(s) in the pressure node plane. Streaming patterns near these active regions are monitored by particle image velocimetry of 1-μm fluorescent latex microparticles. The patterns of erythrocyte movement are examined in the context of the streaming flows. The results contribute to the defining of the physical environment, in which sonoporation of a cell membrane occurs in standing wave fields. Moreover, since Optison® has found wide application *in vivo*, the processes triggered by the contrast agent microbubbles and their possible effects on living cells and tissues need to be clearly understood.

II. EXPERIMENT

A. Acoustic chamber

The cylindrical ultrasound standing wave chamber employed in the experiments, shown in Fig. 1, was described in detail by Khanna *et al.* (2003). Its main features are a PC 26 (Ferroperm, Denmark) disk transducer (nominal resonance frequency 1.5 MHz), a spacer for a water layer, and a 1-mm-thick quartz glass reflector. A stainless-steel layer separates the transducer from the water layer. The transducer's back electrode was etched to 8 mm in diameter to create a principal pressure minimum in the central, axial position in the nodal plane (Whitworth and Coakley, 1992; Spengler *et al.*, 2001).

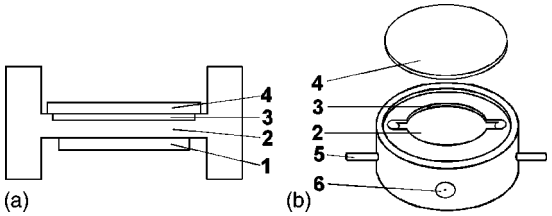


FIG. 1. Perpendicular section (a) and plan view (b) of the cylindrical acoustic resonator: (1) piezoelectric transducer; (2) coupling steel layer; (3) spacer for water layer; (4) quartz glass reflector; (5) water inlet (outlet); (6) electrical connection.

A manually controlled HP function synthesizer provided a sine wave input to an ENI 2100L amplifier. Frequency and applied voltage were controlled by computerized tracking of the driving frequency. The software STAND (Hawkes and Coakley, 1996) scanned the voltage–frequency spectra from a voltmeter and allowed the frequency to be held at a voltage maximum.

B. Measurement of pressure field

The acoustic pressure amplitude was estimated from the balance of the direct radiation and gravitational forces acting on a levitated 45-μm sphere in the standing wave (Spengler *et al.*, 2001).

C. Optical system and video recording

The observation of particle movement in the direction of sound propagation (z) was carried out with an Olympus BX41M epi-fluorescent microscope (Fig. 2). A standard PAL CCD JVC video camera (Victor Company, Japan) was connected via a 0.5 microscope adaptor and the images were recorded onto a standard video tape. A digital high-speed camera (Motionblitz, Germany) (DHSC), capable of capturing images at 500 frames/s, was installed on the second microscope port. Images captured through DHSC were automatically stored as .bmp files, which were directly opened into ADOBE PHOTOTSHOP 5.5 for further analysis.

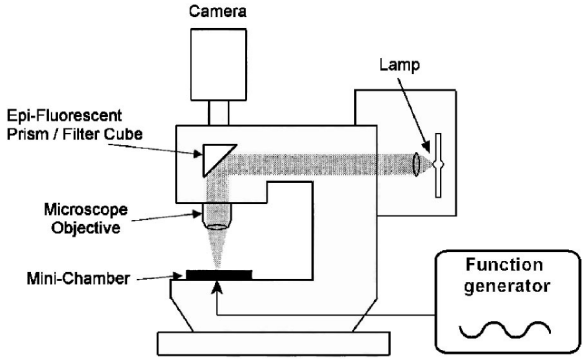


FIG. 2. Experimental setup.

D. Acoustic spectrum analysis

A 5-mm-diameter, 1.33-mm-thick PZ 26 piezoceramic disk (Ferroperm, Denmark) mounted in a cylindrical steel holder was used as a microphone. The assembly was coupled to the side of the chamber by ultrasonic transmission gel (Henleys Medical, UK). The signals from the microphone went to an Agilent E4401 digital spectrum analyzer (Agilent Technologies).

E. Particle image velocimetry analysis

Particle image velocimetry was carried out on recorded video clips showing sequence of events during ultrasound exposure. The recorded video sequences were transferred to a PC video card in digital format with hardware MJPEG data compression (Pinnacle Miro Video DC30+) to be further processed with PIV software. PIV software analyzes time sequences of the images of convected tracer microparticles and produces a vector map of the studied area, which exhibits the direction and velocity of the flow. Two PIV software packages were used for the estimation of velocity fields: PIV SLEUTH (Christensen *et al.*, 2000) and FLOWMANAGER developed by Dantec Dynamics (Denmark).

F. Contrast agent microbubble, erythrocyte, and microparticle source

Human albumin microspheres with octafluoropropane (Optison®) (Mallinckrodt Inc.) were used as contrast agent (CA) source for these experiments. A CA concentration of 75 μl ml^{-1} of suspension was employed. Human red blood cell (RBC) suspensions were prepared by diluting freshly drawn human blood to a final erythrocyte concentration of 10^7 ml^{-1} in phosphate buffered saline with 0.1% bovine serum albumin (PBS/BSA). Fluorescent microparticles (FM) of 1-μm size (Polysciences) were added at a concentration of $1-3\times10^7$ ml^{-1} of suspension.

G. Ultrasound exposure and postsonication analysis

Studies were undertaken involving three different situations where suspensions seeded with CA containing (a) RBC alone; (b) FM alone; or (c) RBC+FM were exposed to 1.56-MHz frequency, continuous USSW at pressures ranging from 0.98 to 1.96 MPa for periods ranging from 1 to 10 min. For all three situations on ultrasound application, the particles, cells, and microbubbles—a product of CA depletion—concentrated in the only (central) pressure node plane ($x-y$ plane) in the chamber. A suspension of CA alone in PBS/BSA was also exposed and observed microscopically for its behavior in the ultrasonic field.

A peristaltic pump (Gilson minipuls 3, France) was used to pump the sample into the chamber. The microscope was prefocused at a selected area. The ultrasound and the video recording began immediately after the pump was switched off (batch mode) (Spengler *et al.*, 2001). An initial frequency scan had identified a frequency at which the transducer voltage was at a maximum. This frequency showed small changes from day to day, in response to factors such as room temperature. All the events were recorded using video cameras by JVC and DHSC.

III. NUMERICAL MODELING

As a first step towards a numerical simulation of the phenomenon of cell circulation (Khanna *et al.*, 2003), a mathematical model proposed recently by Doinikov (2002) can be used that describes the coupled radial and translational dynamics of a gas bubble in a high-intensity plane standing wave. The extension of this model to three-dimensional wave systems yields

$$\left(1-\frac{\dot{R}}{c}\right)R\ddot{R}+\frac{3}{2}\left(1-\frac{\dot{R}}{3c}\right)\dot{R}^2=\left(1+\frac{\dot{R}}{c}\right)\frac{p_l}{\rho_l}+\frac{R}{c\rho_l}\frac{dp_l}{dt}+\frac{\dot{\mathbf{x}}^2}{4}, \tag{1}$$

$$\ddot{\mathbf{x}}+\frac{3\dot{R}}{R}\dot{\mathbf{x}}=\frac{3\mathbf{F}_{\mathrm{ex}}}{2\pi\rho_l R^3}, \tag{2}$$

where $R(t)$ and $\mathbf{x}(t)$ are the instantaneous radius and the position vector of the center of the bubble, respectively, the overdot denotes the time derivative, c is the sound speed in the surrounding liquid, p_l is the liquid pressure at the surface of the bubble, ρ_l is the liquid density, and $\mathbf{F}_{\mathrm{ex}}$ denotes external forces on the bubble, such as the primary Bjerknes force, the viscous drag, and the buoyancy force. Note that the added mass force is not included in $\mathbf{F}_{\mathrm{ex}}$ since it is already given by the left-hand side of Eq. (2). Equation (1) governs the volume oscillation of the bubble and Eq. (2) its translational motion. The term $\dot{\mathbf{x}}^2/4$ on the right-hand side of Eq. (1) provides feedback between the translation and the radial oscillation, and thus makes the equations of motion coupled. The pressure p_l is given by

$$p_l=\left(P_0+\frac{2\sigma}{R_0}\right)\left(\frac{R_0^3-a^3}{R^3-a^3}\right)^{\gamma}-\frac{2\sigma}{R}-\frac{4\eta\dot{R}}{R}-P_0-p_{\mathrm{ac}}, \tag{3}$$

where R_0 is the equilibrium radius of the bubble, P_0 is the hydrostatic pressure, σ is the surface tension, γ is the polytropic exponent of the gas within the bubble, a is the radius of the bubble's van der Waals hard core ($a=R_0/8.54$ for air), η is the liquid viscosity, and p_{ac} is the driving acoustic field. Assuming that the ultrasound field in the resonance chamber can be approximated by the zero cylindrical mode, p_{ac} is written as

$$P_{\mathrm{ac}}=-P_a J_0(k_r r)\cos(k_z z)\sin\omega t, \tag{4}$$

where P_a is the pressure amplitude, J_n is the cylindrical Bessel function of order n, z and r are the cylindrical coordinates in the axial and radial directions, respectively, ω is the angular frequency, and k_z, k_r are the axial and the radial wave numbers, which obey the following identity:

$$k_z^2+k_r^2=\omega^2/c^2. \tag{5}$$

Note that $z=0$ corresponds to the pressure antinode plane (the bottom of the resonator) and $z_n=\pi/2k_z$ to the pressure node plane.

The external forces $\mathbf{F}_{\mathrm{ex}}$ are set equal to the primary Bjerknes force, given by

$$\mathbf{F}_B = -\tfrac{4}{3}\pi R^3 \nabla p_{\mathrm{ac}}, \tag{6}$$

the buoyancy force

$$\mathbf{F}_{\mathrm{buoy}} = \tfrac{4}{3}\pi R^3 \rho_l g, \tag{7}$$

with g denoting the gravity, and the Levich viscous drag, given by

$$\mathbf{F}_{\mathrm{vis}} = -12\pi\eta R(\dot{\mathbf{x}} - \mathbf{v}_{\mathrm{ac}}), \tag{8}$$

where $\mathbf{v}_{\mathrm{ac}}$ denotes the liquid velocity that is generated by the driving acoustic field at the center of the bubble and specified by $\mathbf{v}_{\mathrm{ac}} = \nabla \varphi_{\mathrm{ac}}$, with

$$\varphi_{\mathrm{ac}} = -(P_a/\rho_l \omega) J_0(k_r r)\cos(k_z z)\cos \omega t. \tag{9}$$

If the normal component of the liquid velocity is assumed to vanish at the side wall of the resonator, the corresponding boundary condition is written as

$$J_1(k_r r_c) = 0, \tag{10}$$

where r_c is the inner radius of the resonance chamber. This condition allows us to determine the radial wave number: $k_r = X_{11}/r_c$, where $X_{11} \approx 3.831\,71$ is the lowest solution to the equation $J_1(X_{1m}) = 0$. It also means that the radial distribution of the acoustic pressure inside the resonator passes through zero at a distance r_n from the axis of the resonator, which is equal to $r_n = X_{01}/k_r = r_c X_{01}/X_{11}$, where $X_{01} \approx 2.404\,83$ is the lowest root of the equation $J_0(X_{0m}) = 0$.

Based on these equations, numerical simulations were made. The following material parameters were used: $P_0 = 0.1$ MPa, $\rho_l = 998$ kg/m^3, $\sigma = 0.0725$ N/m, $\eta = 0.001$ kg/(m s), $c = 1500$ m/s, and $\gamma = 1.4$. The driving frequency $f_0 = \omega/2\pi$, the inner radius of the resonator r_c, and the acoustic pressure amplitude P_a were set equal to $f_0 = 1.56$ MHz, $r_c = 7$ mm, and $P_a = 0.98$ MPa.

IV. RESULTS

A. Numerical simulation

The results obtained are presented in Figs. 3–5. Figure 3 illustrates the radial and translational dynamics of a bubble with the equilibrium radius $R_0 = 1.5\,\mu$m. The linear resonance frequency of the bubble is equal to $f_r = 1.85 f_0 = 2.88$ MHz. The radial oscillation of the bubble and its translational motion in the axial direction are displayed in Figs. 3(a) and (b), respectively. Figure 3(c) shows the path of the bubble in the rz coordinates, traversed during 5000 acoustic cycles. As can be seen, the bubble starts near the pressure antinode plane, at the point with the coordinates $r_0 = 0.3 r_n$ and $z_0 = 0.01 z_n$, translates rapidly to the pressure node plane (dashed line), and reciprocates slightly below it, shifting in the radial direction to the region of zero pressure amplitude, i.e., to the circle with $r = r_n$. A closer examination of Figs. 3(a) and (b) discloses some interesting details of this translation. As is seen in Fig. 3(a), at first the bubble undergoes violent radial oscillations, which change the sign of the primary radiation force and make the bubble move towards the pressure node. At the pressure node, the radial oscillations subside, but some peaks occur at regular time intervals. These are the moments of collapse. Comparison of Figs. 3(a) and (b) reveals that the sharp vertical jumps in the translational curve correspond to the moments of collapse in the radial motion. Thus, the pattern of translation appears as follows. When the initial violent radial oscillations die out, the bubble attempts to go back to the pressure antinode, but the radial oscillations again increase rapidly, and at the moment of collapse the bubble is strongly urged back to the pressure node.

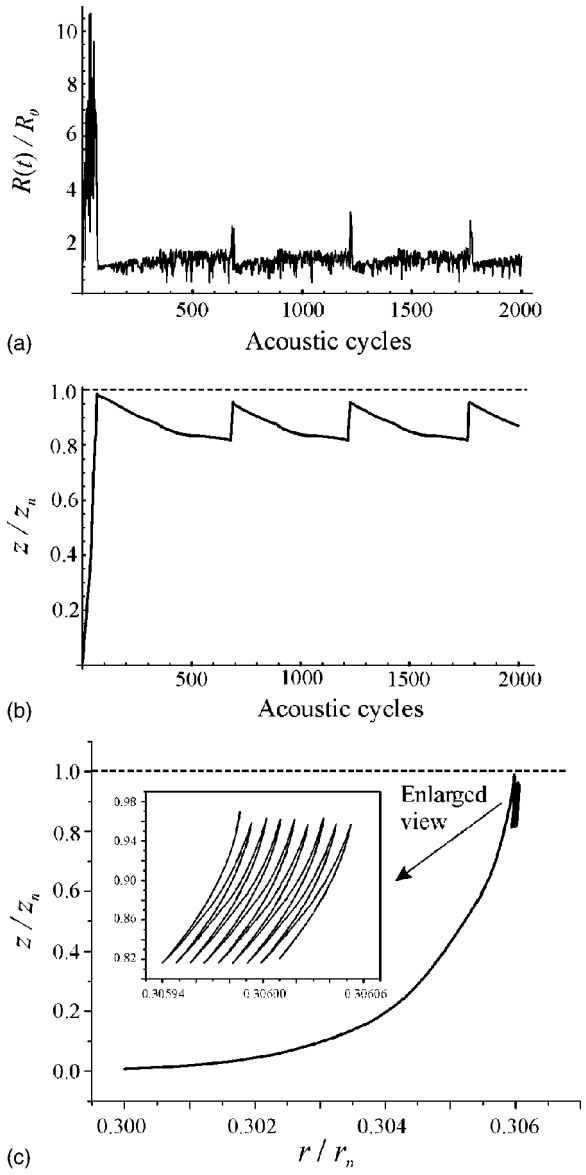


FIG. 3. Simulation of the radial and translational dynamics of a microbubble ($R_0 = 1.5\,\mu$m) in a high-intensity MHz cylindrical standing wave: (a) radial oscillation; (b) translational motion in the axial z direction; (c) bubble trajectory in the rz plane for 5000 acoustic cycles. The coordinates of the initial position of the bubble are $r_0 = 0.3 r_n$ and $z_0 = 0.01 z_n$.

Figure 4 displays Fourier spectra of the radial and translational motions shown in Figs. 3(a) and (b). Both spectra were calculated over the steady-state oscillations. Figure 5 presents the dependence of the spinning frequency f_s versus equilibrium bubble radius R_0.

B. Experiment

1. CA microbubbles microscopic observations

When the suspension containing CA alone was exposed to USSW at a pressure amplitude of 0.98 MPa, the bubbles

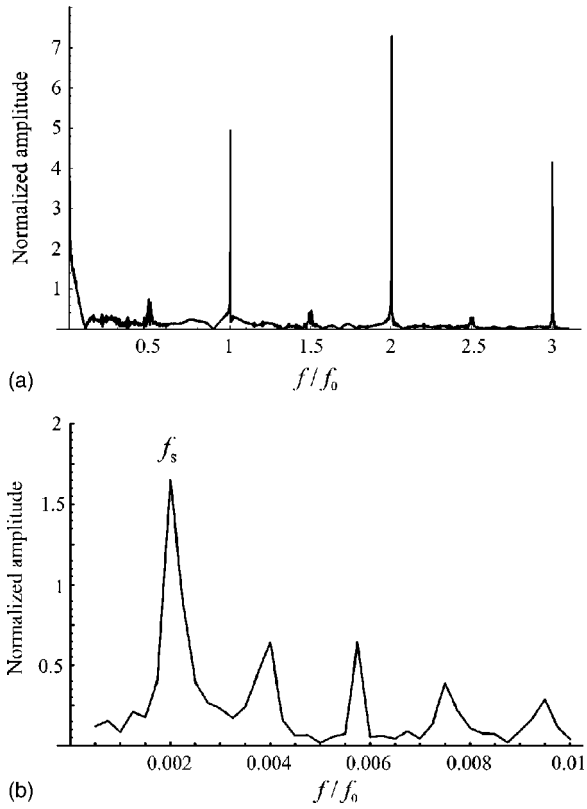


FIG. 4. Calculated Fourier spectra of (a) the radial and (b) the translational oscillations shown in Figs. 3(a) and (b), respectively. f_0 is the driving frequency.

of characteristic contrast agent size disappeared within a single frame (i.e., <40 ms). However, a group of highly active microbubbles was observed at the region of cell aggregation in the node plane [Fig. 6(b)]. The microbubble cloud depleted over a time span of a few seconds [Fig. 6(c)] and was lost completely within 10–15 s with a few microbubbles shooting out of the active region [Fig. 6(d)]. Pressures higher than 0.98 MPa were not tested for CA alone.

2. FM and CA

Multiple centers of microbubble activity were observed on ultrasound exposure. Fluorescent microparticles circulated in planes perpendicular to the node plane in a manner

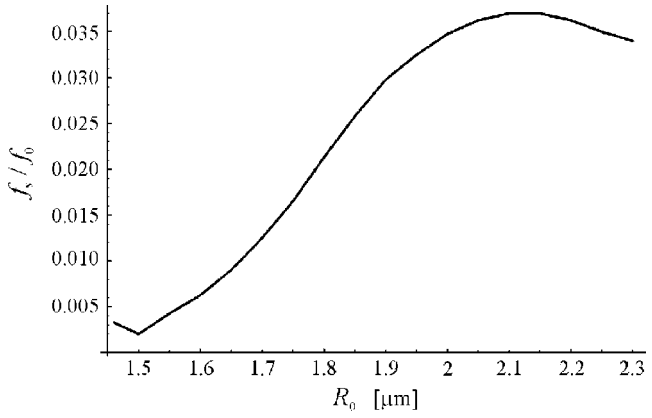


FIG. 5. Calculated bubble spinning frequency f_s versus equilibrium bubble radius R_0. Physical parameters are as in Fig. 3.

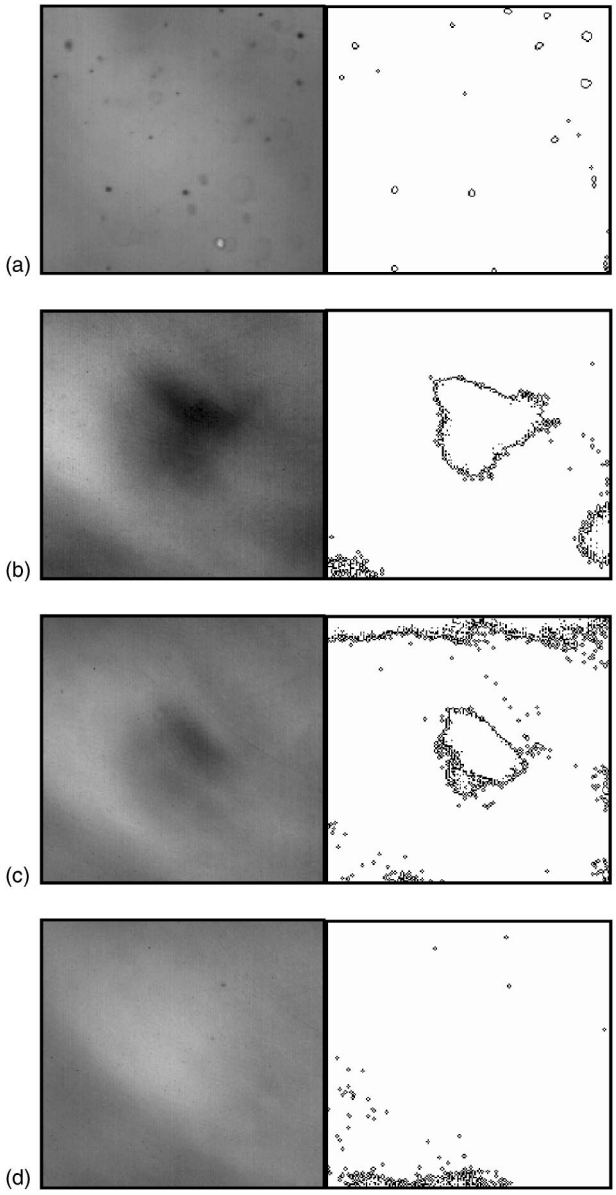


FIG. 6. Contrast agent alone in a 1.5-MHz ultrasound field, 0.98 MPa: (a) before ultrasound application; (b) 2 s; (c) 4 s; and (d) 8 s after ultrasound application; size of the field 0.01 mm^2, ×20 magnification.

similar to that reported earlier for erythrocytes (Khanna *et al.*, 2003). However, on the basis of the volume involved, two different scales of movement were noted. The axes of circulation were in or close to the node (or $x-y$) plane for a smaller (250–300-μm-diameter) scale movement and parallel to, but not necessarily close to, the node plane for a larger (2-mm) scale movement.

Particles in the smaller scale situation circulated in circular or elliptic orbits. Figure 7 presents a microscopic image of such a region and its vector map. Since the experimental observation and video recording were in the *z* direction, this and all the following vector maps are limited to the *x*- and *y* components of the particle velocity. The width of the areas in the presented horizontal plane provides an approximate diameter for the particle orbits. There were a number of small-scale circulation volumes which could migrate freely in any direction within the node plane during experiment. Some of

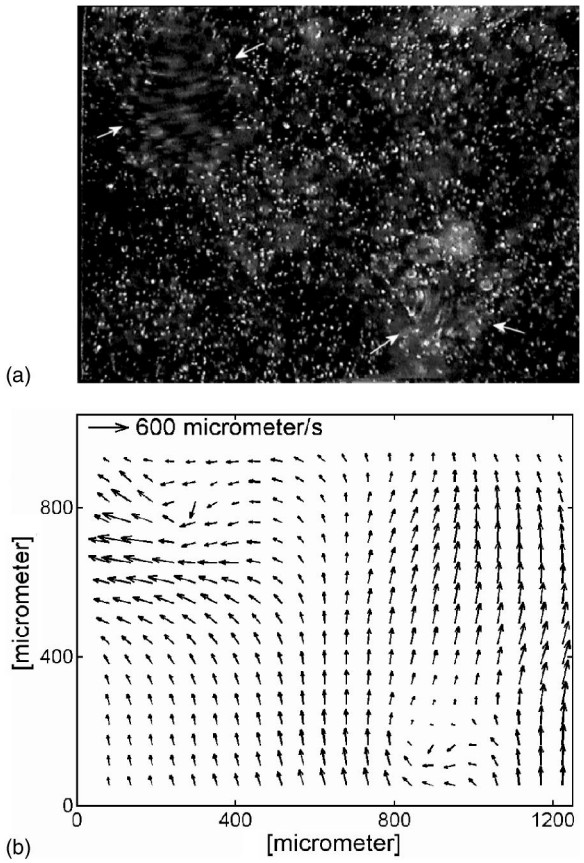


FIG. 7. Two examples of small-scale circulation (arrows) of 1-μm particles and CA about rotation axes in the node plane, 1.96 MPa: (a) image and (b) vector map of x- and y velocity components, $\times 10$ magnification.

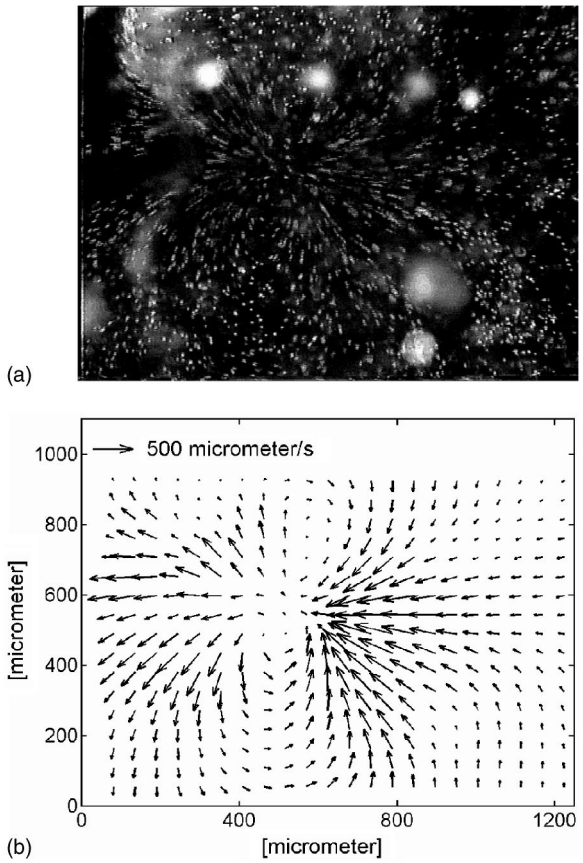


FIG. 8. Large-scale circulation of 1-μm particles and CA (''hourglass'' rotating about an axis parallel to the node plane), 0.98 MPa: (a) microscopic image, large bright spots are out-of-focus gas bubbles on the reflector surface, and (b) vector map of x- and y velocity components, $\times 10$ magnification.

the small-scale circulations stopped occasionally, probably due to a bubble collapse or disappearance from the nodal plane.

The larger 2-mm-diameter circulations usually occupied stable positions in the sonicated volume and could continue for the whole period of sonication. Images from the focused node plane suggested a cylindrical volume rotating about an axis parallel to the node plane. However, when the microscope was focused closer to the base of the chamber, the region involved in circulation had the appearance of a rotating hourglass, its axis being parallel to the node plane. A singular point could be identified as one towards which particle moved from all directions before participating in the circulation (Fig. 8).

Following several minutes of sonication, gas bubbles of up to 150 μm in diameter appeared on the inner surface of the glass reflector. The particle and cell circulations, though, were not dependent on the bubbles and were observed in their absence. There was vigorous motion of the 1-μm particles near the surfaces of most of these bubbles.

At the acoustic pressure amplitude of 1.96 MPa, the particles in the sonicated volume were involved in violent movements.

Acoustic spectrum analysis revealed the appearance of $f_0/2$ and $2f_0$ at 0.98 MPa. At 1.47 MPa, low levels of $3f_0/2$ were also detected and at 1.96 MPa white noise was observed (Fig. 9).

A feature apparently independent of the microbubbles activity was also observed in the chamber: a pair of counter-rotating streaming vortices of 3–4 mm in diameter circulating in the node plane, their axis being in the z direction. The vortices started immediately after ultrasound initiation and ran continuously for the whole period of sonication. Several small-scale circulations in planes perpendicular to the node plane were superimposed on these larger vortices, as shown in Fig. 10.

3. FM+RBC and CA

When suspension containing FM+RBC and CA microbubbles was exposed to ultrasound, small- and large-scale circulations similar to those described above were observed. A typical large-scale circulation is shown in Fig. 11. The acoustic spectrum was also the same as shown for the mixture of microparticles and contrast agent bubbles—Fig. 9.

4. RBC and CA

When RBCs in suspension seeded with CA were exposed to ultrasound, the cells rotated about an axis in the node plane as reported earlier by this group (Khanna *et al.* 2003). Acoustic spectra at different pressures were similar to those of FM and CA bubbles.

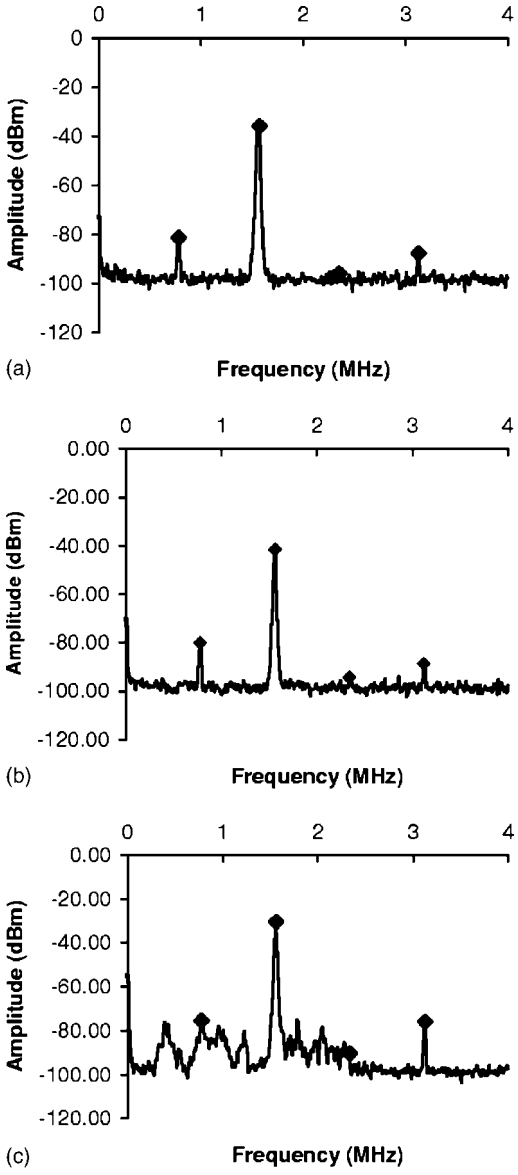


FIG. 9. Measured acoustic spectrum of microparticles and CA: (a) 0.98 MPa; (b) 1.47 MPa; (c) 1.96 MPa.

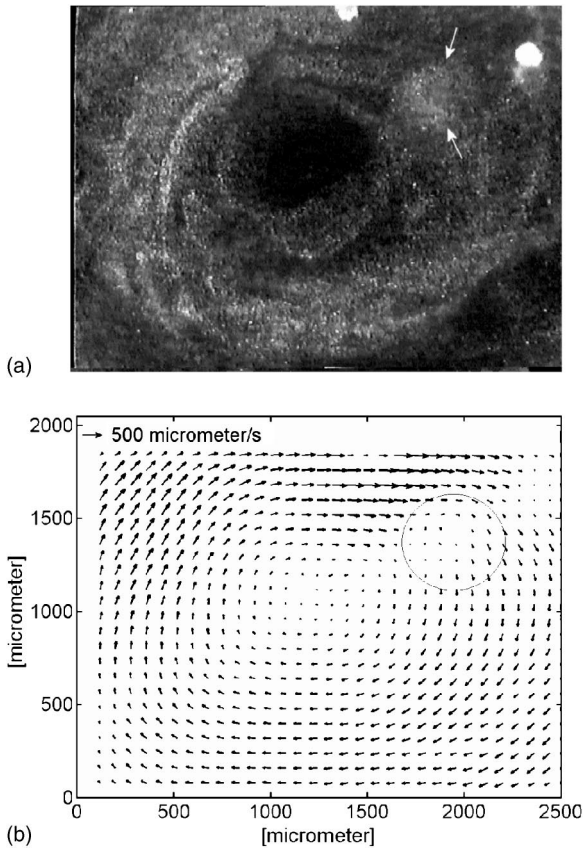


FIG. 10. Streaming vortices in $x-y$ plane (axes in z direction) with a superimposed small-scale particle circulation about an axis in the node plane: (a) image with "white cloud" indicated by arrows and (b) velocity map with encircled streaming pattern disruption; 1.96 MPa, ×5 magnification. The remaining bright regions are out-of-focus gas bubbles on the reflector surface.

V. DISCUSSION

Microscopic observation of ultrasound exposed suspension containing CA alone showed the presence of a group of highly active microbubbles at the region of cell aggregation in the node plane [Fig. 6(b)]. These microbubble clouds depleted over time [Fig. 6(c)] with a few microbubbles shooting out of the active region [Fig. 6(d)]. Since in some cases the phenomenon of particle circulation persists for 15–20 s and then subsides, it may be due to depletion/loss of active microbubbles from the region. Numerical simulation results presented in Fig. 3 demonstrate that, for parameters that conform to the experimental conditions, the theory does predict oscillatory translation of subresonant-size bubbles near the pressure node plane. It has been argued previously that these rapidly oscillating microbubbles might be capable of inducing hydrodynamic drag in the fluid which entrains particles and/or cells resulting in circulation (Khanna *et al.*, 2003).

Significant attention has been directed towards the detection of emissions from microbubbles and contrast agents (Miller, 1998; Chomas *et al.*, 2001). In the present experiments, onset of second harmonic ($2f_0$) and subharmonic ($f_0/2$) signals was closely associated with the circulation of particles and cells. We reported earlier that neither cell circulation nor subharmonic emission was observed in the absence of microbubbles; hence, microbubbles were required for the cell circulation (Khanna *et al.*, 2003).

A comparison of Fig. 9 and Fig. 4(a) shows that the calculated Fourier radial spectrum presents all the experimentally obtained components, although the calculated subharmonic ($f_0/2$) appears weak compared to the experimental spectrum. On the contrary, the theoretical second harmonic is too intensive. This is partly explained by the fact that the test bubble is close enough to the second harmonic resonance ($f_r = 1.85 f_0$). It is also possible that the calculations overestimate the power of the higher harmonics because some dissipative processes were missing. Finally, a considerable contribution to the experimental spectrum may come from bubbles other than those oscillating about the pressure node.

Figure 4(b) shows the calculated spectrum of the translational oscillation. The dominant (lowest) frequency of this spectrum will be referred to hereafter as the microbubble spinning frequency f_s. For the bubble under consideration, $f_s = 0.002 f_0 = 3120$ Hz. Calculations show that the microbubble spinning frequency increases with increasing equi-

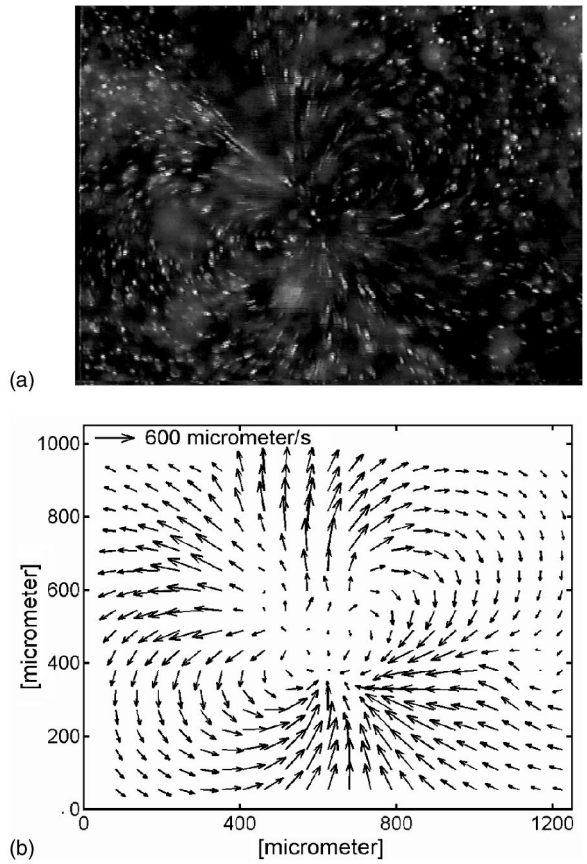


FIG. 11. Large-scale circulation of 1-μm particles, cells and CA (hourglass rotating about its axis parallel to the node plane), 0.98 MPa: (a) image and (b) velocity map.

librium bubble radius (Fig. 5). At the same time the amplitude of the translation decreases as the spinning frequency grows. It should also be noted that the spinning frequency depends on the initial value of the radial coordinate of the bubble, r_0. It decreases with increasing r_0, which is likely to be caused by the spatial decrease of the local driving pressure amplitude. Similar displacement of the starting point in the axial direction results in a very insignificant change of the spinning frequency. That is, if we fix r_0 and vary only z_0, the value of f_s is in fact constant; its change is within the accuracy of the analysis.

We have suggested previously that these rapidly oscillating microbubbles are capable of inducing hydrodynamic drag in the fluid which entrains particles and/or cells resulting in circulation. In the present work, however, when the areas of particles and cell circulation have been observed, the microbubble spinning frequency of the order of 3 kHz could not be experimentally resolved with the present imaging system. It could also be the case that sharp pressure and velocity gradients accompanying the presence of microbubbles at the node plane could drive the vortices that convect the cells and microparticles.

According to the numerical simulation when the equilibrium radius of a bubble exceeds resonance size ($R_{\mathrm{res}} = 2.52\ \mu$m in our case), the bubble is trapped at the pressure node and does not oscillate there. This result is already known from previous analyses (Doinikov, 2002). It is, however, found that the same occurs if the equilibrium bubble radius is smaller than the radius that corresponds to the second harmonic resonance. This latter radius is $R_{02} = 1.40\ \mu$m in our case. Calculations show that bubbles with $R_0 < R_{02}$ move towards the pressure node, execute in some cases one or more translations about it, and then settle there, not translating any more. If this prediction is valid, then the equilibrium radii of active microbubbles in the reported experiments should be in the range of about 1.40 to 2.52 μm. This also suggests that there are two mechanisms of depleting the oscillation region. First, bubbles can just leave this region due to translational motion. Second, their size can become outside the limits indicated above, due to rectified diffusion, coalescence, disintegration, etc., so that they stop actively oscillating.

Two scales of microparticles circulation *viz.*, small and large, were experimentally observed. The velocity of the particles participating in the circulation ranged from 200 to 600 μm/s. This is much greater than 100 μm/s streaming velocity detected by Gormley and Wu (1998) near Albunex spheres in a 160-kHz field. The particle rotation profiles in Figs. 7, 8, and 11 illustrate the scale of the circulations and give the *x*- and *y* components of the particle velocities. It is clear from the microscopy observations that the axes of particle circulation lie in or parallel to the $x-y$ plane. Velocites in the *z* direction are not determined directly in this work. Since observation suggests that the circulations have a near-spherical shape, it is assumed that the measured *x*- and *y* components give an estimate of the velocity component in the *z* direction.

Our large-scale circulation was characterized by a singular point towards which all the particles moved prior to participation in the circulation. The observed pattern might be similar to the streaming with vortex-ring elements near a pair of pulsating bubbles reported by Gormley and Wu (1998). Close observation of the area at bright field illumination in our work at $\times 20$ and $\times 50$ magnifications showed that the singular point was at, or at a small distance from, the base of the chamber. It has not been possible under the current experimental conditions to establish the exact position of it and the object, or objects, which initiated the circulation. The numerical modeling does not predict singular points in the pressure node plane of the zero cylindrical mode. Therefore, in the present simulation the bubble oscillates in one plane, which contains its starting point and is perpendicular to the pressure node plane. This oscillation in the described experiments apparently initiated a small-scale circulation of particles and cells about the node plane.

Streaming vortices circulated in the node plane like the ones shown in Fig. 10 might be caused by the nonuniformity of the lateral pressure distribution in the chamber. That kind of vortex induced in $x-y$ planes in the absence of microbubbles was reported recently for standing wave systems for a similar range of frequencies and acoustic pressures (Kuznetsova and Coakley, 2004). Bubbles would in the present case be expected to oscillate about preferred points in the node plane.

VI. CONCLUSIONS

Microbubbles activity in the pressure node plane has been visualized and recorded. Three-dimentional circulation

of 1-μm fluorescent latex beads and erythrocytes has been detected and the velocity field has been characterized by PIV. Acoustic emissions, i.e., subharmonics, ultraharmonics, and harmonics of the driving frequency were monitored. The numerical simulation predicts the movement of subresonance bubbles at a pressure node plane, and the calculated Fourier radial spectrum presents all the experimentally obtained components. The insight into the location and lifetime of the subresonance microbubbles in the node plane and the establishment of the streaming pattern associated with that will help clarify mechanisms involved in cell membrane poration in stationary wave fields.

ACKNOWLEDGMENT

L.A.K. was supported by the Biotechnology and Biological Sciences Research Council (Research Grant No. 72/E17416).

Apfel, R. E. (**1981**). ''Acoustic cavitation,'' in *Ultrasonics* (Academic, New York), pp. 391–393.

Balducci, A., Grigioni, M., Querzoli, G., Romano, G. P., Daniele, C., D'Avenio, G., and Barbaro, V. (**2004**). ''Investigation of the flow field downstream of an artificial heart valve by means of PIV and PTV,'' Exp. Fluids **36**, 204–213.

Campbell, M., Cosgrove, J. A., Greated, C. A., Jack, S., and Rockliff, D. (**2000**). ''Review of LDA and PIV applied to the measurement of sound and acoustic streaming,'' Opt. Laser Technol. **32**, 629–639.

Carstensen, E. L., Kelly, P., Church, C. C., Brayman, A. A., Child, S. Z., Raeman, C. H., and Schery, L. (**1993**). ''Lysis of erythrocytes by exposure to cw ultrasound,'' Ultrasound Med. Biol. **19**, 147–165.

Chomas, J. E., Dayton, P. A., May, D., and Allen, J. (**2000**). ''Optical observation of contrast agent destruction,'' Appl. Phys. Lett. **77**, 1056–1058.

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.

Christensen, K. T., Soloff, S. M., and Adrian, R. J. (**2000**). ''Integrated particle image velocimetry interrogation/validation software,'' Technical report No 943, Theor. Appl. Mechan., University of Illinois, Urbana-Champaign.

Doinikov, A. A. (**2002**). ''Translational motion of a spherical bubble in an acoustic standing wave of high intensity,'' Phys. Fluids **14**, 1420–1425.

Eller, A. (**1968**). ''Force on a bubble in a standing acoustic wave,'' J. Acoust. Soc. Am. **43**, 170–171.

Frenkel, P. A., Chen, S., Thai, T., Shohet, R. V., and Grayburn, P. A. (**2002**). ''DNA-loaded albumin microbubbles enhance ultrasound-mediated transfection *in vitro*,'' Ultrasound Med. Biol. **28**, 817–822.

Gormley, G., and Wu, J. (**1998**). ''Observation of acoustic streaming near Albunex® spheres,'' J. Acoust. Soc. Am. **104**, 3115–3118.

Hawkes, J. J., and Coakley, W. T. (**1996**). ''A continuous flow ultrasonic cell-filtering method,'' Enzyme Microb. Technol. **19**, 57–62.

Khanna, S., Amso, N. N., Paynter, S. J., and Coakley, W. T. (**2003**). ''Contrast agent bubble and erythrocyte behaviour in a 1.5 MHz standing ultrasound wave,'' Ultrasound Med. Biol. **29**, 1463–1470.

Kim, H. B., Hertzberg, J. R., and Shandas, R. (**2004**). ''Development and validation of echo PIV,'' Exp. Fluids **36**, 455–462.

Kuznetsova, L. A., and Coakley, W. T. (**2004**). ''Microparticle concentration in short path length ultrasonic resonators: Roles of radiation pressure and acoustic streaming,'' J. Acoust. Soc. Am. **116**, 1956 (1966).

May, D. J., and Allen, J. S. (**2002**). ''Dynamics and fragmentation of thick-shelled microbubbles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1400–1410.

Miller, D. L. (**1977**). ''Stable arrays of resonant bubbles in a 1-MHz standing-wave acoustic field,'' J. Acoust. Soc. Am. **62**, 12–19.

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, M. W., Miller, D. L., and Brayman, A. A. (**1996**). ''A review of *in vitro* bioeffects of inertial ultrasonic cavitation from a mechanistic perspective,'' Ultrasound Med. Biol. **22**, 1131–1154.

Miller, M. W., Everbach, E. C., Cox, C., Knapp, R. R., Brayman, A. A., and Sherman, T. A. (**2001**). ''A comparison of the hemolytic potential of Optison and Albunex in whole human blood *in vitro*: Acoustic pressure, ultrasound frequency, donor and passive cavitation detection considerations,'' Ultrasound Med. Biol. **27**, 709–721.

Nakano, A., Sugii, Y., Minamiyama, M., and Niimi, H. (**2003**). ''Measurement of red cell velocity in microvessels using particle image velocimetry (PIV),'' Clin. Hemorheol Microcirc. **29**, 445–455.

Neppiras, E. A., and Coakley, W. T. (**1976**). ''Acoustic cavitation in a focused field in water at 1 MHz,'' J. Sound Vib. **45**, 341–373.

Podell, S., Burraskano, C., Gaal, M., Golec, B., Maniquis, J., and Mehlhaff, P. (**1999**). ''Physical and biochemical stability of Optison®, an injectable ultrasound contrast agent,'' Biotechnol. Appl. Biochem. **30**, 213–223.

Spengler, J. F., Jekel, M., Christensen, K. T., Adrian, R. J., Hawkes, J. J., and Coakley, W. T. (**2001**). ''Observation of yeast cell movement and aggregation in a small-scale MHz-ultrasonic standing wave field,'' Bioseparation **6**, 329–341.

Watanabe, T., and Kukita, Y. (**1993**). ''Translational and radial motions of a bubble in an acoustic standing wave field,'' Phys. Fluids A **5**, 2682–2688.

Whitworth, G., and Coakley, W. T. (**1992**). ''Particle formation in a stationary ultrasonic field,'' J. Acoust. Soc. Am. **91**, 79–85.

Wu, J., Ross, J. P., and Chiu, J. F. (**2002**). ''Reparable sonoporation generated by microstreaming,'' J. Acoust. Soc. Am. **111**, 1460–1464.

Time domain simulation of nonlinear acoustic beams generated by rectangular pistons with application to harmonic imaging

Xinmai Yang and Robin O. Cleveland[a)]
Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215

(Received 10 June 2004; revised 11 October 2004; accepted 13 October 2004)

A time-domain numerical code (the so-called Texas code) that solves the Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation has been extended from an axis-symmetric coordinate system to a three-dimensional (3D) Cartesian coordinate system. The code accounts for diffraction (in the parabolic approximation), nonlinearity and absorption and dispersion associated with thermoviscous and relaxation processes. The 3D time domain code was shown to be in agreement with benchmark solutions for circular and rectangular sources, focused and unfocused beams, and linear and nonlinear propagation. The 3D code was used to model the nonlinear propagation of diagnostic ultrasound pulses through tissue. The prediction of the second-harmonic field was sensitive to the choice of frequency-dependent absorption: a frequency squared f^2 dependence produced a second-harmonic field which peaked closer to the transducer and had a lower amplitude than that computed for an $f^{1.1}$ dependence. In comparing spatial maps of the harmonics we found that the second harmonic had dramatically reduced amplitude in the near field and also lower amplitude side lobes in the focal region than the fundamental. These findings were consistent for both uniform and apodized sources and could be contributing factors in the improved imaging reported with clinical scanners using tissue harmonic imaging.
[DOI: 10.1121/1.1828671]

PACS numbers: 43.25.Cb, 43.35.Bf, 43.80.Qf [MFH] Pages: 113–123

I. INTRODUCTION

High amplitude sound waves are used in many applications. Two examples in biomedical acoustics are focused ultrasound surgery[1–3] and diagnostic imaging.[4,5] At the amplitudes used in these applications, effects associated with nonlinear distortion become important, for example, extra heating in tissue[6] and improved imaging capabilities.[7–9] Realistic simulations for these applications need to account for three-dimensional (3D) propagation because (1) the sources that are used, particularly in imaging, generate acoustic fields that are 3D and (2) tissue is inhomogeneous[10] and even for a perfect axis-symmetric source it will result in acoustic fields that are inherently 3D.

A popular model for the propagation of nonlinear directive sound beams is the Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation.[11,12] This equation accounts for diffraction (in the parabolic approximation), nonlinearity, and thermoviscous absorption. It can be modified to account for arbitrary absorption laws,[13] sound speed inhomogeneities,[14] and media with convection.[15] Numerical solutions of the axis-symmetric version (two spatial dimensions) of the KZK equation for a circular source have been widely investigated.[13,16–19] Numerical solutions for nonlinear axis-symmetric beams not restricted to the paraxial region have also been reported.[20–23] However, calculations for the non axis-symmetric case, for example, the field from a rectangular source, have been limited by the added computational cost of the extra dimension.

Computer technology has now advanced to the point that realistic three-dimensional problems in diagnostic ultrasound can be solved. Christopher[9] reported results using a frequency domain code, which accounts for diffraction exactly using the angular spectrum method, to simulate the propagation of finite-amplitude ultrasound through inhomogeneous tissue. Christopher's algorithm has been used by others to consider harmonic leakage in tissue harmonic imaging[24] and pulse-inversion harmonic imaging.[25] The so-called Bergen code, a frequency domain solution of the KZK equation, has also been used to simulate the propagation of ultrasonic beams of finite amplitude from rectangular and square sources.[26,27] Recently a three-dimensional time-domain code has been shown to be in agreement with measurements made in water of the pressure field generated by a clinical scanner.[28]

The main purpose of the present article is to describe a time-domain numerical code capable of simulating the propagation of ultrasonic beams of finite amplitude from a rectangular source. The 3D time-domain code was adapted from an algorithm developed by Lee and Hamilton for axis-symmetric sources[17]—the so-called ''Texas code.'' The 3D code accounts for the combined effects of diffraction, nonlinearity, thermoviscous absorption, and multiple relaxation phenomena.[29] Relaxation processes allow for frequency-dependent absorption and dispersion effects to be incorporated into the model. The code was used to consider spatial distribution of the fundamental and second-harmonic signal in tissue for an ultrasound-like imaging scenario and the importance of using the appropriate frequency-dependent absorption is demonstrated. The extension of the code to model

a)Electronic mail: robinc@bu.edu

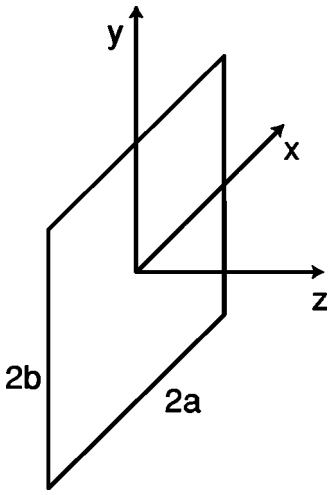


FIG. 1. Geometry and system of coordinates of a rectangular piston.

the propagation of nonlinear beams through an inhomogeneous medium is discussed.

II. MODEL EQUATION AND NUMERICAL METHOD

The KZK equation is a model for the paraxial region of a directive nonlinear sound beam. The original equation was developed for thermoviscous fluids, with a frequency squared absorption, but can be extended to fluids with arbitrary absorption and dispersion.[13] In Cartesian coordinates, the KZK equation, for a relaxing fluid, can be written in the following form:

$$\frac{\partial p}{\partial z}=\frac{c_0}{2}\int_{-\infty}^{t'}\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}\right)dt''+\frac{\beta}{2\rho_0 c_0^3}\frac{\partial p^2}{\partial t'}+\frac{\delta}{2c_0^3}\frac{\partial^2 p}{\partial t'^2}+\sum_v \frac{c_v'}{c_0^2}\int_{-\infty}^{t'}\frac{\partial^2 p}{\partial t''^2}e^{-(t'-t'')/t_v}dt'', \tag{1}$$

where p is the sound pressure, z the coordinate along the axis of the beam, x,y the transverse coordinates (see Fig. 1), $t'=t-z/c_0$ the retarded time, c_0 the small-signal sound speed (in the presence of dispersion this is the equilibrium, $f\to 0$, sound speed), δ the diffusivity of sound, ρ_0 the density, β the coefficient of nonlinearity, t_v the relaxation time, and c_v' the small-signal sound speed increment for each of the relaxation processes v (where $v=1,2,...$). The terms on the right-hand side account for diffraction, nonlinearity, thermoviscous absorption, and relaxation, respectively.

We obtain a numerical solution to Eq. (1) following the method of Lee and Hamilton.[17,29] First, Eq. (1) is transformed into nondimensional form on a Cartesian grid; the rationale for the Cartesian grid is discussed in Sec. II B. The following dimensionless variables are introduced:

$$\sigma=z/d,\quad X=x/a,\quad Y=y/b,\quad \tau=\omega_0 t',\quad P=p/p_0, \tag{2}$$

where d is a characteristic length in the direction of propagation (e.g., the focal length), the quantities a and b are the characteristic lengths in the x and y direction (e.g., the size of the source aperture), ω_0 is a characteristic angular frequency (e.g., working frequency of the source pulse), and p_0 is a characteristic pressure (e.g., peak source pressure). Substitution of Eq. (2) into Eq. (1) yields

$$\frac{\partial P}{\partial \sigma}=\frac{1}{4}\int_{-\infty}^{\tau}\left(\frac{1}{G_x}\frac{\partial^2 P}{\partial X^2}+\frac{1}{G_y}\frac{\partial^2 P}{\partial Y^2}\right)d\tau'+NP\frac{\partial P}{\partial \tau}+A\frac{\partial^2 P}{\partial \tau^2}+\sum_v D_v\int_{-\infty}^{\tau}\frac{\partial^2 P}{\partial \tau'^2}e^{-(\tau-\tau')/\theta_v}d\tau', \tag{3}$$

where the dimensionless coefficients are

$$G_x=k_0a^2/2d,\quad G_y=k_0b^2/2d,\quad N=d/\overline{z},$$
$$A=\alpha_0 d,\quad D_v=k_0 d c_v'/c_0,\quad \theta_v=\omega_0 t_v.$$

Here $k_0=\omega_0/c_0$ is the wave number, $\overline{z}=\rho_0 c_0^3/\beta\omega_0 p_0$ is the plane wave shock formation distance, and $\alpha_0=\delta\omega_0^2/2c_0^3$ is the thermoviscous absorption coefficient (Np/m) at the characteristic frequency.

Equation (3) is solved by marching in the principal propagation direction, σ. Operator splitting[17,30] is employed to separate Eq. (3) into the following:

$$\frac{\partial P}{\partial \sigma}=\frac{1}{4}\int_{-\infty}^{\tau}\frac{1}{G_x}\frac{\partial^2 P}{\partial X^2}d\tau', \tag{4a}$$

$$\frac{\partial P}{\partial \sigma}=\frac{1}{4}\int_{-\infty}^{\tau}\frac{1}{G_y}\frac{\partial^2 P}{\partial Y^2}d\tau', \tag{4b}$$

$$\frac{\partial P}{\partial \sigma}=NP\frac{\partial P}{\partial \tau}, \tag{4c}$$

$$\frac{\partial P}{\partial \sigma}=A\frac{\partial^2 P}{\partial \tau^2}, \tag{4d}$$

$$\left(1+\theta_v\frac{\partial}{\partial \tau}\right)\frac{\partial P}{\partial \sigma}=\theta_v D_v\frac{\partial^2 P}{\partial \tau^2},\quad v=1,2,\ldots. \tag{4e}$$

Each of these equations is solved sequentially at each marching step, $\Delta\sigma$. Equations (4a) and (4b) account for the effect of diffraction, Eq. (4c) the nonlinearity, Eq. (4d) thermoviscous absorption, and Eq. (4e) accounts for the effects of a finite number of individual relaxation processes.

The operator splitting methodology is valid if the step size is small enough that each effect introduces a small correction to the wave form.[30] For this reason it is numerically advantageous if the characteristic distance d is based on the length scale of the dominant process in the problem. If the step size $\Delta\sigma$ is chosen to be small enough to capture the dominant process, then it will also capture the dynamics of the other processes. For the case of ultrasound propagation in tissue (transducer characteristics: 15 mm$\times$35 mm aperture, 3.5 MHz working frequency, and 1 MPa source pressure) the approximate scales are absorption length 60 mm, shock formation distance 30 mm, focal length 70 mm, and Rayleigh distance 1200 mm. We note that the first three length scales have similar values, which means that any of them is an appropriate candidate for the characteristic distance d. For modeling an ultrasound scanner a convenient characteristic distance is the focal length associated with the elevation plane of the transducer because it is normally a fixed value while the other length scales may vary with either the operating frequency of the transducer or the beam forming.

In this work two finite-difference schemes were compared for calculating the diffraction effects, Eqs. (4a) and (4b). The first scheme is the same as that used in the two-dimensional (2D) Texas code and consisted of the implicit backward finite difference (IBFD) method for the first 100 marching steps ($\Delta\sigma=0.001$), in order to damp numerical oscillations, and the Crank–Nicolson finite difference (CNFD) method thereafter ($\Delta\sigma=0.01$).[17,30] An extra check is carried out at each step to ensure that the nonlinear distortion will not allow the wave form to become multivalued.[17] In the 3D code this is implemented by solving Eqs. (4a) and (4b) independently over each incremental step $\Delta\sigma$; as it is not possible to solve both diffraction terms simultaneously and still retain a system of equations that can be expressed as a tri-diagonal matrix, which has significant advantages from a computational point of view.[30]

The second scheme replaces the CNFD method with the Alternating Direction Implicit (ADI) method.[30] In the ADI method each diffraction step is broken into two half-steps: at the first half-step one transverse coordinate is solved implicitly and the other explicitly and this is alternated at the second half-step. In this case it is possible to combine Eqs. (4a) and (4b) to

$$\frac{\partial P}{\partial \sigma}=\frac{1}{4}\int_{-\infty}^{\tau}\left(\frac{1}{G_x}\frac{\partial^2 P}{\partial X^2}+\frac{1}{G_y}\frac{\partial^2 P}{\partial Y^2}\right)d\tau', \tag{5}$$

and still produce a tri-diagonal system at each half-step. This means for both CNFD and ADI it is necessary to solve two diffusion-like equations at each full step. The full step size of the ADI is the same as the CNFD and we found it was still necessary to employ the IBFD method for the first 100 marching steps in order to damp numerical oscillations. For both schemes the other equations are solved in the manner developed in the axis-symmetric Texas code.[17,29] Briefly, Eq. (4c) is solved exactly using the Poisson solution. Equations (4d) and (4e) are solved using initially an IBFD method followed, after the same 100 step transition, by the CNFD method.

A. Source conditions

The source condition for the KZK equation is determined by defining the time wave forms at $\sigma=0$ for all X and Y. In general the code is capable of having arbitrary excitation applied at the source location. Three source conditions were considered in this study.

(1) An unfocused uniformly excited piston, where the source condition (in dimensional variables) was expressed in the following form:

$$p=p_0 f(t)g(x,y) \quad \text{at } z=0, \tag{6}$$

where $f(t)$ was the source wave form and $g(x,y)$ the amplitude shading function. For uniform shading the function g is the two-dimensional top hat function, which in dimensionless co-ordinates is

$$g(X,Y)=\begin{cases}1, & -1\leqslant X\leqslant 1 \text{ and } -1\leqslant Y\leqslant 1\\ 0, & \text{otherwise.}\end{cases} \tag{7}$$

(2) A focused sound beam, where the source condition can be written within the parabolic approximation as follows:

$$p=p_0 f(t+x^2/2c_0 d_X+y^2/2c_0 d_Y)g(x,y) \quad \text{at } z=0.$$

In dimensionless variables this source condition takes the form:

$$P=f(\tau+G_{XS}X^2+G_{YS}Y^2)g(X,Y) \quad \text{at } \sigma=0, \tag{8}$$

where $G_{XS}=G_X d/d_X$ and $G_{YS}=G_Y d/d_Y$ ensure the correct phase in the source wave form to effect focusing at distance d_X in the xz plane and d_Y in the yz plane. The shading function $g(X,Y)$ was taken as a top-hat function.

(3) An apodized focused sound beam, where the focusing was effected using Eq. (8) but the amplitude was shaded using a cosine function in the X direction and thus impacting primarily the scan plane. The shading function used was

$$g(X,Y)=g_1(X)g_2(Y),$$

$$g_1(X)=\begin{cases}\cos\left(\frac{\pi}{2}X\right), & -1\leqslant X\leqslant 1\\ 0, & \text{otherwise}\end{cases} \tag{9}$$

$$g_2(Y)=\begin{cases}1, & -1\leqslant Y\leqslant 1\\ 0, & \text{otherwise.}\end{cases}$$

Apodization is commonly used in imaging systems to reduce side lobe levels.

B. Boundary conditions

In the lateral directions the Cartesian grid extended from $X_{\min}$ to $X_{\max}$ in the x axis with uniform spacing ΔX and from $Y_{\min}$ to $Y_{\max}$ in the y axis with uniform spacing ΔY. A zero pressure boundary condition was applied along the edges of the numerical domains, that is, at $X=X_{\max}$, $X=X_{\min}$, $Y=Y_{\max}$, and $Y=Y_{\min}$. The zero pressure boundary condition was simple to implement but did result in reflections from the edge of the numerical grid that can interfere with the desired signals in the central part of the grid. The time domain code described here used pulses and in this work the numerical boundaries were placed far enough away that the reflections did not interfere with the signals of interest. This is analogous to an experimental system where physical boundaries need to be far enough away to prevent real echoes affecting measurements. The trade-off with this approach is that the outer boundaries may need to be placed at very large distances to ensure that reflections do not interfere with the field of interest, which can result in a large computational cost particularly for long tone bursts.

One method of reducing the size of the grid and yet still avoid reflections is to use a transformed spatial grid whereby the outer boundary is allowed to move as the sound beam either focuses or spreads. A reflection is still generated but the boundary is placed such that the amplitude of the reflection is small. For example, in the case of an unfocused piston a spreading grid that approximates the field of a circular source in the linear limit has been used[31] and for highly focused beams a mixed converging/diverging grid has been used.[32,33] In this work we anticipate applying the algorithm

to the problem of biomedical ultrasound imaging which is a low focusing problem (typical gains around 3 to 10) and transforming the equation does not provide a significant advantage.[34] Therefore, the KZK equation was solved directly on a Cartesian grid without transformation.

We note that an alternative method to reduce boundary reflections is to employ an absorbing boundary layer at the outer edge of the grid. In this case the outer edge of the domain is approximated to appear as an infinite space. Examples include using the plane wave impedance condition,[35] application of a tapered spatial window,[26,36] or a perfectly matched layer which can be designed to absorb the sound before reflections occur.[37] Absorbing boundary layers allow the size of the numerical grid to be reduced but they add a level of complexity to the code. For the simulations investigated here, which involved short time pulses, it was possible to obtain solutions in a timely manner without implementing an absorbing boundary layer.

In special cases where it is known a priori that the field has symmetry, internal boundaries can be introduced to reduce the size of the numerical domain. For the case of symmetric sources, e.g., square or rectangular, propagating into a homogeneous or layered medium, it is possible to halve the size of the numerical domain for each axis of symmetry by placing an artificial rigid boundary along the axis. Therefore, in the case of a square or rectangular piston it is necessary to solve for only one quadrant of the $X-Y$ space. At each axis of symmetry one of the following boundary conditions was applied: $\partial P/\partial X|_{X=0}=0$ or $\partial P/\partial Y|_{Y=0}=0$. In the numerical implementation the second-order derivative at boundary was determined from a Taylor series expansion around the axis:

$$P(\Delta X)=p|_{X=0}+\Delta X\left(\frac{\partial P}{\partial X}\right)_{X=0}+\frac{(\Delta X)^2}{2}\left(\frac{\partial^2 P}{\partial X^2}\right)_{X=0}+O((\Delta X)^3).$$

Setting the first-order derivative to zero yields

$$\left.\frac{\partial^2 P}{\partial X^2}\right|_{X=0}=2\,\frac{P(\Delta X)-P(0)}{(\Delta X)^2}+O(\Delta X). \tag{10}$$

We note that this approximation is correct only to first-order in space, whereas all other finite-difference approximations are correct to second-order in space. Increasing the accuracy of this expression to second-order would result in losing the tri-diagonal nature of the algorithm, so the first-order expression was used in our simulations. An identical trade-off is made in the axis-symmetric Texas code. The 3D code was written to be able to include either axis of symmetry as required.

III. NUMERICAL RESULTS

The results from the code were benchmarked to other solutions in the literature for both circular pistons and rectangular pistons. The code was then used to simulate the propagation of a diagnostic ultrasound beam in tissue.

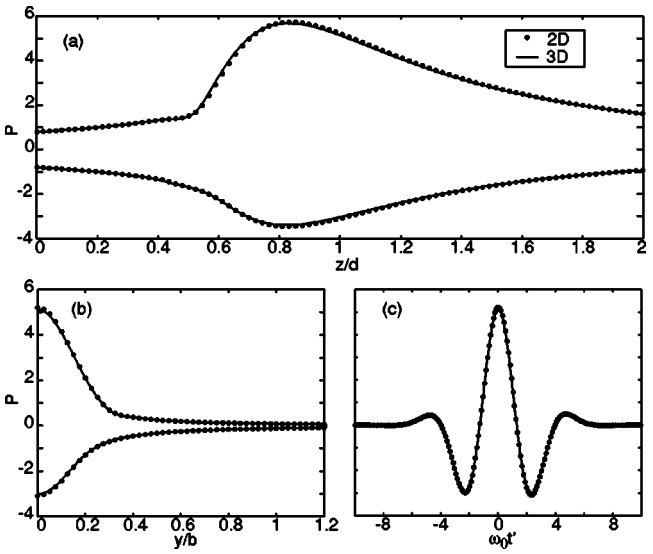


FIG. 2. Linear propagation from a circular piston: 2D vs 3D. (a) axial amplitude (z axis), (b) transverse beam plots at $\sigma=1/0$, (c) wave forms at $\sigma=1.0$.

A. Focused circular source: Linear and nonlinear

First, results of the 3D code were compared with the 2D Texas code for a focused circular symmetric source of radius a and focal length d. The source wave form was a smooth pulse given by[17]

$$P=\exp[-(\tau/n\pi)^{2m}]\sin\tau. \tag{11}$$

In this section the number of cycles $n=0.9549$ and the envelope exponent $m=1$ resulted in a short pulse about one-cycle in duration. The focusing gain of the transducer was taken to be $G=5.0$. The time window stretched from $\tau_{\min}=-20\pi$ to $\tau_{\max}=20\pi$ with a sampling rate $\Delta\tau=2\pi/60$, that is, there were 60 points per cycle and 1200 points in the time window. In the 2D code the grid had a maximum radius $\rho_{\max}=r/a=4$ and there were 160 grid points in radial direction ($\Delta\rho=0.025$). In the 3D simulation the symmetry of the problem allowed us to solve the problem on a quadrant with a lateral grid given by $0\leqslant X\leqslant X_{\max}$ or $0\leqslant Y\leqslant Y_{\max}$ with $X_{\max}=Y_{\max}=4$ and 160 grid points in the X and Y directions ($\Delta X=\Delta Y=0.025$). The step size was $\Delta\sigma=0.001$ for IBFD and $\Delta\sigma=0.01$ for CNFD. The memory requirement of the axis-symmetric code was on the order of 800 000 unknowns per marching step, for the 3D code the requirement was 500 000 000 unknowns per marching step. The memory and computational requirements became almost 1000 times larger by solving this particular problem in 3D.

The first scenario considered a linear lossless problem ($A=0,N=0$). Figure 2 compares axial propagation curves, focal beam patterns, and focal wave forms from the two codes. There is excellent agreement between the results of the 2D and 3D codes. A slight discrepancy exists in the positive peak of the wave form in Fig. 2(c). The discrepancy was the result of the extra dimension of discretization used in the 3D code (recall the 2D code assumed cylindrical symmetry and introduced no discretization error in the θ coordinate) and we found it could be reduced by using more lateral grid points in the 3D code.

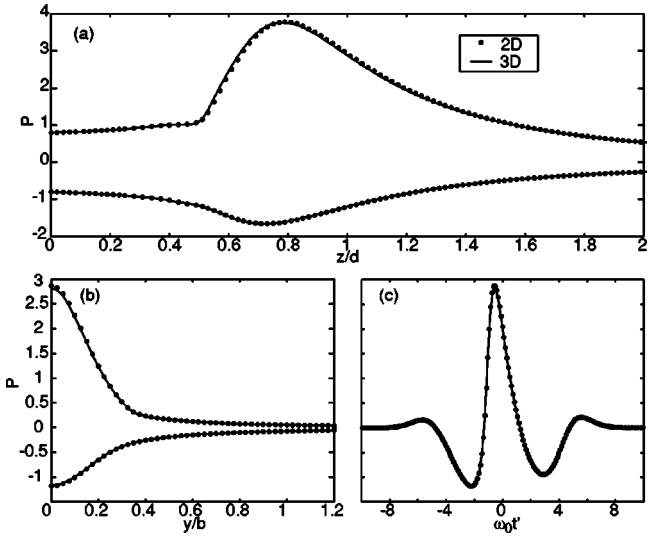


FIG. 3. Nonlinear propagation from a circular piston: 2D vs 3D. (a) Axial amplitude (z axis), (b) transverse beam plots at $\sigma=1/0$, (c) wave forms at $\sigma=1.0$.

Figure 3 shows a similar comparison for the same source condition and grid but for a case with nonlinearity and absorption present. The dimensionless nonlinearity and absorption coefficients were $N=1.0$ and $A=0.5$. The agreement is again excellent except in Fig. 3(c) where the 3D code slightly underpredicted the peak positive pressure due to the extra discretization in the 3D code.

B. Unfocused rectangular source: Linear and nonlinear

For the case of a uniform unfocused rectangular piston, we compared the result of a linear, lossless calculation using the 3D code with the results of the Rayleigh integral. We used the solution and excitation pulse of Ullate and San Emeterio.[38] The excitation was a short pulse of approximately 1.5 cycles at 3 MHz (see Fig. 4) which propagated in water ($c_0=1500$ m/s). The aperture size was 24×15 mm ($a=12$ mm and $b=7.5$ mm) with a resulting Rayleigh distance of 720 mm which, in the absence of a focal length, was used as the characteristic distance. Neither absorption nor nonlinearity was calculated in the simulations. The time window was from $\tau_{\min}=-20\pi$ to $\tau_{\max}=20\pi$ with 30 points per cycle, that is, $\Delta\tau=\pi/15$. The outer edges of the numerical grid were at $X_{\max}=Y_{\max}=4$ and there were 800 grid points in both the x and y directions ($\Delta X=\Delta Y=0.005$). The step size was $\Delta\sigma=5\times10^{-5}$ for IBFD and $\Delta\sigma=5\times10^{-4}$ for CNFD.

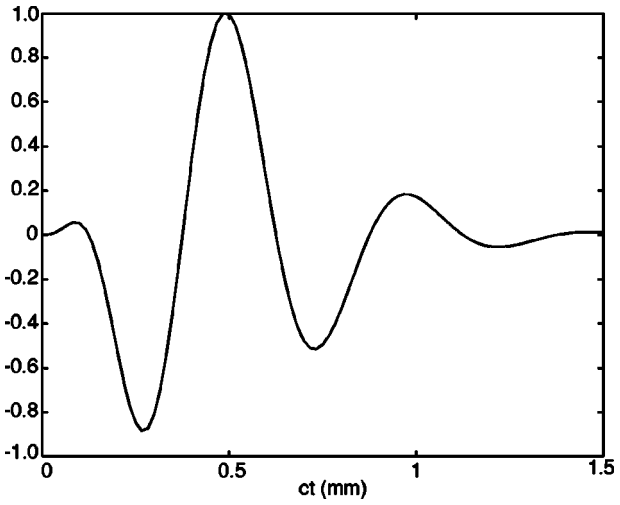


FIG. 4. Source velocity of rectangular piston for comparison with a Rayleigh integral solution (Ref. 38). The central frequency is $f_0=3$ MHz.

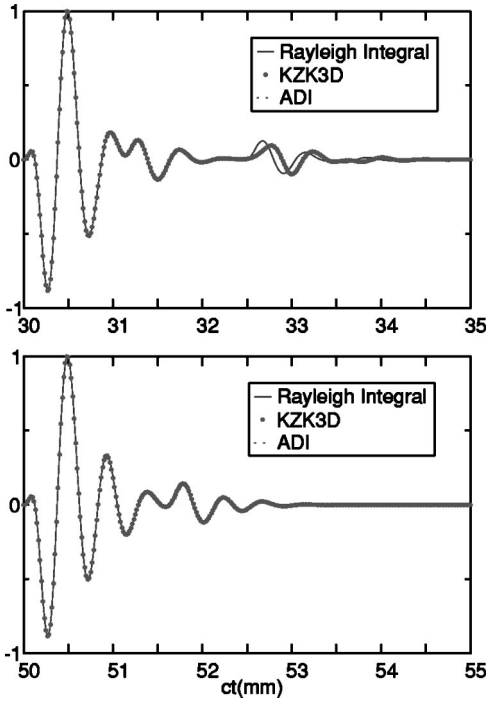


FIG. 5. Predicted axial wave forms at $z=30$ mm (upper) and $z=50$ mm (lower) for linear propagation from a rectangular aperture—comparison of the Rayleigh integral, KZK3D (CNFD) and KZK3D (ADI). The KZK code agrees well with the analytical solution except for the edge wave at $z=30$ mm where the parabolic approximation should start to break-down. There was no significant difference between the CNFD and ADI methods.

Figure 5 compares on-axis wave forms at $z=30$ mm and $z=50$ mm with the Rayleigh integral results. At $z=50$ mm there is excellent agreement between the exact solution and the KZK code. At $z=30$ mm, which is well within the near field of this transducer, there is excellent agreement for the portion of the wave that can be identified as arriving from the center of the transducer, but a small error in the arrival time of the edge wave. This discrepancy was due to the use of the parabolic approximation that is incorporated into the KZK equation. At a range of 30 mm the half-aperture angle to the corner of the source was 25° which is beyond what is usually acceptable for the parabolic approximation to be valid and, although rectangular apertures are less sensitive to errors in the edge wave,[26] the error manifests itself in the arrival time of the edge wave.

In Fig. 5 we also show results obtained using the ADI method on the same grid. The curves completely overlap with the results from IBFD/CNFD. However, the ADI method required about 10% more CPU time than the IBFD/CNFD, and nearly twice as much computational memory. Therefore we chose to use IBFD/CNFD algorithms for our code.

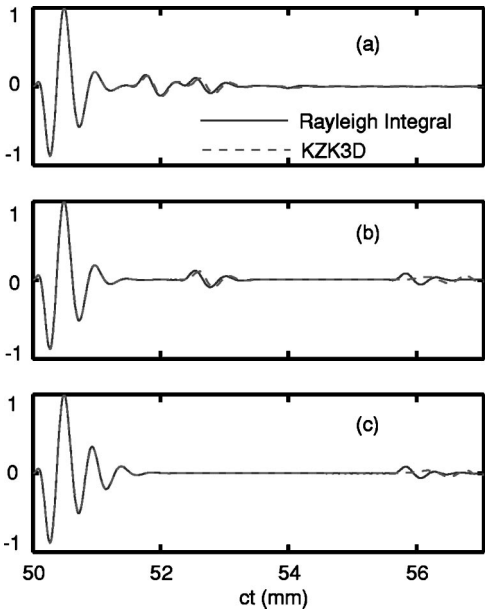


FIG. 6. Comparison of off-axis wave forms from a rectangular aperture at $z=50$ mm for (a) $x=0$, $y=b$, (b) $x=a$, $y=b$, and (c) $x=a$, $y=0$. There is generally excellent agreement between the analytical solution and the KZK3D code—the small discrepancies in the edge wave are attributable to the parabolic approximation.

Figure 6 shows three off-axis wave forms ($x=0$, $y=b$; $x=a$, $y=b$ and $x=a$, $y=0$) at a distance of $z=50$ mm. The agreement is generally excellent between the analytical solution and the KZK code with the exception of slight discrepancies in the arrival of the final edge wave. This is because even though at $z=50$ mm the parabolic approximation was fine on axis, the propagation of sound from the furthest points of the source occurred at too great an angle for the parabolic approximation to hold for field points that were substantially off axis.

For the nonlinear pressure field of rectangular apertures, we compared our code with results from the Bergen code for a square aperture given by Baker *et al.*[26] The parameters used in the simulation were $c_0=1486$ m/s, $\rho_0=1000$ kg/m^3, $\beta=3.5$, fundamental frequency 2.25 MHz, and $\alpha_0 d=0.094$. The time window was from $\tau_{\min}=-20\pi$ to $\tau_{\max}=20\pi$ with $\Delta\tau=\pi/32$ (64 points per cycle), the extent of the lateral grid was $X_{\max}=Y_{\max}=4$ with grid spacing $\Delta X=\Delta Y=0.025$ (160 points in each direction) and $\Delta\sigma=10^{-3}$ for IBFD and $\Delta\sigma=10^{-2}$ for CNFD. The time-domain code was used to simulate cw performance by using a 10 cycle long pulse [Eq. (10) with $n=10$ and $m=11$]. The amplitude of the harmonics in the time-domain code was determined by calculating the discrete Fourier transform of the center cycle in the pulse. Figure 7 shows the amplitude of the fundamental, second harmonic and third harmonic along the axis for a square aperture with amplitude of 220 kPa. We found excellent agreement in the predicted amplitude of the harmonics by the KZK code and the Bergen code.

C. Focused rectangular source: Nonlinear propagation in tissue

In this section results are presented for the propagation of a nonlinear ultrasound beam through tissue. In particular

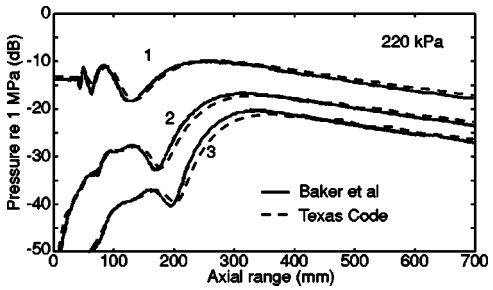


FIG. 7. Amplitude of the fundamental, second, and third harmonic as a function of propagation distance from a rectangular aperture. The results from this code compare favorably with those of Baker *et al.* (Ref. 26) for the same parameters.

we wish to investigate the spatial distribution of the fundamental and second-harmonic signals in tissue for application to diagnostic imaging. Tissue harmonic imaging (THI) is currently implemented on most mid-range and high-end commercial scanners, and in this mode images are formed from the reflected signals at the second harmonic rather than the fundamental. THI is currently the imaging mode of choice in cardiology where it has been shown to improve border delineation,[39,40] and measurements of heart function.[41] THI also appears to be promising in imaging other organs in the body, for example, it offers improved detection of cysts in the breast[42] and lesions in the liver[43] and kidney.[44]

THI is reported to provide enhancement in the spatial resolution of targets (''detail contrast'')[45] which has been attributed to the smaller focal spot,[24,39] immunity to aberration effects generated by near field inhomogeneities,[9,24] and nonlinear interaction at interfaces.[46] THI is also reported to provide superior ''contrast resolution'' as it is less sensitive to clutter.[24,39] Clutter refers to received echo signals that are generated outside the region of interest, for example, reverberation in the near field due to the presence of a fat layer near the skin or scattering by objects in the side-lobes. We investigated possible reasons for the image enhancement based on the distribution of the spatial distribution of the fundamental and second harmonic. We also evaluated the effect of source apodization and different tissue absorption laws on the predictions.

The source was modeled as a rectangular source with dimensions of 15 mm in the scan plane (x axis) and 10 mm in the elevation plane (y axis). We modeled the source as being continuously phased, that is, no individual array elements, however, the code is capable of modeling arbitrary conditions. The source was given two different geometrical focal distances: $d_x=80$ mm in the scan plane (normally done by electronic focusing in an imaging system) and $d_y=50$ mm in the elevation plane (normally a fixed focus in an imaging system) as most imaging is done with different focal lengths in the scan and elevation planes. The source was uniformly excited with a short pulse [Eq. (11) with $n=3$ cycles and $m=4$] and a center working frequency of 2 MHz. This wave form is similar to what will be used by a clinical scanner in THI mode, which uses more cycles than would be used in regular B-scan to improve the spectral purity in the source wave form. The pressure at the source was modeled

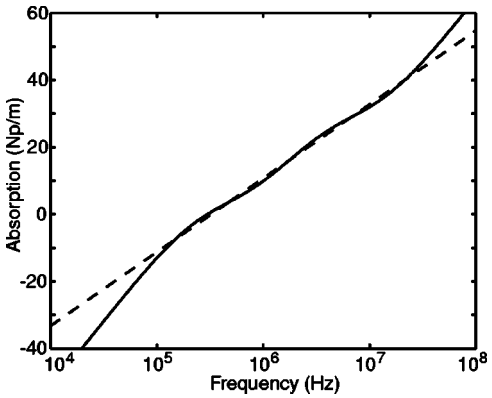


FIG. 8. Comparison of absorption curves for: two-relaxation process fit to a $f^{1.1}$ law (solid line) and thermoviscous absorption f^2 (dashed line).

to have an amplitude of 0.5 MPa resulting in a mechanical index (MI) in the tissue of approximately 0.7 (note this was based on the peak negative pressure predicted by the simulation and not from derating the peak pressure obtained from a simulation through water).

The tissue was modeled as a homogeneous medium with density $\rho_0 = 1050$ kg/m^3 and sound speed $c_0 = 1540$ m/s, and coefficient of nonlinearity $\beta = 5$. The absorption was assumed to obey the following power law:

$$\alpha = \alpha_0 \left(\frac{f}{f_0} \right)^{\eta}, \tag{12}$$

where $\alpha_0 = 3.4538$ Np/m, $\eta = 1.1$, and $f_0 = 1$ MHz (equivalent to 0.3 dB/cm at 1 MHz). The time-domain code cannot model such frequency-dependent absorption exactly; however, this behavior can be approximated by using a number of relaxation processes. In this simulation we used two relaxation processes to model the absorption properties of the tissue, that is, the effective absorption coefficient was

$$\alpha(f) = \frac{c'_{R1}(1 - jf)}{f^2 + f_{R1}^2} + \frac{c'_{R2}(1 - jf)}{f^2 + f_{R2}^2} + \alpha_{\mathrm{TV}} f^2. \tag{13}$$

The parameters c'_{R1}, f_{R1}, c'_{R2}, f_{R2}, and α_{TV} were chosen to give a best fit, in a least squares sense, to the absorption law given in Eq. (12) over the frequency band of 100 kHz to 30 MHz. The fminsearch function from MATLAB (The Mathworks, Natick, MA) was used to determine the parameters for the two relaxation processes and the thermoviscous absorption coefficient. The fitted parameters were: relaxation frequencies $f_{R1} = 237.0$ kHz and $f_{R2} = 3.749$ MHz; relaxation dispersion $(c'/c0)_{R1} = 1.3990 \times 10^{-3}$ and $(c'/c0)_{R2} = 1.6597 \times 10^{-3}$, and the thermoviscous absorption coefficient $\alpha_{\mathrm{TV}} = 1.625 \times 10^{-13}$ Np/m/Hz^2. The comparison of the power-law absorption and the fit using relaxation processes is shown in Fig. 8. There is excellent agreement over the frequency band of interest. We note that the relaxation operator also correctly accounts for dispersion, that is, the Kramers–Kronig relations are satisfied by the absorption given in expression in Eq. (13).[47]

The dimensionless parameters for this simulation were $N = 0.6554$, $G_X = 2.869$, $G_Y = 1.275$, $A = 0.0520$, $D_1 = 0.9133$, $\theta_1 = 8.4377$, $D_2 = 1.0835$, and $\theta_2 = 0.5334$. The focusing at the source was $G_{XS} = 2.869$ and $G_{YS} = 2.04$. The

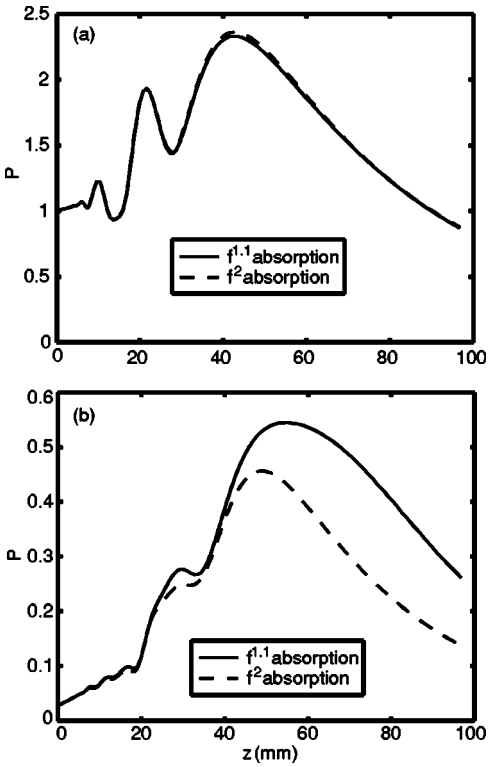


FIG. 9. Simulation of nonlinear propagation in a tissue-like medium. Axial plots of the amplitude of (a) the fundamental and (b) the second harmonic. The solid line is the result for $f^{1.1}$ absorption and the dashed line for f^2 absorption.

time window was from $\tau_{\min} = -20\pi$ to $\tau_{\max} = 20\pi$ with $\Delta\tau = 2\pi/64$, $X_{\max} = Y_{\max} = 4$, and $\Delta X = \Delta Y = 0.01$. The step size was $\Delta\sigma = 10^{-5}$ for IBFD and $\Delta\sigma = 10^{-2}$ for CNFD.

In Fig. 9 we present plots of the axial amplitude of the fundamental and second harmonic in the tissue. In this case the amplitudes were determined by carrying out a discrete Fourier transform on a five-cycle window around the pulse. The solid lines are the predictions for $f^{1.1}$ absorption and the dashed lines for f^2 absorption. The coefficient of the f^2 absorption law was chosen to yield the same absorption as the $f^{1.1}$ law at the fundamental frequency of 2 MHz. The results show that the absorption law had little impact on the prediction of the fundamental. This is not unexpected as the two absorption laws match at this frequency. However, there was a substantial difference in the prediction of the second harmonic, in particular the peak amplitude of the second harmonic for the f^2 simulation was reduced by more than 20% and the location of the peak shifted toward the source by 6 mm. Both of these phenomena are consistent with the fact that f^2 absorption will attenuate the second harmonic more strongly than $f^{1.1}$ absorption. We note that the amplitude of the fundamental in the $f^{1.1}$ case was slightly lower in amplitude in the focal region which is because the second harmonic has higher amplitude for the $f^{1.1}$ case and this resulted in more efficient transfer of energy from the fundamental to the third and higher harmonics.

Figure 10 shows contour plots of the amplitude of the fundamental and second harmonic of the pulse in the scan plane. The solid lines are the predictions for $f^{1.1}$ absorption and the dashed lines for f^2 absorption. We see that the off-

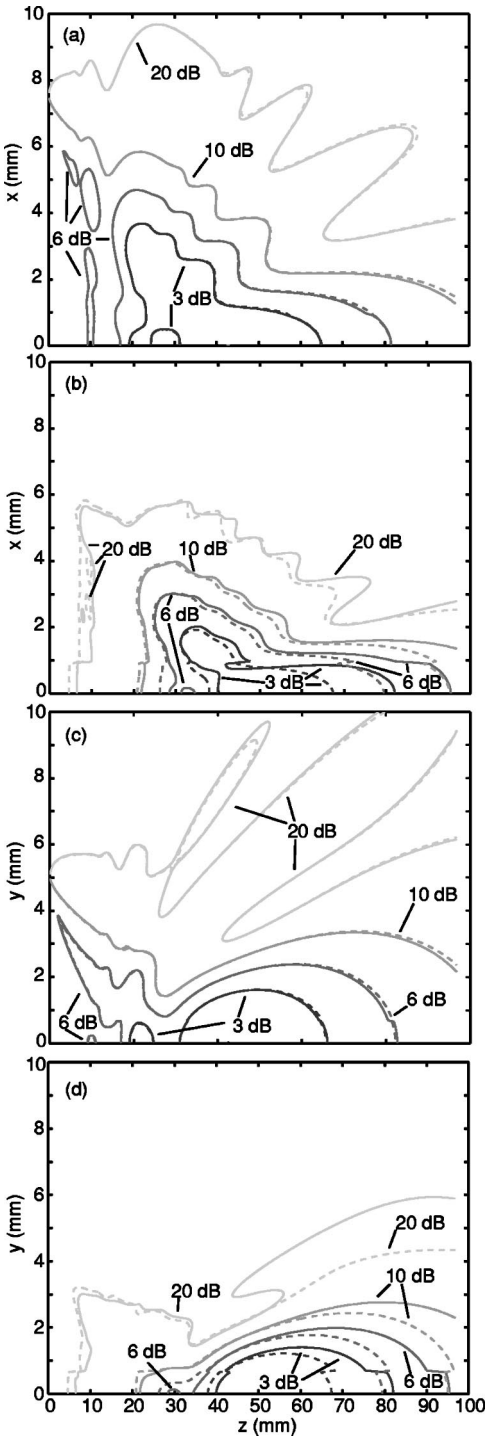


FIG. 10. Simulation of nonlinear propagation in a tissue-like medium. Contour plots of the amplitude of (a) the fundamental in the scan ($x-z$) plane, (b) the second harmonic in the scan plane, (c) the fundamental in the elevation ($y-z$) plane, and (d) the second harmonic in the elevation plane. The reference value was the peak amplitude for each condition. The solid line is the result for $f^{1.1}$ absorption and the dashed line for f^2 absorption.

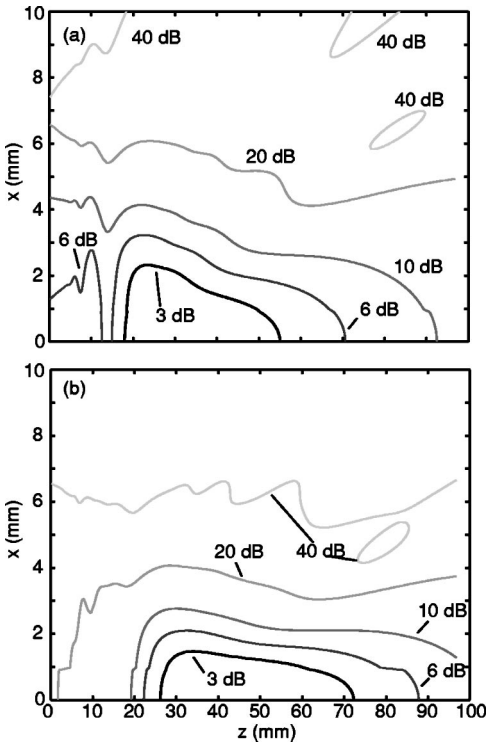


FIG. 11. Simulation of nonlinear propagation in a tissue-like medium ($f^{1.1}$ absorption) with a cosine amplitude shading of the source in the scan plane. Contour plots of (a) the fundamental and (b) the second harmonic in the scan ($x-z$) plane.

axis distribution of the second harmonic was not as sensitive as the on-axis distribution to the use of an incorrect absorption. In comparing Figs. 10(a) and 10(b), we note that the second-harmonic signal in the scan plane was more spatially confined than the fundamental. The distance to the -3 dB contour was ~ 0.9 mm for the second harmonic and ~ 1.3 mm for the fundamental which could contribute to improvements in detail resolution on-axis. Further, the -20 dB contour of the second harmonic barely extends more than 5 mm from the axis where as for the fundamental it extends up to 10 mm away which means that the second harmonic should suffer less from clutter signals due to off-axis scatterers. Lastly, in the near field of the transducer the levels of the second harmonic (~ -10 dB at $z=20$ mm) are much less than those of the fundamental (~ -3 dB at $z=20$ mm) and therefore the second-harmonic signal should suffer from less clutter in the presence of reverberating layers. Figures 10(c) and 10(d) show the contour plots in the elevation plane, and similar features can be observed in these plots. We note that repeating these simulations with a larger grid $X_{\max}=Y_{\max}=6$ produced no changes to the results, and this indicates that the data were not affected by reflections from the edge of the numerical grid.

Finally, we demonstrate the impact of apodization on the field of the fundamental and second harmonic for propagation through tissue (with $f^{1.1}$ absorption law). The source condition described by Eq. (9) was employed. Figure 11 shows contour plots of the fundamental and the second harmonic in the scan plane. Compared to Fig. 10, one observes that the apodization greatly reduced the amplitude of the side lodes of the fundamental as expected. The apodization reduced the ''side-lobes'' of the second-harmonic signal as well. Contour levels are shown down to -40 dB and for the second-harmonic signal the -40 dB contour rarely extends to approximately 6 mm from the axis whereas for the fundamental the -20 dB contour lies around 6 mm off axis and the -40 dB contours extends mostly beyond the range

shown. The effect of the source apodization on the acoustic field in the elevation plane was much less dramatic (data not shown), which is consistent with the fact that no apodization was implemented in that axis. Therefore, even though the beam pattern of the fundamental signal is improved by apodization, the second harmonic still has a narrower beam (with lower side lobes) and much less near field structure than the fundamental.

IV. CONCLUSIONS AND DISCUSSION

A time-domain algorithm that solves the KZK equation in three spatial dimensions has been described. The code includes the effects of diffraction, absorption, nonlinearity, and relaxation on the propagation of the finite-amplitude sound beams and is capable of describing the propagation of arbitrary sound beams in a wide variety of media. The algorithm used an operator-splitting algorithm to integrate the KZK equation. The 2D diffraction term was solved using a combined IBFD/CNFD method which we found to produce similar results to the ADI method but at lower computational cost. Results obtained from the algorithm were in excellent agreement with linear and nonlinear fields from a focused circular source; and the linear and nonlinear fields of unfocused rectangular sources. We note that the 3D code is computationally intensive but can be parallelized very effectively. Our simulations were carried out on an IBM pSeries 690 machine with 16 parallel processors; the tissue simulations shown in Figs. 10 and 11 each required a total of 9 CPU-hours and executed in about 0.6 wall-clock hours, which is close to the theoretical minimum of 0.56 wall-clock hours.

The code was used to simulate the field from a diagnostic ultrasound scanner. It was observed that the choice of absorption law had a strong effect on the prediction of the second-harmonic field. In comparing the spatial distribution of the fundamental and second-harmonic fields, we observed that the second-harmonic field was more sharply focused and that the second harmonic had much lower signal amplitude in the near field. These phenomena were present with both uniform and apodized sources. Both of these effects could be contributing factors to the improved imaging ability attributed to tissue harmonic imaging. The tighter focus should improve detail contrast, in particular by improving the lateral resolution. Clutter will be reduced both due to the narrower beam, which results in lower pressure levels off-axis, and due to the lower levels in the near-field, which results in less reverberation.

One effect not modeled with the 3D code at present is the effect of an inhomogeneous medium. The current model can be extended to account for small fluctuations in ambient properties. For variations in sound speed one may add a small fluctuation $c'(x,y,z)$ that can vary in space to the background sound speed c_0. Within the ordering frame work of the KZK equation this introduces one extra term which essentially alters the phase to correct for the variation in sound speed.[14] If the medium also has small variations in the density change to $\rho_0(x,y,z)$ then a new term $\nabla p \cdot \nabla \rho_0/\rho_0$ is introduced to the full wave equation.[48] In the parabolic approximation, to leading order, one finds that the density term reduces to

$$\frac{1}{c_0\rho_0}\frac{\partial\rho_0}{\partial z}\frac{\partial p}{\partial t'}-\frac{1}{\rho_0}\left(\frac{\partial\rho_0}{\partial x}\frac{\partial p}{\partial x}+\frac{\partial\rho_0}{\partial y}\frac{\partial p}{\partial y}\right). \tag{14}$$

Small scale variations in other parameters, i.e., nonlinearity, absorption, and dispersion, do not introduce extra terms into the second-order wave equation. For an inhomogeneous medium Eq. (1) can be rewritten in the following form:

$$\begin{aligned}\frac{\partial p}{\partial z}=&\frac{c_0}{2}\int_{-\infty}^{t'}\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial y^2}\right)dt''+\frac{\beta}{2\rho_0 c_0^3}\frac{\partial p^2}{\partial t'}+\frac{\delta}{2c_0^3}\frac{\partial^2 p}{\partial t'^2}\\&+\sum_v\frac{c_v'}{c_0^2}\int_{-\infty}^{t'}\frac{\partial^2 p}{\partial t''^2}e^{-(t'-t'')/t_v}dt''+\frac{c'}{c_0^2}\frac{\partial p}{\partial t}\\&+\frac{1}{2\rho_0}\frac{\partial\rho_0}{\partial z}p-\frac{c_0}{2\rho_0}\int_{-\infty}^{t}\frac{\partial\rho_0}{\partial x}\frac{\partial p}{\partial x}+\frac{\partial\rho_0}{\partial y}\frac{\partial p}{\partial y}dt.\end{aligned} \tag{15}$$

The added terms can be incorporated in the current numerical scheme. The phase-speed term associated with sound speed could be included into the nonlinear term, as an amplitude-independent effect, and calculated using the Poisson solution. The transverse density fluctuation could be incorporated into the diffraction term using finite-differences to evaluate the first-order derivatives. The density fluctuation in the axial direction is the only term that requires a new equation to be included into the operator splitting paradigm:

$$\frac{\partial p}{\partial z}=\frac{1}{2\rho_0}\frac{\partial\rho_0}{\partial z}p.$$

This is an ordinary first-order differential equation that can be approximated with a finite-difference solution

$$p(x,y,z+\Delta z)=p(x,y,z)\frac{2\rho_0(x,y,z+\Delta z)}{\rho_0(x,y,z+\Delta z)+\rho_0(x,y,z)}.$$

This corresponds to the plane wave pressure transmission coefficient for the propagation between two media of different densities.

In conclusion the extension of the Texas code to 3D provides an effective tool for modeling arbitrary finite-amplitude beams. In its current form it can account for propagation of arbitrary beams in a thermoviscous and relaxing fluid. We speculate that it may be advantageous over frequency domain codes in problems that involve short pulsed wave forms, for example, shock wave therapy and the forward propagation problem in ultrasound imaging. The present formulation can be extended to account for propagation in inhomogeneous media.

ACKNOWLEDGMENTS

The authors would like to thank Yuan Jing for providing comments on this manuscript. Computing resources were provided by the Scientific Computing and Visualisation group at Boston University. Partial funding was provided by National Institutes of Health through Grant No.

P01-DK53881 and by CenSSIS, the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award No. EEC-9986821).

[1] A. J. Coleman and J. E. Saunders, ''A review of the physical properties and biological effects of the high amplitude acoustic fields used in extracorporeal lithotripsy,'' Ultrasonics **31**, 75–89 (1993).
[2] K. Hynynen, ''Demonstration of enhanced temperature elevation due to nonlinear propagation of focused ultrasound in a dog's thigh in vivo,'' Ultrasound Med. Biol. **13**, 85–91 (1987).
[3] S. Vaezy, R. Martin, H. Yaziji, P. Kaczkowski, G. Keilman, S. Carter, M. Caps, E. Y. Chi, M. Bailey, and L. A. Crum, ''Hemostasis of Punctured Blood Vessels Using High-Intensity Focused Ultrasound,'' Ultrasound Med. Biol. **24**, 903–910 (1998).
[4] E. L. Carstensen, W. K. Law, N. D. McKay, and T. G. Muir, ''Demonstration of nonlinear acoustic effects at biomedical frequencies and intensities,'' Ultrasound Med. Biol. **6**, 359–368 (1980).
[5] H. C. Starritt, M. A. Perkins, F. A. Duck, and V. F. Humphrey, ''Evidence for ultrasonic finite-amplitude distortion in muscle using medical equipment,'' J. Acoust. Soc. Am. **77**, 302–306 (1985).
[6] D. R. Bacon and E. L. Carstensen, ''Increased heating by diagnostic ultrasound due to nonlinear propagation,'' J. Acoust. Soc. Am. **88**, 26–34 (1990).
[7] B. Ward, A. C. Baker, and V. F. Humphrey, ''Nonlinear propagation applied to the improvement of resolution diagnostic medical ultrasound,'' J. Acoust. Soc. Am. **101**, 143–154 (1997).
[8] M. A. Averkiou, D. N. Roundhill, and J. E. Powers, ''A new imaging technique based on the nonlinear properties of tissues,'' Proceedings of the IEEE Ultrasonics Symposium, IEEE Catalog No. 97CH36118, 1997, vol. 2, pp. 1561–1566.
[9] T. Christopher, ''Finite amplitude distortion-based inhomogeneous pulse echo ultrasonic imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 125–139 (1997).
[10] T. D. Mast, L. M. Hinkelman, M. J. Orr, V. W. Sparrow, and R. C. Waag, ''Simulation of ultrasonic pulse propagation through the abdominal wall,'' J. Acoust. Soc. Am. **102**, 1177–1190 (1997).
[11] E. A. Zabolotskaya and R. V. Khokhlov, ''Quasi-plane waves in the nonlinear acoustics of confined beams,'' Sov. Phys. Acoust. **15**, 35–40 (1969).
[12] V. P. Kuznetsov, ''Equation of nonlinear acoustics,'' Sov. Phys. Acoust. **16**, 467–470 (1970).
[13] S. I. Aanonsen, T. Barkve, J. Naze Tjøtta, and S. Tjøtta, ''Distortion and harmonic generation in the nearfield of a finite amplitude sound beam,'' J. Acoust. Soc. Am. **75**, 749–768 (1984).
[14] B. Lipkens, P. Blanc-Benon, L. Dallois, M. F. Hamilton, and D. T. Blackstock, ''Propagation of finite amplitude sound through turbulence: Modeling with geometrical acoustics and the parabolic approximation,'' J. Acoust. Soc. Am. **111**, 487–498 (2002).
[15] V. A. Khokhlova, M. V. Averianov, Ph. Blanc-Benon, and R. O. Cleveland, ''Propagation of nonlinear acoustic signals through inhomogeneous moving media,'' Proceedings of the IEEE Ultrasonics Symposium, Montreal, Canada (in press).
[16] B. E. McDonald and W. A. Kuperman, ''Time domain formulation for pulse propagation including nonlinear behavior at a caustic,'' J. Acoust. Soc. Am. **87**, 1473–1481 (1990).
[17] Y. S. Lee and M. F. Hamilton, ''Time-domain modeling of pulsed finite-amplitude sound beams,'' J. Acoust. Soc. Am. **97**, 906–917 (1995).
[18] M. A. Averkiou and M. F. Hamilton, ''Nonlinear distortion of short pulses radiated by plane and focused circular pistons,'' J. Acoust. Soc. Am. **102**, 2539–2548 (1997).
[19] Y. A. Pishchalnikov, O. A. Sapozhnikov, and V. A. Khokhlova, ''A Modification of the Spectral Description of Nonlinear Acoustic Waves with Discontinuities,'' Acoust. Phys. **42**, 362–367 (1996).
[20] T. Christopher and K. J. Parker, ''New approaches to nonlinear diffractive field propagation,'' J. Acoust. Soc. Am. **90**, 488–499 (1991).
[21] J. Tavakkoli, D. Cathignol, R. Souchon, and O. A. Sapozhnikov, ''Modeling of pulsed finite-amplitude focused sound beams in time domain,'' J. Acoust. Soc. Am. **104**, 2061–2072 (1998).
[22] G. Wojcik, J. Mould, Jr, F. Lizzi, N. Abboud., M. Ostromogilsky, and D. Vaughan, ''Nonlinear modeling of therapeutic ultrasound,'' Proceedings of IEEE Ultrasonics Symposium, IEEE Catalog No. 95CH35844, vol. 2, 1617–1621 (1995).
[23] I. M. Hallaj and R. O. Cleveland, ''Single pulse simulation of the pressure and temperature fields in thermoviscus fluids using the FDTD method,'' Acoust. Res. Lett. Online http://ojps.aip.org/arlo published in J. Acoust. Soc. Am. **105**, L7–12 (1999).
[24] C.-C. Shen and P.-C. Li, ''Harmonic leakage and image quality degradation in tissue harmonic imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 728–736 (2001).
[25] R. J. Zemp, J. Tavakkoli, and R. S. Cobbold, ''Modeling of nonlinear ultrasound propagation in tissue from array transducers.'' J. Acoust. Soc. Am. **113**, 139–152 (2003).
[26] A. C. Baker, A. M. Berg, A. Sahin, and J. N. Tjøtta, ''The nonlinear pressure field of plane, rectangular apertures: Experimental and theoretical results,'' J. Acoust. Soc. Am. **97**, 3510–3517 (1995).
[27] M. D. Cahill and A. C. Baker, ''Numerical simulation of the acoustic field of a phased-array medical ultrasound scanner,'' J. Acoust. Soc. Am. **104**, 1274–1283 (1998).
[28] A. Bouakaz, C. T. Lancee, and N. de Jong, ''Harmonic ultrasonic field of medical phased arrays: Simulations and measurements,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 730–735 (2003).
[29] R. O. Cleveland, M. F. Hamilton, and D. T. Blackstock, ''Time-domain modeling of finite-amplitude sound in relaxing fluids,'' J. Acoust. Soc. Am. **99**, 3312–3318 (1996).
[30] W. F. Ames, *Numerical Methods for Partial Differential Equations*, 3rd ed. (Academic, New York, 1992).
[31] M. F. Hamilton, J. Naze Tjøtta, and S. Tjøtta, ''Nonlinear effects in the far field of a directive sound source,'' J. Acoust. Soc. Am. **78**, 202–216 (1985).
[32] T. S. Hart and M. F. Hamilton, ''Nonlinear effects in focused beams,'' J. Acoust. Soc. Am. **84**, 1488–1496 (1988).
[33] J. Naze Tjøtta, S Tjøtta, and E. H. Vefring, ''Effects if focusing on the nonlinear interaction between two collinear finite amplitude sound beams,'' J. Acoust. Soc. Am. **89**, 1017–1027 (1991).
[34] B. Ystad and J. Berntsen, ''Numerical solution of the KZK equation for focusing sources,'' Acta Acustica **3**, 323–330 (1995).
[35] A. Taflove, *Computational Electrodynamics: The Finite-difference Time Domain Method* (Artech, Norwood, MA, 1995), pp. 158–160.
[36] T. Christopher, ''Source pre-biasing for improved second harmonic bubble-response imaging,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 556–563 (1999).
[37] Q.-H. Liu and J. Tao, ''The perfectly matched layer for acoustic waves in absorptive media,'' J. Acoust. Soc. Am. **102**, 2072–2082 (1997).
[38] L. G. Ullate and J. L. San Emeterio, ''A new algorithm to calculate the transient near-field of ultrasonic phased arrays,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 745–753 (1992).
[39] H. Becher, K. Tiemann, C. Pohl, N. C. Nanda, M. A. Averkiou, J. E. Powers, and B. Lüderitz, ''Improvement in the endocardial border delineation using tissue harmonic imaging,'' Echocardiography **15**, 511–517 (1998).
[40] R. J. Graham, W. Gallas, J. S. Gelman, L. Donelan, and R. E. Peverill, ''An assessment of tissue harmonic versus fundamental imaging modes for echocardiographic measurements,'' J. Am. Soc. Echocardiogr. **14**, 1191–1196 (2001).
[41] G. A. Whalley, G. D. Gamble, H. J. Walsh, S. P. Wright, S. Agewall, N. Sharpe, and R. N. Doughty, ''Effect of tissue harmonic imaging and contrast upon between observer and test-retest reproducibility of left ventricular ejection fraction measurement in patients with heart failure,'' Eur. J. Heart Fail. **6**, 85–93 (2004).
[42] K. Kubota, N. Hisa, Y. Ogawa, and S. Yoshida, ''Evaluation of tissue harmonic imaging for breast tumors and axillary lymph nodes,'' Oncol. Rep. **9**, 1335–1338 (2002).
[43] S. Tanaka, O. Oshikawa, T. Sasaki, T. Ioka, and H. Tsukuma ''Evaluation of tissue harmonic imaging for diagnosis of focal liver lesions,'' Ultrasound Med. Biol. **26**, 183–187 (2000).
[44] T. Schmidt, C. Hohl, P. Haage, M. Blaum, D. Honnef, C. Weibeta, G. Staatz, and R. W. Gunther, ''Diagnostic accuracy of phase-inversion tissue harmonic imaging versus fundamental B-mode sonography in the evaluation of focal lesions of the kidney,'' Am J Roentgenol. **180**, 1639–1647 (2003).
[45] J. E. Browne, A. J. Watson, N. M. Gibson, N. J. Dudley, and A. T. Elliott, ''Objective measurements of image quality,'' Ultrasound Med. Biol. **30**, 229–237 (2004).
[46] C. Pislaru, J. D. Hooge, S. V. Pislaru, E. Brandt, R. Cipic, C. E.

Angermann, F. J. van de Werf, B. Bijnes, M. C. Herregods, and G. R. Sutherland, ''Is there a change in myocardial nonlinearity during the cardiac cycle?'' Ultrasound Med. Biol. **27**, 389–398 (2001).
[47] T. L. Szabo, ''Time domain wave equations for lossy media obeying a frequency power law,'' J. Acoust. Soc. Am. **96**, 491–500 (1994).
[48] I. M. Hallaj, R. O. Cleveland, P. Barbone, S. G. Kargl, and R. A. Roy, ''Amplitude degradation of time-reversed pulses in nonlinear absorbing thermoviscous fluids,'' Ultrasonics **38**, 885–889 (2000).

Slow dynamics and anomalous nonlinear fast dynamics in diverse solids

Paul Johnson[a)]
Geophysics Group, University of California, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Alexander Sutin[b)]
Davidson Laboratory, Stevens Institute of Technology, Hoboken, New Jersey 07030

(Received 19 April 2004; revised 27 September 2004; accepted 3 October 2004)

Results are reported of the first systematic study of anomalous nonlinear fast dynamics and slow dynamics in a number of solids. Observations are presented from seven diverse materials showing that anomalous nonlinear fast dynamics (ANFD) and slow dynamics (SD) occur together, significantly expanding the nonlinear mesoscopic elasticity class. The materials include samples of gray iron, alumina ceramic, quartzite, cracked Pyrex, marble, sintered metal, and perovskite ceramic. In addition, it is shown that materials which exhibit ANFD have very similar ratios of amplitude-dependent internal-friction to the resonance-frequency shift with strain amplitude. The ratios range between 0.28 and 0.63, except for cracked Pyrex glass, which exhibits a ratio of 1.1, and the ratio appears to be a material characteristic. The ratio of internal friction to resonance frequency shift as a function of time during SD is time independent, ranging from 0.23 to 0.43 for the materials studied. [DOI: 10.1121/1.1823351]

PACS numbers: 43.25.Dc, 43.25.Ba, 43.25.Ed, 43.25.Gf [MFH] Pages: 124–130

I. INTRODUCTION

The application of an oscillatory-wave or impulsive source at small strains to a solid that manifests nonlinear mesoscopic elastic behavior leads to a modulus and quality factor (inverse dissipation, Q) decrease.[1] Simultaneously, the elastic wave present in the material distorts, manifest by wave speed decrease, and the creation of harmonics and wave modulation.[1] Further, conditioning, an effect of strain memory is induced, manifest, for example, by resonant curve sweeps being different depending on whether the sweep is upgoing or downgoing in frequency.[2,3] These manifestations are *signatures of anomalous nonlinear fast dynamics* (ANFD), also termed *nonclassical nonlinear fast dynamics*.[4] The modulus and dissipation changes do not return to the initial values, the equilibrium state, after wave excitation terminates, but recover over 10^3-10^4 s with a log(time) dependence. This behavior is called *slow dynamics* (SD).[3]

In previous work,[1,5] it was speculated that SD always occurs in tandem with ANFD in nonlinear mesoscopic elastic materials. This has never been demonstrated aside from in a handful of materials. What has been demonstrated is that many materials exhibit ANFD, including sandstones and limestones,[2] damaged solids,[6,7] and concrete.[8] Slow dynamics have been described in sandstone and concrete[9] and in cracked sintered metal.[10] In fact, no systematic studies exist that characterize the connection between ANFD/SD. ANFD/SD has only been described in Berea sandstone,[3] and cracked Pyrex glass[11] to our knowledge.

The purpose of this paper is to report new measurements of ANFD/SD in diverse solids and establish a quantitative connection between ANFD and SD. In Sec. II, we describe SD/ANFD and provide relevant background information. In Secs. III and IV we present the experiment and results, and in Sec. V we discuss the results, and then conclude.

II. NONLINEAR DYNAMICS

A. Nonlinear fast dynamics in classical (atomic elastic) materials

The classical nonlinear theory for ''atomic elasticity'' (anharmonicity) used to describe nonlinear behavior in air, water, single crystals, etc., is thoroughly described in the literature,[12] and we will provide only a brief overview. The classical theory begins with the expansion of the elastic strain energy, E, in powers of the strain tensor, ϵ_{ij}. The expansion coefficients designate the components of the second-order elastic tensor and the third-order elastic tensor, respectively. These tensors are characterized, respectively, by 21 and 56 independent components for an arbitrary anisotropic medium (in the lowest-order, triclinic material symmetry) and by two and, respectively, three components in the highest-order symmetry (isotropic material).

The equation of motion in Lagrangian coordinates is

$$\rho\ddot{u}_i=\frac{\partial\sigma_{ij}}{\partial x_j}, \tag{1}$$

where ρ, σ_{ij}, and $\ddot{u}$ designate the density, the stress tensor, and the particle acceleration, respectively.

A one-dimensional, longitudinal wave (P-wave) propagating in an isotropic medium can exist with only nonzero components $\sigma_{xx}=\sigma$ and $u_x=u$ or $\epsilon_{xx}=\epsilon=\partial u/\partial x$. The corresponding equation of motion can be written as

$$\rho_0\frac{\partial^2\epsilon}{\partial t^2}=\frac{\partial^2\sigma(\epsilon)}{\partial x^2}. \tag{2}$$

a)Electronic address: paj@lanl.gov
b)Electronic address: asutin@stevens-tech.edu

From the energy expansion, the stress-strain relation (also known as the equation of state,) can be written as

$$\sigma = M(\epsilon + \beta\epsilon^2 + \delta\epsilon^3 + \cdots), \tag{3}$$

where M is the elastic modulus, and β and δ are nonlinear coefficients that can be expressed in terms of combinations of the higher order elastic moduli. This equation implies characteristic scaling relations. These are: (a) the amplitude of the second harmonic scales as ϵ^2; (b) the amplitude of the third harmonic scales as ϵ^3; and (c) the change in resonance frequency of the fundamental-mode resonance peak in a Young's mode experiment, for example, scales as ϵ^2.

B. Anomalous nonlinear fast dynamics

The behavior of nonlinear mesoscopic elastic (NME) materials is described in many articles.[1,5] These materials, the prototypical material being rock, behave in unexpected manners. At lower strain amplitudes (order 10^{-6} and lower) it has recently been shown that at least some nonlinear mesoscopic elastic materials (e.g., Berea sandstone) behave ''classically''[13] (obey Landau theory).[12] At higher strain amplitudes (order 10^{-6} and above) there is ample evidence suggesting that classical theory does not describe wave behaviors,[5,1,7] based on (i) observations of the scaling of harmonics and resonant peak shift with strain amplitude and (ii) the presence of conditioning. A mean field theory developed by McCall and Guyer[14] has been broadly applied to both quasistatic and dynamic behavior in these materials, over part of the dynamic strain range.[1,6,7,15] In this region, hysteresis in the equation of state has been appealed to and can be included in the following manner:

$$\sigma = M(\epsilon + \beta\epsilon^2 + \cdots) + \hat{\alpha}[\epsilon, \mathrm{sign}(\dot{\epsilon})], \tag{4}$$

where $\hat{\alpha}$ is a function describing hysteresis in $\sigma - \epsilon$, a ''nonclassical'' effect.[1] This function depends on whether stress is increasing or decreasing and thus the sign term, $\dot{\epsilon} = \partial\epsilon/\partial t$. A specific form of α should follow from the material physics, not yet established.

The above approach does not include conditioning. Conditioning is an induced, continuous state of disequilibrium during dynamic excitation of NME materials. Meaning, that when the sample is excited by an elastic wave, the modulus, and Q decrease immediately; however, if the excitation continues, the modulus and Q continue to decrease for some period of time until the material reaches a new equilibrium state where no more change takes place. Conditioning is considered by many to be the manifestation of slow dynamics during wave excitation, although this has not been quantitatively demonstrated. Existing theories that address conditioning include those described by Scalerandi and Delsanto[4] and Vakhnenko *et al.*[16] The recent model of Vaknenko *et al.* is based on the physics of bond rupture as the underlying mechanism of ANFD/SD. The ramifications of conditioning in ANFD are only beginning to be unraveled experimentally however, and will not be addressed further here, except to say our group is working on separating ANFD and SD at present. This issue does not affect the results described in this paper, but will ultimately affect the accompanying theory.

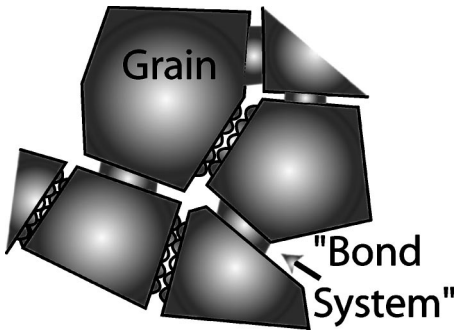


FIG. 1. The hard/soft paradigm of nonclassical materials.

C. Slow dynamics

We currently believe that, in the strain regime where ANFD takes place, SD also appears, manifest by a modulus recovery over $10^3 - 10^4$ s that is linear with $\log(t)$, commencing when a relatively strong wave excitation terminates. Whether the two behaviors appear simultaneously with strain level has not yet been experimentally verified, however, and this is a topic our group is currently working on. SD differs from the well-known phenomenon of creep observed in quasistatic experiments in rock and some metals[17] although they may well be related. SD is a mechanically induced behavior at very small dynamic strain levels, that displays asymmetry with respect to the driving excitation stress. That is, the modulus always displays a quasistatic decrease in response to fast, symmetric oscillatory driving stress. The observed asymmetry is contrary to other physical systems that display proportion to the $\log(t)$ such as creep in rock[18] and metals,[19] magnetic fields, in spin glasses,[20] etc., that respond symmetrically to the driving stress.

The materials that exhibit ANFD and SD have in common a small volume of elastically soft constituents (e.g., bonds in a sintered ceramic) where the SD and ANFD originate, distributed within a rigid matrix (e.g., grains in the ceramic) as is shown in Fig. 1. We refer to this as the ''hard/soft'' paradigm. The soft portion of the material, the ''bond system''[1] is distributed throughout, but within a small fraction of the total volume, less than 1%. In cracked materials the bond system is localized.

III. EXPERIMENT AND DATA ANALYSIS

Our first step was to test a large number of samples (50–100) of a variety of solids to determine whether or not they exhibited ANFD/SD. The samples included numerous rock types, metals, and ceramics. Seven materials seen to exhibit ANFD were selected for quantitative study because they represent the enormous contrasting physical, geometrical, and chemical diversity of the class. These include a pearlite/graphite mixed phase metal (commonly called ''gray iron''), a 99.9% pure alumina ceramic (crucible material), a perovskite ceramic ($LaCoO_3$, used in oxygen separation and solid oxide fuel cells) a feldspathic quartzite, a sintered metal (used in bearings), a sample of Carrera marble (metamorphosed limestone) composed principally of calcite, and Pyrex glass containing a localized crack. Sample shape, dimension and physical characteristics are given in Table I.

TABLE I. Sample information for the seven materials studied. OD and ID refer to the outside and inside diameters, respectively, and L is sample length.

Material	Sample shape	Dimensions (mm^3)
Pearlite/graphite metal	Parallelepiped	100×14×9
Alumina ceramic	Rod	279×19
Quartzite	Rod	117×31.7
Pyrex containing cracks	Rod	284×16
Marble	Rod	101×22
Sintered metal	Ring	OD 41, ID 29, L 25
Perovskite ceramic	Parallellapiped	47.3×6.6×6.6

The experimental configuration used for ANFD and SD measurements is shown in Fig. 2. This setup provides the means to conduct ANFD and SD tests successively for each sample under the same conditions using the same equipment. For ANFD measurements, nonlinear resonant ultrasound spectroscopy (NRUS) was employed.[1,5,6] Using NRUS, the sample was continuously probed using a swept-frequency wave at an eigenmode of the material, applying progressively increasing drive levels. For these measurements we used the resonant ultrasound spectroscopy device manufactured by Dynamic Resonance Systems, Inc. The time-average amplitude for of sweep frequency is recorded and stored on computer. Samples were excited using piezoceramics bonded with epoxy. We used a laser vibrometer manufactured by Polytech for noncontact signal detection. Particle velocity $\dot{u}$ was measured with the Polytech, and $\hat{\epsilon}$, the dynamic strain amplitude, was calculated from this and wavespeed c,[1] $\hat{\epsilon}=\dot{u}/c$.

The linear (low amplitude) resonance frequency f_0 and its corresponding strain $\hat{\epsilon}_0$ were obtained by fitting the resonance peak with a lorentzian. Q, the most commonly applied measure of wave dissipation,[21] was obtained as follows. The Q of a linear elastic system is, $Q_0=2\pi f_0 W/\Delta W$. ΔW is the energy loss per cycle, W is elastic energy stored at maximum stress and strain during harmonic excitation, and f_0 is the mode frequency.[21] This also means that Q_0 is proportional to the wave amplitude at resonance for a given applied force. In a dynamic resonant system, $Q_0=2\pi f_0 W/\delta W=f_0/\delta f$, where δf is the frequency width between half power (3 dB in amplitude) about the resonance peak at f_0 on a power-frequency plot. When the system is linear we can apply a Lorentzian fit to extract Q_0. In a nonlinear system, however, $Q_0=f_0/\delta f$, cannot be used. In fact, definitions of nonlinear dissipation are normally obtained from the ringdown of a freely resonating body.[22–24] Since we have a continuously driven system, we must apply a different approach. In the elastically nonlinear case one can measure the power required to maintain a constant amplitude of oscillation.[22,24]

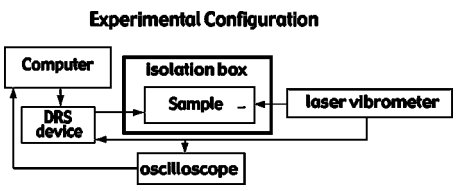


FIG. 2. Experimental configuration for ANFD/SD studies.

This is equivalent to measuring the amplitude of oscillation at constant applied force. In our system, the applied force F of the transducer attached to a solid sample is

$$F=\frac{m\ddot{u}}{2}, \tag{5}$$

where $\ddot{u}$ is acceleration of the transducer and m is its mass (we use no backload on the transducer). The applied force is constant over the frequency range of the experiment because we are operating well below the resonance frequency of the transducer. Further the force is proportional to the applied voltage V:

$$F=\frac{md_{33}\omega^2 V}{2}=k_1 V, \tag{6}$$

where d_{33} is the piezoelectric constant (in units of m/V) and ω^2 appears from calculating displacement u from acceleration $\ddot{u}$, and $k_1=md_{11}\omega^2/2$. The linear relation between force and voltage was verified experimentally.

At the far end of the bar, the detected strain amplitude $\hat{\epsilon}$ is related to the drive force by the amplitude dependent $Q(\hat{\epsilon})$ of the bar

$$\hat{\epsilon}=k_2 F Q(\hat{\epsilon}), \tag{7}$$

where k_2 is the transfer function of the bar.

Substituting Eq. (6) into Eq. (7) yields

$$\hat{\epsilon}=KVQ(\hat{\epsilon}), \tag{8}$$

where $K=k_1k_2$.

An advantage of this method is that k_1 and k_2 do not have to be measured. We can obtain K from the low amplitude measures of strain, voltage and Q_0 [obtained from the half width of the low amplitude (linear) resonance curve]

$$K=\frac{\hat{\epsilon}_0}{V_0 Q_0}. \tag{9}$$

Substitution yields the equation used for estimation of the amplitude dependent attenuation

$$\Delta(1/Q(\hat{\epsilon}))=\frac{1}{Q(\hat{\epsilon})}-\frac{1}{Q_0}=\frac{1}{Q_0}\left(\frac{V\hat{\epsilon}_0}{V_0\hat{\epsilon}}-1\right), \tag{10}$$

which we use to obtain $\Delta(1/Q(\hat{\epsilon}))$ from measurements.

Numerous observations show that the change in $Q(\hat{\epsilon})$ and resonance frequency $f(\hat{\epsilon})$ scales linearly with strain amplitude[1,5,7] above approximately 10^{-6}, in the ANFD regime. In this strain region, we define α_Q as the parameter relating strain and nonlinear energy loss

$$\frac{1}{Q(\hat{\epsilon})}-\frac{1}{Q_0}=\alpha_Q\hat{\epsilon}. \tag{11}$$

Likewise, we define a parameter describing the frequency shift. In the strain range where ANFD is invoked, we have empirically that

$$\frac{f(\hat{\epsilon})-f_0}{f_0}=\frac{\Delta f(\hat{\epsilon})}{f_0}=\alpha_f\hat{\epsilon}, \tag{12}$$

where α_f relates frequency shift and strain amplitude. The P-M space theory predicts the above two relations as well[14]

as does the theory described in Ref. 25. α_f ranges from 10^2 to 10^4 for the materials where it has been measured.[7] We will apply the earlier equations to calculate the characteristic parameters α_f and α_Q from the experimental data.

Recall that SD is the return to equilibrium after large amplitude dynamic excitation. In our experiments, we first conduct a RUS measurement to probe the equilibrium state, then condition the sample at large amplitude, then probe the sample using RUS over a period of time to observe SD.

Specifically, the equilibrium, low amplitude (linear) resonance frequency f_0 of the sample is measured using a standard, low amplitude, swept-frequency resonance using the DRS device before high-amplitude disturbance. The peak is extracted from a lorenzian fit. The sample is then driven at fixed frequency near a resonance and large strain amplitude corresponding to the maximum amplitudes shown in Table II for 2 min to induce material softening. Immediately upon termination of the drive, the RUS measurement recommences at very low strain amplitude (order microstrain) in order to probe the recovery of the resonant peak frequency and Q as a function of elapsed time.

Here $V=V_0$ due to the fact that we probe with low amplitude frequency-sweeps, so Eq. (10) becomes

$$\frac{1}{Q(t)}-\frac{1}{Q_0}=\frac{1}{Q_0}\left(\frac{\hat{\epsilon}_0}{\hat{\epsilon}(t)}-1\right), \tag{13}$$

where $\hat{\epsilon}_0$ is the final, equilibrium strain. Thus it is only necessary to measure the strain amplitude at each frequency peak in the recovery. These measurements are obtained by measuring the peak of a lorenzian fit for each successive resonance curve taken in time.

The effect of temperature is significant and can influence the SD result if one is not extremely careful.[13] Thus precautions were taken to minimize these effects by employing a temperature stable environment using an insulated box constructed from styrofoam.

Materials tested that do not exhibit ANFD/SD include Pyrex glass, polycarbonate, PVC, 5180 steel, tantalum, aluminum, lucite, martinsitic 5180 steel, and pearlitic 5180 steel. A complete list of materials tested in the reconnaissance experiments is available from the authors.

IV. RESULTS

Figure 3 shows raw resonance curves for ANFD/SD obtained from one of the materials (feldspathic quartzite). Figure 3(a) shows amplitude/frequency response of the strain for progressively increasing levels of excitation. Material softening appears as a change in the resonance frequency with drive voltage. Figure 3(b) shows the same data normalized to the amplitude of the applied voltage. This manner of plotting is valuable in simultaneously isolating resonance frequency change and nonlinear dissipation change with strain level. The fact that the amplitude of the normalized peak decreases with increasing drive level and the resonance peak simultaneously broadens indicates the material nonlinear dissipation is increasing with material softening. Figure 3(c) shows the amplitude/frequency response measured at successively increasing times during the recovery process, the SD, for the same material. Recall the drive is fixed at a very small level (assumed linear elastic) during the SD measurements. The recovery is linear with the logarithm of time, as was seen in Berea sandstone.[3]

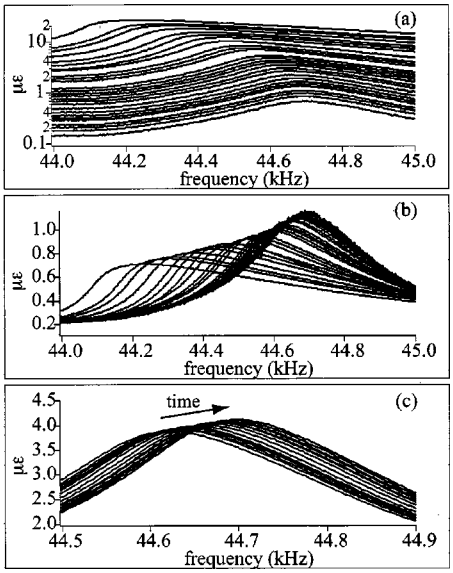


FIG. 3. ANFD and SD observation in the quartzite. (a) shows the resonance curves, (b) shows resonance curves normalized to drive voltage, and (c) shows the SD recovery. Amplitude is in microstrain.

Figures 4(a) and (b) illustrate the change in frequency $(\Delta f/f_0)$ and $Q((1/Q-1/Q_0)$ dependence on strain for all samples obtained from the raw resonance curves. In the low amplitude drive region the dependencies are classically nonlinear [Eq. (3)], and they exhibit nonclassical behavior $[(\Delta f/f_0)$ and $1/Q-1/Q_0$ scale linear with strain] at the higher amplitudes. From the strain region where frequency scales linear with strain amplitude, α_{ANFD} and α_{SD} were extracted.

Table II shows the resonance frequency used, the maximum strain, the material wave speed, and Q_0 for each sample. Table III shows the values of α_f and α_Q extracted from the measurements.

Figures 4(b) and (d) show combined results of frequency and dissipation for SD. In SD, the frequency and $1/Q$ recovers linearly with the log of time.

In order to characterize the relative changes of Q and resonance frequency with strain level in ANFD, we introduce γ_{ANFD}, the ratio of the α parameters: that is, the ratio of the energy loss to the frequency shift ($\gamma_{\mathrm{ANFD}}=\alpha_Q/\alpha_f$). γ_{ANFD} ranges from 0.28 to 1.10 meaning the change in frequency is less than the change in attenuation except in the Pyrex sample. In Berea sandstone, γ_{ANFD} was found to be 0.3.[26]

We note one other observation of interest. We calculated the ratio of the dissipation change to the frequency shift during the SD recovery to see if it had variation between materials. This ratio, termed γ_{SD}, is

$$\gamma_{\mathrm{SD}}=\left(\frac{1}{Q(t)}-\frac{1}{Q_0}\right)\frac{f_0}{\Delta f(t)}. \tag{14}$$

Here $V=V_0$ due to the fact that we probe with low amplitude frequency-sweeps. Substituting using Eq. (13) we get

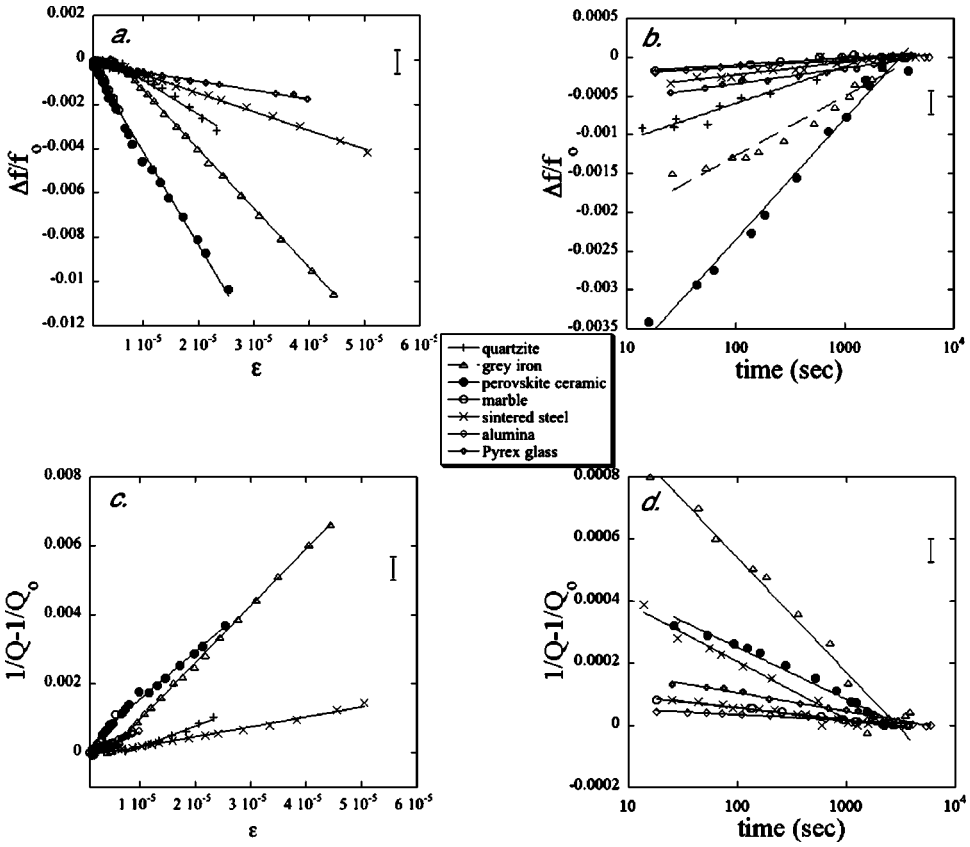


FIG. 4. ANFD and SD in all samples. (a) and (b) show the change in resonance frequency $f-f_0$ normalized to the linear resonance frequency f_0 for ANFD and SD, respectively. (c) and (d) show the behavior in $1/Q$ for ANFD–SD, respectively. The error bars shown are based are extremely conservative: they are obtained from the frequency sampling rate of the experiment, carried through all of the calculations. The lines in the figures are present to guide the eye (they are linear fits in the case of ANFD and logarithmic fits for SD). These data were used to calculate γ_{ANFD} and γ_{SD} in Table III. The high-strain-amplitude data, where $f-f_0$ and $1/Q$ are linear proportional to strain, were used to calculate α in Table III.

$$\gamma_{\mathrm{SD}}=\frac{f}{\Delta f(t)}\left(\frac{1}{Q_0}\frac{\hat{\epsilon}_0}{\hat{\epsilon}(t)}-1\right), \tag{15}$$

where $\hat{\epsilon}_0$ is the final, equilibrium strain. The measured values of this are presented in Table III as well. We find that γ_{SD} are very similar for all seven materials (0.23–0.46). The SD must be due entirely to the relaxation of stored strain assuming the probe wave does not influence the total system strain energy.

V. DISCUSSION

Our primary purpose in this work was to illustrate new observations of ANFD/SD. This is the first systematic study describing ANFD/SD in numerous materials. We find that wherever ANFD are observed, SD are also observed, and vice-versa. We describe results in seven of these materials that represent the huge diversity of the class. We further find that there is a relationship between resonance frequency shift and wave dissipation in ANFD that is similar, but not the same, in each of seven materials. This relation is between 0.28 and 0.66, except for cracked glass which exhibits a ratio of 1.1. The ratios appear to be material dependent. Cracked glass is known to exhibit unusual mechanical behavior in general. We also measure the ratio of dissipation to frequency shift during slow dynamics. This ratio, termed γ_{SD} was between 0.23 and 0.43.

The parameter γ_{ANFD} is predicted by different models. The P-M space[14] model predicts γ_{ANFD} of approximately one, and other hysteretic empirical models such as that described in Ref. 26 leave some flexibility in this value.

In our view, there are currently two credible theoretical approaches that link ANFD/SD. The first is the approach employing a variation of the P-M space model described by Scalerandi and Delsanto.[4] The second is a model described by Vakhnenko *et al.*[16] that treats the elastic nonlinearity of the solid as an ensemble of rupturing bonds.

TABLE II. Sample elastic linear and nonlinear properties for the seven materials studied [Note that Q_0 is the Q of the loaded (source+sample) system]. Wave speed for sintered metal not measured.

Material	f_0 (kHz)	Max strain (10^{-6})	Wave speed (m/s)	Q_0
Pearlite/graphite metal	44.7	44.0	3225	487
Alumina ceramic	9.12	10.0	4940	415
Quartzite	33.6	2.3	3826	167
Pyrex containing cracks	20.0	35.0	4700	662
Marble	86.3	5.6	4030	353
Sintered metal	9.4	5.0		346
Perovskite ceramic	60.0	26.0	1670	140

TABLE III. Values of γ for ANFD and SD in seven materials. α is not shown for the sintered metal part because the strain amplitudes were not directly calculable due to sample geometry, meaning α could not be extracted.

Material	Mean α_Q	Mean α_f	Mean γ_{ANFD}	Mean γ_{SD}
Perlite/graphite metal	487 ± 37	167 ± 25	0.35 ± 0.03	0.25 ± 0.02
Alumina ceramic	95 ± 4	63 ± 14	0.66 ± 0.12	0.26 ± 0.05
Quartzite	145 ± 10	41 ± 7	0.28 ± 0.02	0.36 ± 0.02
Pyrex containing cracks	81 ± 10	84 ± 6	1.10 ± 0.09	0.30 ± 0.02
Marble	370 ± 65	149 ± 45	0.42 ± 0.04	0.43 ± 0.5
Sintered metal			0.31 ± 0.03	0.25 ± 0.03
Perovskite ceramic	240 ± 31	164 ± 11	0.66 ± 0.04	0.23 ± 0.01

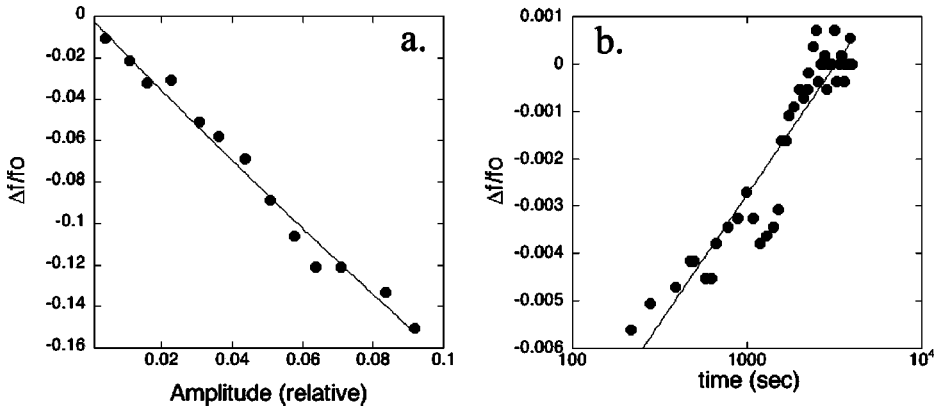


FIG. 5. Model frequency results of ANFD (a) and SD (b).

The reason so few models exist, and in fact that the most widely applied model (P-M space) is phenomenological, is that the physical basis for ANFD/SD is lacking because the origin(s) of these behaviors are unknown. In fact, the physical origins appear to be represented by the spectrum of materials presented in this paper. For instance, dislocations in the metals and glassy dynamics in amorphous materials are likely candidates for the physical origin in these instances; but in cracked glass, sintered metal, or rock, it is not at all clear what could be responsible for the observed behaviors. However, due to the fact that we cannot address the actual physics, we will appeal to the model presented in Refs. 4 and 27 which links ANFD/SD.

The model is based on the local interaction simulation approach[28] applied with a spring model and a version of the Preisach–Mayergoyz (PM) space model of elasticity.[14] In the model the material is described as a sequence of linear elastic elements (the ''hard'' portion—the grain system) and the interstices (the ''soft'' bond system) are described by a large number of hysteretic elastic units that can behave either rigidly or elastically depending on the local pressure. These units correspond to the bond system, localized in the damaged material or distributed in the others. The model includes thermally activated random transitions between different states (rigid or elastic).

This approach is incorporated into PM space where the dynamic wave disturbance is tracked providing the means to predict the dynamic system stress-strain behavior, the ANFD. The SD arise from the fact that the PM space is altered after a wave disturbance, translating to a different bulk material (modulus) state (see Fig. 4 in Scalerandi and Delsanto).[4] The altered material state returns to its initial state with (bulk) log time behavior by thermally activated jumps in state within the hysteretic elastic units from the portion of PM space that has been altered.

From the model, ANFD and SD emerge. Figure 5 shows model results extracted from Scalerandi and Delsanto[4] where the frequency behaviors in ANFD and SD are predicted. They exhibit the same scaling behavior in ANFD as the results shown here, and the log (time) recovery of SD. The general approach appears to capture the properties of ANFD and SD, and could potentially be modified to predict material characteristic behaviors.

We note that, due to the effects of conditioning during ANFD, the rate of conducting a resonance frequency sweep has effects on the shape of the resonance curve.[2] The rate can also affect the resonance peak frequency. We studied the effect by varying the sweep rates and measuring the effect on the calculation of α_f, α_Q, and γ_{ANFD}. We found that the sweep rate can produce variations up to a few percent in these parameters (A quantitative study of sweep-rate dependence is in progress.). We therefore used constant sweep rates in measurements described here. In addition, we conducted the measurements of ANFD at amplitudes much less than those used for the conditioning/SD measurements, where the effects of conditioning are minimized.

VI. CONCLUSIONS

This is the first systematic study showing the relation of ANFD and SD in a large group of solids confirming that, to date, ANFD and SD appear together. We find that, for the seven different materials studied here, γ_{ANFD}, the ratio of internal friction to the resonance frequency shift, has the same order (0.28–0.63). In SD measurements we find that the ratio of internal friction to frequency shift as a function of time during SD, γ_{SD}, has the same order also (0.23–0.43). We do not know the physical basis of ANFD and SD but the fact that they occur together and have the same order of ratio between elastic and anelestic modulus changes point to same physical mechanism for the both effects. Much of our current work is focused on this question.

ACKNOWLEDGMENTS

The work was supported by the US DOE Office of Basic Energy Science. The authors are grateful to invaluable input from S. Habib, M. Scalerandi, P.-P. Delsanto, K. E.-A. Van Den Abeele, D. Pasqualini, K. Heitmann, J. Tencate, R. A. Guyer, and T. Darling.

[1] L. A. Ostrovsky and P. Johnson, ''Dynamic nonlinear elasticity in geomaterials,'' Riv. Nuovo Cimento **24**, 1–46 (2001).
[2] P. A. Johnson, B. Zinszner, and P. N. J. Rasolofosaon, ''Resonance and elastic nonlinear phenomena in rock,'' J. Geophys. Res. **101**, 11553–11564 (1996).
[3] J. TenCate and T. Shankland, ''Slow dynamics in the nonlinear elastic response of Berea sandstone,'' Geophys. Res. Lett. **23**, 3019–3022 (1996).
[4] P.-P. Delsanto and M. Scalerandi, ''Modeling nonclassical nonlinearity, conditioning, and slow dynamics effects in mesoscopic elastic materials,'' Phys. Rev. B **68**, 064107–064116 (2003).
[5] R. A. Guyer and P. Johnson, ''Nonlinear mesoscopic elasticity: Evidence for a new class of materials,'' Phys. Today **52**, 30–36 (1999).

[6] K. E.-A. Van Den Abeele, J. Carmeliet, J. A. Ten Cate, and P. A. Johnson, ''Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, Part II: Single-mode nonlinear resonance acoustic spectroscopy,'' Res. Nondestruct. Eval. **12**, 31–42 (2000).
[7] P. A. Johnson, B. Zinszner, P. Rasolofosaon, K. Van Den Abeele, and F. Cohen-Tenoudji, ''Dynamic measurements of the nonlinear elastic parameter alpha in rock under varying conditions,'' J. Geophys. Res. **109**, 10129–10139 (2004).
[8] J. C. Lacouture, P. Johnson, and F. Cohen-Tenoudji, ''Study of critical behavior in concrete during curing by application of dynamic linear and nonlinear means,'' J. Acoust. Soc. Am. **113**, 1325–1332 (2003).
[9] J. TenCate, D. E. Smith, and R. Guyer, ''Universal slow dynamics in granular solids,'' Phys. Rev. Lett. **85**, 1020–1023 (2000).
[10] P. A. Johnson, A. Sutin, and J. TenCate, ''Development of nonlinear slow dynamical damage diagnostics (S3D) for application to nondestructive evaluation,'' Proceedings of the World Congress on Ultrasonics 2003, pp. 129–132 (http://www.sfa.asso.fr/wcu2003/procs/website/index.html)
[11] V. Zaitsev, B. Castagnede, and V. Gusev, ''Thermoelastic mechanism for logarithmic slow dynamics and memory in elastic wave interactions with individual cracks,'' Phys. Rev. Lett. **90**, 075501–075504 (2003).
[12] L. D. Landau and E. M. Lifschitz, *Theory of Elasticity* (Pergamon, New York, 1986).
[13] J. TenCate, D. Pasqualini, S. Habib, K. Heitmann, P. Johnson, and D. Higdon, ''Nonlinear and nonequilibrium dynamics in geomaterials,'' Phys. Rev. Lett. **93**, 065501–065504 (2004).
[14] K. R. McCall and R. Guyer, ''Equation of state and wave propagation in hysteretic nonlinear elastic materials,'' J. Geophys. Res. **99**, 23887–23897 (1994).
[15] K. E.-A. Van Den Abeele, J. Carmeliet, P. A. Johnson, and B. Zinszner, ''Influence of water saturation on the nonlinear elastic mesoscopic response in Earth materials and the implications to the mechanism of nonlinearity,'' J. Geophys. Res. **107**, 101029–101040 (2002).
[16] O. Vakhnenko, V. Vakhnenko, T. J. Shankland, and J. TenCate, ''Strain-induced kinetics of intergrain defects as the mechanism of slow dynamics in the nonlinear resonant response of humid sandstone bars,'' Phys. Rev. E **70**, 15602–15604 (2004).
[17] U. Hunsche and N. Cristescu, *Time Effects in Rock Mechanics* (Wiley, New York, 1998).
[18] B. I. Pandit and J. C. Savage, ''An experimental test of Lomnitz's theory of internal friction in rocks,'' J. Geophys. Res. **78**, 6097–6099 (1973).
[19] *Creep Behavior of Advanced Materials for the 21st Century*, edited by R. Mishra, A. K. Mukherjee, and K. L. Murty (Minerals, Metals & Materials Society, London, 1999).
[20] K. H. Fischer and J. A. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge, 1991), p. 8.
[21] N. Toksoz and D. Johnston, *Seismic Wave Attenuation* (Geophysics Reprints Series No. 2, Tulsa, 1982).
[22] D. O. Thompson and F. M. Glass, ''Elastic constant-internal friction spectrometer,'' Rev. Sci. Instrum. **29**, 1034–1040 (1958).
[23] A. L. Audenino and P. M. Calderale, ''Measurement of non-linear internal damping in metals: Processing of decay signals in a uniaxial stress field,'' J. Sound Vib. **198**, 395–409 (1996).
[24] R. De Batist, *Internal Friction of Structural Defects in Crystalline Solids* (North Holland, London, 1972).
[25] V. E. Nazarov, A. V. Radostin, L. A. Ostrovsky, and I. A. Soustova, ''Wave processes in media with hysteretic nonlinearity: Part 2,'' Acoust. Phys. **49**, 444–448 (2003).
[26] R. Guyer, J. TenCate, and P. A. Johnson, ''Hysteresis and the dynamic elasticity of consolidated granular materials,'' Phys. Rev. Lett. **82**, 3280–3283 (1999).
[27] M. Nobili and M. Scalerandi, ''Temperature effects on the elastic properties of hysteretic elastic media: Modeling and simulation,'' Phys. Rev. B **69**, 104105–104116 (2004).
[28] R. S. Schechter, H. H. Chaskelis, R. B. Mignogna, and P. P. Delsanto, ''Real-time parallel computation and visualization of ultrasonic pulses in solids,'' Science **265**, 1188–1192 (1994).

Time reversal of ocean noise

Philippe Roux[a] and W. A. Kuperman
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

(Received 9 July 2004; revised 5 September 2004; accepted 29 September 2004)

It has recently been shown [Roux *et al.*, ''Extracting coherent wave fronts from acoustic ambient noise in the ocean,'' J. Acoust. Soc. Am. **116**, 1995–2003 (2004)] that the time-averaged correlation of ocean noise between two points yields a deterministic waveguide arrival structure embedded in the time-domain Green's function. By performing a set of these correlations between a vertical receive array and a single receiver, transfer functions necessary for time reversal can be extracted from ocean noise. Theory and simulation demonstrate this process and data of opportunity from the North Pacific Acoustic Laboratory experiment confirm the expected performance of a noise-based time reversal mirror. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1834616]

PACS numbers: 43.30.Pc, 43.50.Rq, 43.60.Tj [WLS] Pages: 131–136

I. INTRODUCTION

The experimental implementation of an acoustic time reversal mirror (TRM) is now well established both in underwater acoustics[1–3] and in ultrasonics.[4–6] Time reversal is based on reciprocity and time symmetry of the wave equation and the *coherent* recombination of acoustic fields. The extension of the TRM process to *a priori* incoherent noise fields is, therefore, not immediately intuitive. However, since it has recently been shown[7] that the time-averaged correlation function of ocean noise between two points yields deterministic wave fronts related to the structure of the time domain Green's function (TDGF) between those two points, a noisebased TRM would, nevertheless, appear a possibility. Indeed, theory and simulation demonstrate this process[8] and data of opportunity from the North Pacific Acoustic Laboratory experiment[9] confirm the expected performance of a noise-based TRM (NTRM).

The basic configuration of a TRM experiment in the ocean is shown in Fig. 1(a). A probe source S emits a pulse that is received on an array [Fig. 1(b)], time-reversed, and re-emitted. After time reversal, the resulting field focuses back at the probe source position [Fig. 1(c)]. For a noise-based TRM, the noise-extracted TDGF between a surrogate probe source S′ and one element of the array [Fig. 1(d)] is obtained passively via a noise correlation process[10] as indicated in Fig. 1(e). Indeed, the time-averaged cross correlation of ambient noise yields a two-sided symmetric signal that corresponds to an incoming and outgoing wave between S′ and the array [Fig. 1(e)]. As explained in Sec. III of this paper, the outgoing wave is composed of amplitude-shaded noise-extracted TDGFs, the shading being the signature of predominant surface noise sources in the ambient noise field [Fig. 1(d)]. As with a regular TRM, a NTRM experiment then consists of reemitting the noise-extracted TDGFs from the array [Fig. 1(f)]. At this point, there are two options for carrying out the second step of the time-reversal process:

(1) Active time reversal: implement the time reversal process by actually transmitting the time-reversed noise-extracted TDGFs from the array.
(2) Passive time reversal: implement the time reversal process numerically by correlating the noise-based TDGFs obtained from different time-domain ambient noise realizations.

We show in this paper from numerical and experimental data that both active and passive NTRM focus the time-reversed noise correlation function back to the surrogate probe source S′. The potential value of passive NTRM is that it suggests a totally passive (source less) matched field tomography based on the degradation associated with a time changing environment. On the other hand, as shown in Sec. III, because the noise-extracted TDGFs are different from the actual TDGFs, the focus that results from passive NTRM, active NTRM, and classical TRM may differ.

The main advantage of NTRM over TRM relies in the passive way the noise-extracted TDGFs are acquired with a NTRM, for which no probe source is required. However, a NTRM needs a connection between S′ and the array, which may not be practical with distant arrays. Another drawback of NTRM is the necessary time average to observe the convergence of the ambient noise correlation process to the noise-extracted TDGFs. Actually, a NTRM experiment will only be possible when the time scale of the medium fluctuations will be longer than the convergence time of the noise correlation process.[8] This issue is addressed with experimental data in Sec. IV. The paper is structured as follows: In Sec. II, we present the theory of the noise-based TRM. Section III shows results from numerical simulations while Sec. IV confirms the formulation with experimental data.

II. THEORY OF A NOISE-BASED TRM

In the following we show that we obtain a TRM from the noise correlation process. First we write down the classical TRM result; next we review the structure of the output

[a]Electronic mail: proux@ucsd.edu

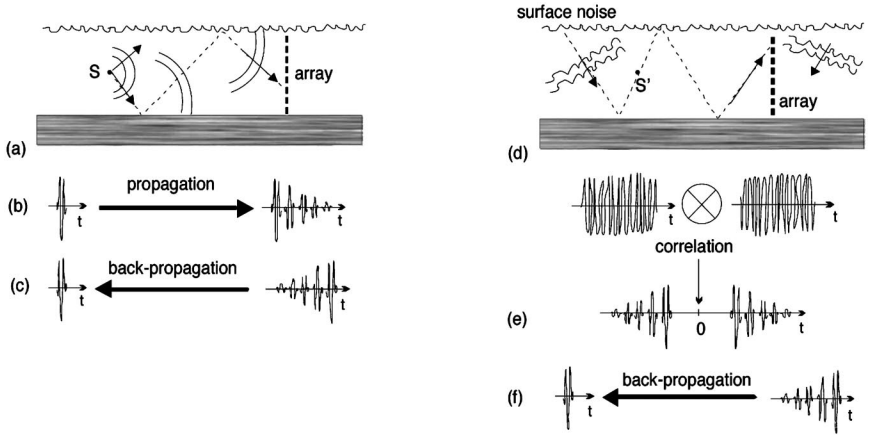


FIG. 1. (a) Classical configuration for a time-reversal mirror (TRM). (b) A probe source S transmits a signal to a source receive array where it is time reversed and retransmitted at each element of the array. (c) The backpropagated field focuses at the source position. (d) Configuration for a noise-based time reversal mirror (NTRM). A noise-extracted TDGF is obtained by retaining the positive time side of the noise correlation function (e) between a surrogate probe source S′ and each element of the array (Ref. 9). Though different in amplitude, this signal is phase-matched to the signal obtained on the array from the probe source in (b). The subsequent backpropagation (f) is the same as in the classical TRM. A variation is to computationally perform (f) but with an independent set of noise-extracted TDGFs taken at another time interval. The degree to which the ocean is temporally stable will be reflected in the quality of the focus.

of the temporal noise correlation. Finally we insert this result into the classically derived TRM result to obtain the noise-based TRM process.

A. Classical TRM

For the geometry shown in Fig. 1, the phase conjugated (frequency analog of time reversal) pressure field P_{pc} from a vertical line of J discrete sources, is classically defined as[11,12]

$$P_{\mathrm{pc}}(r,z,\omega)=\sum_{j=1}^{\mathrm{J}} G_{\omega}(r,z;R,z_j)G_{\omega}^{*}(R,z_j;0,z_{\mathrm{ps}}), \tag{1}$$

where $G_{\omega}^{*}(R,z_j;0,z_{\mathrm{ps}})$ represents the complex-conjugated frequency-dependent Green's function from the probe source at depth z_{ps} to each element of the array at range R and depth z_j. Similarly $G_{\omega}(r,z;R,z_j)$ is the Green's function from each array element back to range r and depth z around the probe source position.

We obtain the time domain version of Eq. (1) by first Fourier synthesizing a pulse at the jth array element,

$$P(R,z_j,t)=\int G_{\omega}(R,z_j;0,z_{\mathrm{ps}})S(\omega)e^{-i\omega t}d\omega, \tag{2}$$

where $S(\omega)$ is the Fourier transform of the probe source pulse. For constant $S(\omega)$, Eq. (2) is proportional to the time domain Green's function, $G_t(R,z_j;0,z_{\mathrm{ps}})$. This expression incorporates all waveguide effects, including time elongation due to multipath propagation. For convenience, we take the time origin such that $P(R,z_j,t)=0$ outside the time interval $(0,\tau)$. The time-reversed signal that will be used to excite the jth transmitting array element is $P(R,z_j,T-t)$ such that $T>\tau$. This condition is imposed by causality. The signal has to be completely received before it can be time reversed. Then, we have

$$\begin{aligned}P(R,z_j,T-t)&=\int G_{\omega}(R,z_j;0,z_{\mathrm{ps}})S(\omega)e^{-i\omega(T-t)}d\omega\\&=\int [G_{\omega}^{*}(R,z_j;0,z_{\mathrm{ps}})e^{i\omega T}S^{*}(\omega)]e^{-i\omega t}d\omega,\end{aligned} \tag{3}$$

where the sign of the integration variable, ω, has been reversed and the conjugate symmetry of the frequency-domain Green's function and the probe source signal have been used. The quantity in square brackets in Eq. (3) is the Fourier transform of the signal received by the jth array element after time reversal. The time delay confirms the equivalence of time reversal and phase conjugation in their respective time and frequency domains.

Noting that the bracketed quantity in Eq. (3) contains the frequency domain representation of the signal retransmitted by the jth array element, Fourier synthesis can be used to obtain the time-domain representation of the field produced by the TRM. From Eq. (1), we have

$$\begin{aligned}P_{\mathrm{TRM}}(r,z,t)=\sum_{j=1}^{J}\int G_{\omega}(r,z;0,z_j)G_{\omega}^{*}(R,z_j;0,z_{\mathrm{ps}})\\\times e^{i\omega T}S^{*}(\omega)e^{-i\omega t}d\omega.\end{aligned} \tag{4}$$

This expression can be used to show that the TRM produces focusing in time as well as in space. Focusing in time occurs because a matched filter occurs. To understand this, examine the TRM field at the focal point [that is, take $r=0$, $z=z_{\mathrm{ps}}$ in Eq. (4)]. The time-domain equivalent of Eq. (4) is

$$\begin{aligned}P_{\mathrm{TRM}}(0,z_{\mathrm{ps}},t)=\frac{1}{(2\pi)^2}\sum_{j=1}^{J}\int\int G_{t''+t'}(0,z_{\mathrm{ps}};R,z_j)\\\times G_{t'}(R,z_j;0,z_{\mathrm{ps}})S(t''+T-t)dt'dt'',\end{aligned} \tag{5}$$

where $G_{t'}(R,z_j;0,z_{\mathrm{ps}})$ represents the TDGF at time t' from the probe source at depth z_{ps} to each array element at range R and depth z_j. Similarly $G_{t''+t'}(0,z_{\mathrm{ps}};R,z_j)$ is the TDGF at

time $t''+t'$ from each array element back to the source at depth z_{ps}. If reciprocity holds in the medium, the TDGF satisfies the condition:

$$G_{t''+t'}(0,z_{ps};R,z_j)=G_{t''+t'}(R,z_j;0,z_{ps}). \quad (6)$$

Using Eq. (6), we observe that the integration over t' in Eq. (5) is an autocorrelation of the Green's function. This operation is matched filtering, with the filter matched to the impulse response from the probe source to the jth array element. The summation is then performed on the J elements of the TRM. Finally, note that the integral over t'' in Eq. (5) is a convolution of each matched-filtered channel impulse response with the time reversed and delayed probe pulse. The signal $S(t''+T-t)$ is the time reversed version of the original probe source signal and the derivation of Eq. (5) uses the causality requirement, $T>\tau$.

B. The TDGF from noise correlation

We omit repeating the well-known normal mode waveguide theory.[13] The frequency domain correlation function of ocean surface generated noise between the probe source and one element of the array is approximated by[10]

$$C_\omega(R,z_{ps},z_j)=\frac{\pi q(\omega)^2}{2\rho^2 k^2}\sum_m \frac{[u_m(z')]^2}{\alpha_m\kappa_m} \times u_m(z_{ps})u_m(z_j)J_0(\kappa_m R). \quad (7)$$

The derivation of Eq. (7) assumes that the modal attenuation coefficients α_m are much smaller than the smallest separation between the real part κ_m of the wave numbers (which is common in underwater acoustics); $u_m(z)$ are the depth dependent mode shape; $q(\omega)$ is the strength of the noise sources, and z' is the depth of the noise layer. Since the Bessel Function $J_0(\kappa_m R)$ is the sum of two Hankel functions of opposite sign, $J_0(\kappa_m R)=\frac{1}{2}(H_0^{(1)}(\kappa_m R)+H_0^{(1)}(-\kappa_m R))$, Eq. (7) can be rewritten as in[9,12]

$$C_\omega(R,z_{ps},z_j)=\frac{\pi q(\omega)^2}{4\rho^2 k^2}\sum_m \frac{[u_m(z')]^2}{\alpha_m\kappa_m}u_m(z_{ps})u_m(z_j) \times[H_0^{(1)}(\kappa_m R)+H_0^{(1)}(-\kappa_m R)] \equiv C_\omega^{(\mathrm{in})}(R,z_{ps},z_j)+C_\omega^{(\mathrm{out})}(R,z_{ps},z_j). \quad (8)$$

Aside from the amplitude shading represented by the fraction in the modal sum of Eq. (8), the frequency domain noise correlation function is composed of the ingoing $C_\omega^{(\mathrm{in})}$ and outgoing $C_\omega^{(\mathrm{out})}$ waveguide Green's function. The shading term $[u_m(z')]^2/\alpha_m\kappa_m$ indicates that the modes amplitude results from near surface, dipole-like excitation of sources distributed over an infinite surface. Since this shading term is real, the structure of the resulting time domain correlation function will have the same temporal structure as the TDGF, but with wave front amplitude determined by the shading. We illustrate this in Fig. 2 where the TDGF in the waveguide is plotted together with the outgoing part of the time domain noise correlation function as computed from the Fourier transform of the frequency domain noise correlation function. Both signals have been weighted by a Gaussian-shaped spectrum between 70 and 130 Hz. The noise correlation function has been computed from a full-field integration code[10,14] in order to further confirm the precise relationship with the TDGF. As expected from Eq. (8), Figs. 2(a) and (b) exhibit the same wave front structure with different amplitude modulation. Figure 2(c) further confirms that the TDGF and the noise correlation function can be superimposed aside from the relative amplitudes of their modal components. Because time reversal is known to be more sensitive to phase than amplitude mismatch,[6] the noise correlation function may be used as a powerful approximation of the TDGF to perform time reversal. Finally, note that the dependence of the noise correlation function versus the noise layer depth z' is studied in detail in Ref. 8.

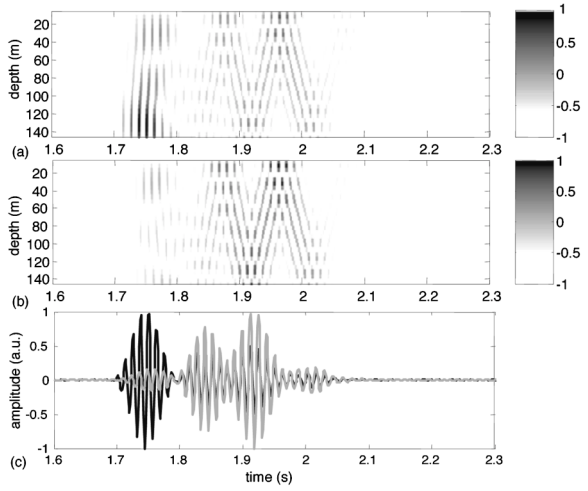


FIG. 2. Spatial-temporal representation of the acoustic field in a 150-m-deep shallow water waveguide. (a) The classical TDGF is computed between 70 and 130 Hz for a 125-m-deep point source as received on a vertical array at range 2500 m. (b) The noise-extracted TDGF for the same geometry and frequency bandwidth showing greater amplitudes of higher modes (i.e., later arrivals) due to the dipole shading of the original noise sources (Ref. 9). The contrast of the gray colorbar has been enhanced to better demonstrate that the temporal structures of both panels are the same. (c) Classical TDGF (black line) and noise-extracted TDGF (gray line) obtained at the source depth. The sound speed profile in the waveguide decreases linearly from 1500 m/s at the surface to 1470 m/s at the bottom. The bottom sound speed, density, and attenuation are 1800 m/s, 1800 kg/m^3, and 0.1 dB/λ, respectively.

C. Noise-based TR

Since there are no factors that would alter the mathematics used in going to the time domain representation of the noise correlation function, we can immediately consider the Fourier transform of Eq. (8) to be the two sided version of the noise-extracted TDGF. Consequently, after selecting the outgoing wave of the time domain noise correlation function, the passive noise-based time reversal analog of Eq. (5) is

$$P_{\mathrm{NTRM}}^{\mathrm{passive}}(0,z_{ps},t)=\frac{1}{(2\pi)^2}\sum_{j=1}^{J}\int\int C_{t''+t'}^{(\mathrm{out})}(R,z_{ps},z_j)C_{t'}^{(\mathrm{out})} \times(R,z_j,z_{ps})S(t''+T-t)dt'dt'', \quad (9)$$

with

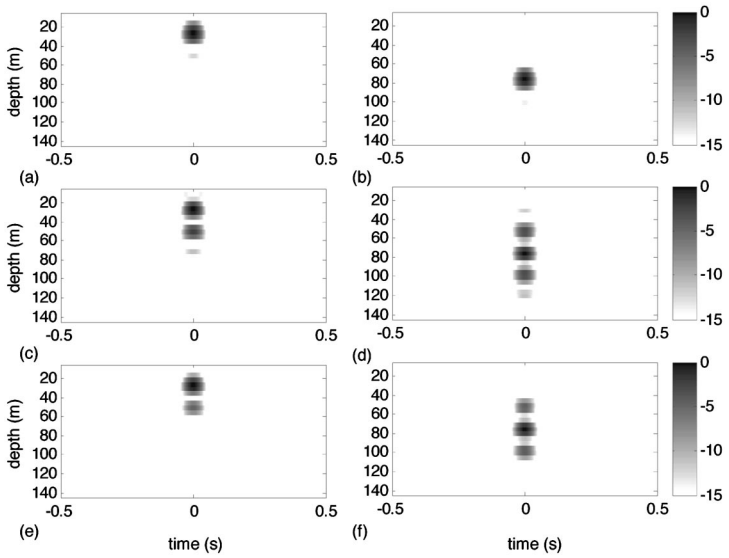


FIG. 3. Spatial-temporal representation of time reversal focal spots computed between 70 and 130 Hz for two different probe source or surrogate probe source positions at depth 25 and 75 m, respectively. (a), (b) Classical TRM; (c), (d) passive NTRM; (e), (f) active NTRM. The latter simulates an actual retransmission for the second step of the TRM process. In each case, time reversal is performed from a 29-element array that covers the whole water column. The waveguide characteristics are the same as in Fig. 2.

$$C_t^{(\mathrm{out})}(R,z_{\mathrm{ps}},z_j)=\int C_\omega^{(\mathrm{out})}(R,z_{\mathrm{ps}},z_j)S(\omega)e^{-i\omega t}d\omega. \tag{10}$$

We retain the source spectrum term in Eq. (4) as a reminder that the noise correlation will cover a finite bandwidth; typically we take the noise spectrum to be constant over this bandwidth. Equation (9) represents the passive NTRM process in which the noise data are used for both the forward and backward components of the TRM process. For active NTRM, in which the time reversal is actually performed by retransmission at the array, we have

$$P_{\mathrm{NTRM}}^{\mathrm{active}}(0,z_{\mathrm{ps}},t)=\frac{1}{(2\pi)^2}\sum_{j=1}^{j}\int\int G_{t''+t'}(0,z_{\mathrm{ps}};R,z_j)\times C_{t'}^{(\mathrm{out})}(R,z_j,z_{\mathrm{ps}})S(t''+T-t)dt'dt''. \tag{11}$$

Comparing Eqs. (5), (9), and (11), we see that TRM is not equivalent to passive/active NTRM because the noise correlation function (which outgoing part is defined as the noise-extracted TDGF) differs from the actual TDGF.

III. SIMULATION OF A NOISE BASED TRM

Figure 3 shows the results of NTRM simulation where we use the correlation of noise between a receiver that acts as a surrogate probe source and an array of receivers that performs the role of a TRM. We show the spatial temporal time reversal focal spots corresponding to both Eqs. (10) (passive NTRM) and (11) (active NTRM) as well as to Eq. (5), i.e., the classical TRM procedure. The simulation is full field[14] so as not to exclude the continuous spectrum part of the waveguide Green's function. The focusing examples are for a surrogate probe source at two different depths and the three cases discussed above: classical TRM, passive and active NTRM.

The main difference between the NTRM focal spots and the classical TRM is the sidelobe structure as shown in greater detail in Fig. 4. The classical TRM has minimum sidelobes and the next best performance is active NTRM. The difference between classical and the passive noise-based TRM arises from the fact that the surface noise modal distribution is weighted toward higher modes (see Fig. 2 or Ref. 7) and therefore cannot be used to fully reconstruct a point source in the diffraction limit. The reason that the active NTRM has better sidelobe performance than the passive NTRM is that the former is closely related to the classical TRM with higher modes being shaded only in the second step of the TRM process rather than having the full, effective two-way shading of the latter.

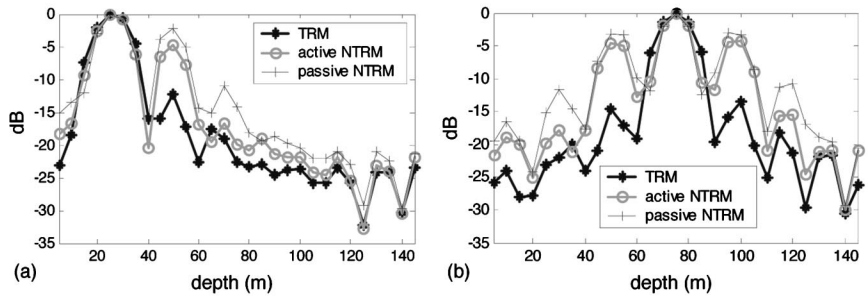


FIG. 4. The depth-dependent focal spot structure for the examples in Fig. 3. (a) and (b) correspond to probe source or surrogate probe source at depth 25 and 75 m, respectively. Classical TRM exhibits lower sidelobes. Active NTRM retransmission has the second best focal spot structure; in this case, the amplitude shading of the noise-extracted TDGFs only impacts the forward propagation. For the passive NTRM, this shading is present in both forward and backward propagations, leading to higher sidelobes.

IV. EXPERIMENTAL DEMONSTRATION OF A NOISE-BASED TRM

The experimental requirement for a NTRM is the ability to perform correlations of noise at distant hydrophones; hence there is a requirement for highly accurate time synchronization between receivers. The North Pacific Acoustic Laboratory experiments[7,9] provide such data to demonstrate the feasibility of implementing a noise-based TRM. We have already shown[7] that even though the NPAL data were basically shipping and not surface-generated noise, the long time correlation process leads to comparable results for both types of data.

Figure 5 is a schematic of the NPAL experiment in the context of a NTRM. With data, we work directly in the time domain so that the inverse Fourier transform of Eq. (8) is simply

$$C_t(R,z_{\mathrm{ps}},z_j)=\int_0^{T_N} N(R,z_j,t')N(0,z_{\mathrm{ps}},t+t')dt'$$

$$\equiv C_t^{(\mathrm{in})}(R,z_{\mathrm{ps}},z_j)+C_t^{(\mathrm{out})}(R,z_{\mathrm{ps}},z_j), \qquad (12)$$

where $N(0,z_{\mathrm{ps}},\mathrm{t})$ and $N(R,z_j,t)$ are the ambient noise received simultaneously at time t on two receivers on different arrays. T_N is the time interval on which ambient noise has been recorded. In the case of the NPAL data, we used time segments $T_N=5$ min. In the case of shipping noise, this time interval appears to be long enough in the studied frequency bandwidth [70–130 Hz] to allow the convergence of the noise correlation function to the noise-extracted TDGF as written in the lower part of Eq. (12). From the two-sided noise correlation function, we retain only the positive time side, $C_t^{(\mathrm{out})}(R,z_{\mathrm{ps}},z_j)$ for the NTRM implementation.

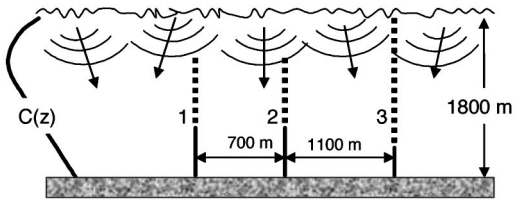


FIG. 5. Schematic of the NPAL experimental setup used for constructing the NTRM process. Array 3 (composed of 40 hydrophones with a 35 m array pitch) is twice as long as arrays 1 and 2 (20 hydrophones with a 35 m array pitch). All the arrays are synchronized permitting accurate cross-correlation. The time-domain noise correlation function between arrays 3 and 2 and arrays 1 and 2 will be used to construct the noise-extracted TDGFs on arrays 1 and 3 with the surrogate probe sources being located on array 2.

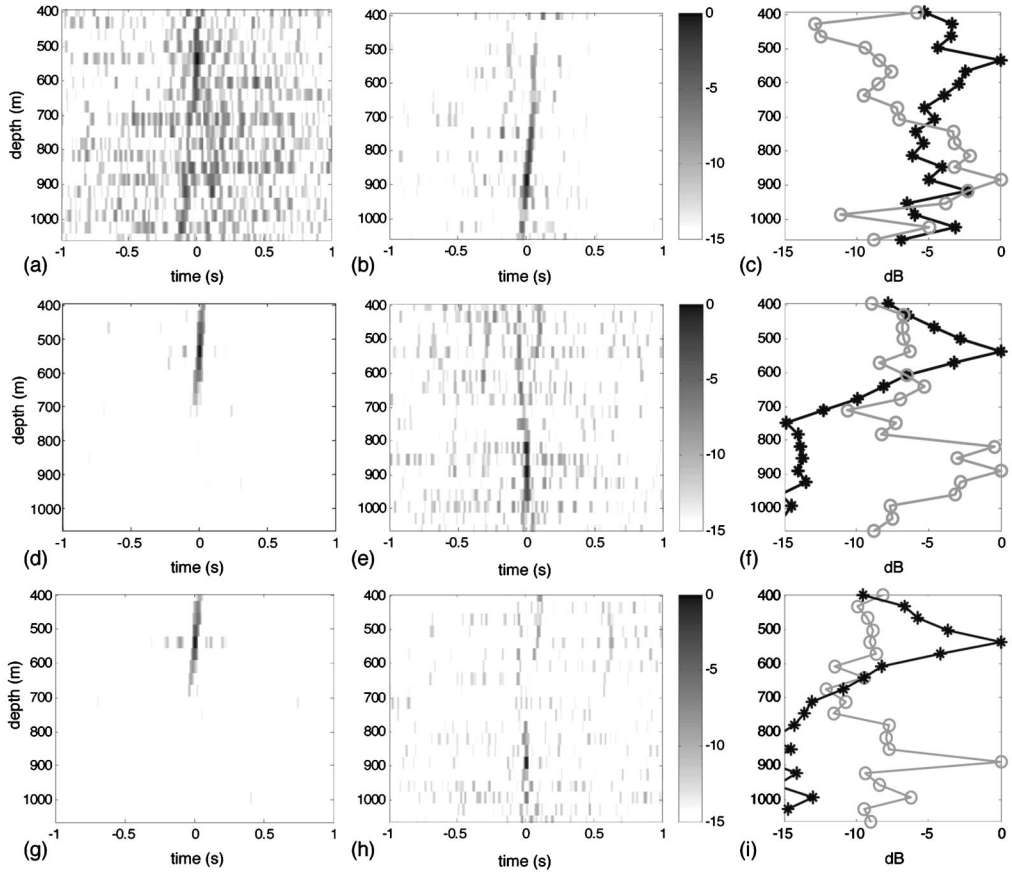


FIG. 6. Spatial temporal representation of the focal spots obtained with passive NTRM from ambient noise data recorded during the NPAL experiment as configured in Fig. 5. We use two different 5 min time intervals for the noise correlation function providing the forward and backward propagating noise-extracted TDGFs. The surrogate probe sources on array 2 are located at depths 536 and 887 m. (a) and (b) Passive NTRM performed from array 1 onto array 2 at the two depths, respectively. (d) and (e) Passive NTRM performed from array 3 onto array 2 at the two depths, respectively. Note that the focus is better because array 3 is larger than array 1. (g) and (h) Passive NTRM performed simultaneously from arrays 1 and 3 onto array 2 at the two depths, respectively. The combined focus at the deeper depth (h) displays a much sharper focus than the single array passive NTRM results of (b) and (e); this is strong evidence that we have constructed a *coherent* NTRM process. (c), (f), (i) The right-hand panels show the depth-dependent focal spot structures at array 2 at depth 536 m (in black), and at depth 887 m (in gray).

Figure 6 shows the spatial temporal focal spot in the case of the passive NTRM. In order to avoid circular reasoning in constructing this result, passive NTRM is performed in two steps. First, $T_N = 5$ min of data are used to obtain the noise-extracted TDGFs by correlating ambient noise between one receiver on array 2 and the other two arrays. In a second step, a different interval $T_N = 5$ min is used to construct the noise-extracted TDGFs that will be used to virtually backpropagate the time-reversed field to the receiver on array 2. The time-reversal results are consistent with what we expect from the classical TRM process. For the larger array refocusing onto array 2 [Figs. 6(d)–(f)] we have a better focus than from the smaller one [Figs. 6(a)–(c)]. Further, the combination of both arrays refocusing at array 2 gives the best focus [Figs. 6(g)–(i)]. For the latter case, the sharper focus observed on the deeper array element can only be explained by a constructive interference between the time reversed fields issued from arrays 1 and 3. This confirms that a TRM-like coherent process occurred from arrays 1 and 3 back to array 2.

V. DISCUSSION

The successful oceanic implementation of NTRM relies on the efficiency of the ambient noise cross-correlation process to extract coherent wave fronts between a surrogate probe source and a hydrophone array. As detailed in Ref. 8, the noise correlation process yields a good estimation of the TDGF when a series of requirements are fulfilled. First, the use of time-synchronized receivers and the absence of strong motions between them (relative to the wavelength) are required. The dynamic range of the hydrophones should be such that low ambient noise can be distinguished from electronic noise. This means also an adequate balance between the medium attenuation and the range separating the pseudosource and the array. As such, the NPAL data provided an excellent example of multiarray noise measurements satisfying these conditions. Second, the choice of the recording time is driven by the need to record a sufficient number of uncorrelated noise events. Thus, hydrophones with large frequency bandwidth work in favor of the noise-correlation process. Naturally, the recording time should also be smaller than the fluctuations time-scale of the environment. Third, the noise correlation process will converge faster toward the noise-extracted TDGF when a homogeneous distribution, in space and time, of uncorrelated noise sources are present. When this condition is not satisfied (ship passing by, burst of sound from marine mammals), temporal or frequency equalization will be necessary to remove high energy single events.

In conclusion, we have shown in this paper by theory and simulation that a noise-based time-reversal mirror is feasible. We have also used data of opportunity to confirm our results. The data confirm that ambient noise from two separate time intervals can be used to construct the required transfer functions for the propagation and backpropagation stages of the time-reversal process. The reason we use the word ''confirm'' and not ''prove'' with respect to the data is that there is a time-averaged correlation requirement. This time interval is dependent on specific environmental and shipping conditions[7] that we do not know for certain in the case of the data presented in this paper. Actually, the convergence of the noise correlation process in general is still very much a research issue.[8,15] However, the multiarray demonstration of a passive NTRM as explained earlier is further compelling evidence of the basic process. Theory and simulation prove the feasibility of NTRM within their respective domains. What still remains to be experimentally demonstrated is active NTRM, in which the noise data on a TRM representing a surrogate probe source can be re-emitted by a set of sources and refocused at the surrogate probe source.

ACKNOWLEDGMENT

This research was supported by the Office of Naval Research.

[1] A. Parvulescu and C. S. Clay, ''Reproducibility of signal transmissions in the ocean,'' Radio Electron. Eng. **29**, 223–228 (1965).
[2] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and Darell Jackson, ''Phase Conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror,'' J. Acoust. Soc. Am. **103**, 25–40 (1998).
[3] S. Kim, G. G. 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).
[4] M. Fink, C. Prada, F. Wu, and D. Cassereau, ''Self focusing with time reversal mirror in inhomogeneous media,'' Proceedings of IEEE Ultrasonics Symposium, 2, Montreal, 1989, pp. 681–686.
[5] M. Fink, ''Time reversed acoustics,'' Phys. Today **50**, 34–40 (1997).
[6] P. Roux and M. Fink, ''Time-reversal in a waveguide: Study of the spatial and temporal focusing,'' J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[7] P. Roux, W. A. Kuperman, and the NPAL Group, ''Extracting coherent wave fronts fom acoustic ambient noise in the ocean,'' J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[8] K. G. Sabra, P. Roux, and W. A. Kuperman, ''Time delays structure of the long-time ambient noise cross-correlation function in an oceanic waveguide,'' J. Acoust. Soc. Am. (in press).
[9] The ATOC Consortium (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. Dzieciuch, M. G. Forbes, C. Hill, B. M. Howe, J. Marshall, D. Menemenlis, J. A. Mercer, K. Metzger, W. Munk, R. C. Spindel, D. Stammer, P. F. Worcester, and C. Wunsch), Science **281**, 1327–1332 (1998). The North Pacific Acoustic Laboratory (NPAL) experiments were designed to study coherence of acoustic signal propagating long distances in the ocean. The acoustic source was 3000 km from the arrays. The NPAL group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. Munk, R. C. Spindel, and P. F. Worcester) provided us with noise data from their receiver array during times when their source was not transmitting. Their array technology is the same used in the Acoustic Thermometry of the Ocean Climate experiments.
[10] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[11] D. R. Jackson and D. R. Dowling, ''Phase conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181 (1991).
[12] D. R. Dowling and D. R. Jackson, ''Narrow-band performance of phase-conjugate arrays in dynamic random media,'' J. Acoust. Soc. Am. **91**, 3257–3277 (1992).
[13] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, Woodbury, NY, 1994).
[14] H. Schmidt and W. A. Kuperman, ''Estimation of surface noise source level from low frequency seismoacoustic ambient noise measurements,'' J. Acoust. Soc. Am. **84**, 2153–2162 (1988).
[15] R. L. Weaver and O. I. Lobkis, ''Elastic wave thermal fluctuations, ultrasonic waveforms by correlation of thermal phonons,'' J. Acoust. Soc. Am. **113**, 2611–2621 (2003).

Compressional and shear wave properties of marine sediments: Comparisons between theory and data

Michael J. Buckingham[a)]
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0238

(Received 7 June 2004; accepted for publication 26 August 2004)

According to a recently developed theory of wave propagation in marine sediments, the dispersion relationships for the phase speed and attenuation of the compressional and the shear wave depend on only three macroscopic physical variables: porosity, grain size, and depth in the sediment. The dispersion relations also involve three (real) parameters, assigned fixed values, representing microscopic processes occurring at grain contacts. The dispersion relationships are compared with extensive data sets, taken from the literature, covering the four wave properties as functions of all three physical variables. With no adjustable parameters available, the theory matches accurately the trends of all the data sets. This agreement extends to the compressional and shear attenuations, in that the theory accurately traces out the *lower bound* to the widely distributed measured attenuations: the theory predicts the *intrinsic* attenuation, arising from the irreversible conversion of wave energy into heat, whereas the measurements return the *effective* attenuation, which includes the intrinsic attenuation plus additional sources of loss such as scattering from shell fragments and other inhomogeneities in the medium. Provided one wave or physical property is known, say the compressional speed or the porosity, all the remaining sediment properties may be reliably estimated from the theory. [DOI: 10.1121/1.1810231]

PACS numbers: 43.30.Ma [AIT] Pages: 137–152

I. INTRODUCTION

Over recent decades, extensive data sets have been published by Hamilton *et al.*[1–8] and Richardson *et al.*[9–12] on the wave properties of surficial, unconsolidated marine sediments. It is evident from the data that an unconsolidated sediment is capable of supporting two types of propagating wave, a compressional (longitudinal) wave and a much slower shear (transverse) wave. Although there have been several attempts to detect a third type of wave, the ''slow'' compressional wave of the Biot theory,[13,14] all have returned a negative result, including the most recent experiment by Simpson *et al.*[15] Based on this evidence, it is tacitly assumed throughout the following discussion that the slow wave in an unconsolidated sediment is negligible if not absent altogether.

It is well established from the published data that the phase speed and the attenuation of both the compressional and shear wave depend, in a more or less systematic way, on the physical properties of the sediment, principally the porosity, the bulk density, the mean grain size, and the depth beneath the seafloor. Indeed, both Hamilton[3,7,16] and Richardson[9,11,12] have developed a set of empirical regression equations, each one of which expresses a wave property (e.g., the compressional phase speed) in terms of a physical property of the medium (e.g., the porosity).

Although regression equations can be satisfactorily fitted to the data, thus providing a useful predictive tool, they give little insight into the physical mechanisms underlying the observed inter-relationships between wave parameters and physical properties. To achieve an understanding of the observed dependencies, it is necessary to turn to a theoretical model of wave propagation in the medium and, in particular, to the dispersion relationships predicted by the model. Such a model has recently been developed by Buckingham[17] on the basis of a specific form of dissipation arising at grain-to-grain contacts. This grain-shearing (G-S) model is intended to represent wave propagation in an unconsolidated granular medium, that is to say, a material in which the mineral grains are in contact but unbonded. By definition, this condition is taken to mean that the mineral matrix has no intrinsic strength or, equivalently, that the elastic (bulk and shear) frame moduli are identically zero.

The absence of an elastic frame in the G-S model contrasts with the starting assumption in Biot's classic theory[13,14] of wave propagation in porous media. Biot treated the medium as though it possessed an elastic mineral frame, an essential assumption in his analysis since the elasticity is the means by which a shear wave is supported in his model. In the G-S model, on the other hand, it is not necessary to postulate the existence of an elastic frame because elastic-type behavior emerges naturally from the analysis as a result of the intergranular interactions themselves: grain-to-grain sliding introduces a degree of rigidity into the medium, which, amongst other effects, automatically leads to the support of a shear wave. The lack of elastic (bulk and shear) frame moduli in the G-S model is consistent with the commonplace observations that an unconfined pile of sand grains shows no resistance to deformation (indicating no restoring force) and that sand grains may be picked off a pile (indicating no tensile strength). Moreover, even though the grains

[a)]Also affiliated with the Institute of Sound and Vibration Research, The University, Southampton SO17 1BJ, England. Electronic mail: mjb@mpl.ucsd.edu

remain unbonded, loose sand in a container shows increasing resistance to penetration (shear strength), which can only arise from intergranular interactions. Such interactions form the essence of the G-S wave-propagation model.[17]

The purpose of this article is to compare the theoretical properties of compressional and shear waves in marine sediments, as expressed in the dispersion relationships of the G-S model, with an extensive set of data culled from the open literature. Most of the data examined here were obtained from *in situ* measurements in silicilastic sediments.

Besides frequency, the G-S dispersion expressions for the phase speed and attenuation depend explicitly on the mean grain size, the porosity, the bulk density (which is strongly correlated with the porosity and hence is not an independent variable), and the overburden pressure (which translates into depth in the sediment). In addition, the G-S dispersion relationships involve three unknown constants which characterize the microscopic processes that occur as adjacent grains slide against one another during the passage of a wave. Once the numerical values of these three constants have been determined, by comparison with three spot-frequency data points, all the functional dependencies of the G-S theory may be evaluated and compared with the data. Thus, the predicted relationships between wave properties (e.g., shear speed) and physical properties (e.g., depth in sediment) may be compared directly with the corresponding data sets that have appeared in the literature. Such comparisons are examined in some detail in this article, along with the dispersion curves (i.e., phase speeds and attenuations versus frequency) predicted by the G-S theory, which are shown to match the available data over wide frequency ranges.

II. THE G-S DISPERSION RELATIONS

According to Buckingham,[17] the compressional-wave speed, c_p, and attenuation, α_p, are given by the expressions

$$c_p = \frac{c_o}{\mathrm{Re}\left[1 + \dfrac{\gamma_p + (4/3)\gamma_s}{\rho_o c_o^2}(j\omega T)^n\right]^{-1/2}}, \tag{1}$$

and

$$\alpha_p = -\frac{\omega}{c_o}\mathrm{Im}\left[1 + \frac{\gamma_p + (4/3)\gamma_s}{\rho_o c_o^2}(j\omega T)^n\right]^{-1/2}, \tag{2}$$

where $j = \sqrt{-1}$. The corresponding expressions for the shear-wave speed, c_s, and attenuation, α_s, are

$$c_s = \sqrt{\frac{\gamma_s}{\rho_o}}\frac{(\omega T)^{n/2}}{\cos\left(\dfrac{n\pi}{4}\right)}, \tag{3}$$

and

$$\alpha_s = \omega\sqrt{\frac{\rho_o}{\gamma_s}}(\omega T)^{-n/2}\sin\left(\frac{n\pi}{4}\right). \tag{4}$$

Between them, Eqs. (1)–(4) constitute the dispersion relationships predicted by the G-S theory. Both dispersion pairs are causal, satisfying the Kramers–Kronig relationships. Several familiar parameters appear in the G-S dispersion expressions: the angular frequency, ω, the bulk density of the medium, ρ_o, the sound speed in the absence of grain-to-grain interactions, c_o, and an arbitrary time $T = 1$ s, introduced solely to avoid awkward dimensions that would otherwise arise when the frequency is raised to a fractional power.

Less familiar are the three remaining parameters, γ_p, γ_s, and the (positive, fractional) index n, which between them represent the effects of grain-to-grain interactions on the wave speeds and attenuations. From the way that they appear in the dispersion relations, and the fact that they have dimensions of pressure, it is evident that the two (real) parameters γ_p and γ_s are compressional and shear moduli, respectively, providing a measure of the normal and tangential stresses associated with intergranular sliding. In fact, γ_p and γ_s are closely analogous to the Lamé parameters of elasticity theory. There is no such analogy, however, for the dimensionless index n, which is a measure of the degree of strain hardening that is postulated to occur at intergranular contacts as grain-to-grain sliding progresses. Details of the intergranular sliding and strain-hardening mechanisms may be found in Buckingham.[17]

If n were zero, the compressional and shear attenuations would both vanish and the expressions for the two waves speeds would be independent of frequency. Of course, dissipation is never completely absent, so n is always finite, but it is small compared with unity, taking a value close to 0.1 for a typical sand sediment. With such a low value for n, the two G-S expressions for the wave speeds exhibit logarithmic dispersion, at levels of the order of 1% (compressional) and 10% (shear) per decade of frequency, and the associated attenuations both scale essentially as the first power of frequency, f. These simple frequency-dependencies derive directly from straightforward approximations[17] for the exact expressions in Eqs. (1)–(4).

From the point of view of wave propagation, two important physical properties of sediments are the porosity, N, and the mean grain diameter, u_g. In addition, the intergranular interactions, and hence also the wave properties, are sensitive to the overburden pressure, which scales with the depth, d, in the sediment. All three parameters appear in the G-S dispersion relations, the porosity through c_o and ρ_o, while the grain size and the depth in the sediment, both raised to fractional powers, appear in the expressions given below for γ_p and γ_s. These two moduli also show a weak but significant dependence on the porosity, N.

It is well known that the bulk density, ρ_o, may be expressed in terms of the porosity as a weighted mean of the density of the pore water, ρ_w and the density of the mineral grains, ρ_g

$$\rho_o = N\rho_w + (1-N)\rho_g. \tag{5}$$

Similarly, the bulk modulus of the medium, κ_o, may be expressed in terms of the porosity as a weighted mean

$$\frac{1}{\kappa_o} = N\frac{1}{\kappa_w} + (1-N)\frac{1}{\kappa_g}, \tag{6}$$

where κ_w and κ_g are the bulk moduli of the pore water and the mineral grains, respectively. If the sediment were a simple suspension in which grain-to-grain interactions were absent, the speed of sound would be c_o, which depends on the porosity through Wood's equation[18]

$$c_o = \sqrt{\frac{\kappa_o}{\rho_o}}. \tag{7}$$

In the limit of low frequency, the G-S expression for the compressional wave speed in Eq. (1) reduces to c_o, while Eqs. (2)–(4) show that α_p, c_s, and α_s all approach zero. Thus, according to the G-S dispersion relations, in the low-frequency limit, the effects of grain-to-grain interactions are negligible and the medium acts as a simple suspension.

Turning to the compressional and shear moduli, γ_p and γ_s, their dependencies on grain size, depth in the sediment, and porosity are established by treating the mineral grains as elastic spheres, which deform slightly under the influence of the overburden pressure. At the point where two grains touch, a small, tangential circle of contact is formed, the radius of which is given by the Hertz theory[19] of elastic bodies in contact. Assuming that the two grains are identical spheres of diameter u_g, Young's modulus E_g and Poisson's ratio θ_g, the radius, a, of the circle of contact is

$$a = \sqrt[3]{\frac{3}{8} F \frac{(1-\theta_g^2)}{E_g} u_g}, \tag{8}$$

where F is the force pressing the spheres together. At depth d in the sediment, the force F scales with the overburden pressure, P, which is defined[20] as the excess pressure arising from the difference between the bulk density of the sediment and the density of the pore water

$$P = (\rho_o - \rho_w) g d = (1-N)(\rho_g - \rho_w) g d, \tag{9}$$

where g is the acceleration due to gravity. The expression on the right of Eq. (9) derives from Eq. (5) for the bulk density, which accounts for the dependence of the overburden pressure on the porosity. It follows that the radius of the circle of contact may be expressed as

$$a \propto [(1-N) d u_g]^{1/3}, \tag{10}$$

where the constant of proportionality involves the elastic properties of the mineral grains but is independent of the bulk properties of the two-phase medium.

Now, the moduli γ_p and γ_s scale with the rates at which sliding events occur within the circle of contact of radius a. With many micro-asperities distributed randomly over each face of the contact area, the rate of sliding will scale with the number of asperities available to slip against one another. In the case of the compressional modulus, most of the sliding takes place around the perimeter of the circle of contact, where the normal pressure is a minimum. Thus, the number of asperities involved in the sliding process is expected to scale with the perimeter of the circle of contact, and hence γ_p is proportional to a. Shearing, on the other hand, involves a slip of the two flat faces of the surface of contact against each other, in which case the number of available asperities scales with area of the circle of contact, and therefore γ_s is proportional to a^2. According to these arguments, the two moduli may be expressed in terms of the porosity, grain size, and depth in the sediment as follows:

$$\gamma_p = \gamma_{po} \left[\frac{(1-N) u_g d}{(1-N_o) u_{go} d_o} \right]^{1/3}, \tag{11}$$

and

$$\gamma_s = \gamma_{so} \left[\frac{(1-N) u_g d}{(1-N_o) u_{go} d_o} \right]^{2/3}, \tag{12}$$

where the terms in brackets have been constructed to be dimensionless, with the factors γ_{po} and γ_{so} taking numerical values that are independent of the porosity, grain size, and depth in the sediment. By introducing the three reference parameters, porosity N_o, grain size u_{go}, and depth d_o, into the denominators of Eqs. (11) and (12), awkward dimensions are avoided when the terms in square brackets are raised to a fractional power. It should be clear that the values of these reference parameters may be chosen for convenience; they do not represent additional unknowns. To distinguish them from the compressional and shear moduli, the parameters γ_{po} and γ_{so}, respectively, will be referred to as compressional and shear coefficients. Obviously, γ_{po} and γ_{so} are the values of γ_p and γ_s when $N = N_o$, $u_g = u_{go}$, and $d = d_o$.

In making the comparisons between the G-S theory and data, the values of the compressional and shear coefficients, as well as the strain-hardening index, n, are held fixed for all sediments, from coarse sands to the finest silts and clays. In principle, however, γ_{po}, γ_{so}, and n could depend weakly on the elastic properties of the mineral grains and also on such factors as the micro-roughness of the areas of contact. Thus, in practice, the values of these three parameters could differ slightly from one sediment to another but, judging from the good quality of the match between theory and data, as demonstrated below, it would seem that such variations are very minor.

The functional dependence of the G-S dispersion relations on the macroscopic, geoacoustic parameters of a sediment are given explicitly and completely by the algebraic expressions in Eqs. (1) to (12). From a casual inspection, it is evident that the selection of geoacoustic parameters appearing in the G-S theory is quite different from that in Biot's theory.[13,14] According to Table 1.1 in Stoll,[21] the Biot theory involves the permeability, the viscosity of the pore fluid, the pore-size parameter, the structure factor, the (complex) shear modulus of the skeletal frame, and the (complex) bulk modulus of the skeletal frame, a total of eight parameters, all of which are absent from the G-S formulation of the dispersion relations. Instead, the G-S theory includes the three (real) parameters representing microscopic processes occurring at grain contacts, γ_{po}, γ_{so}, and n. Through the compressional and shear moduli [Eqs. (11) and (12)], the G-S theory also includes the grain size and the depth in the sediment, neither of which appears explicitly in Biot's theory. The porosity, the bulk density, and the physical properties (densities and bulk moduli) of the individual constituents of the biphasic medium are common to both theories.

According to Eq. (5), the bulk density and the porosity are linearly related and hence are fully correlated. In fact,

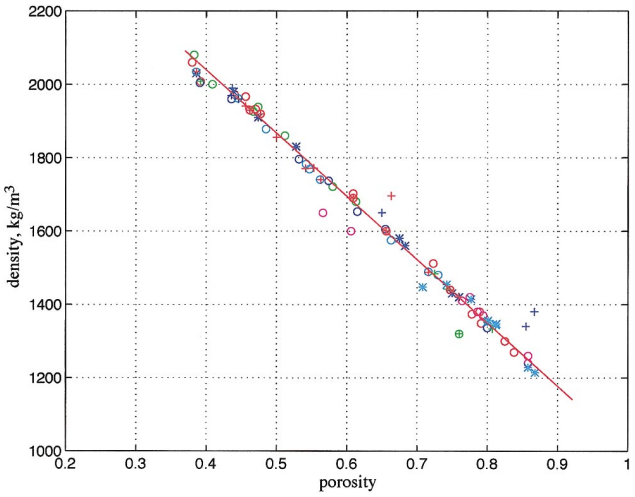


FIG. 1. Compilation of density versus porosity data, mostly from Hamilton. The red line is from Eq. (5) using the values of ρ_w and ρ_g listed in Table II. (See Table I for the key to the symbols in this and all subsequent figures.)

Richardson *et al.*[10–12,22] have used this relationship to compute the bulk density from laboratory measurements of the porosity. Hamilton,[3,7,8] on the other hand, measured the density directly from weight and volume, rather than computing it by substituting his independently measured porosity into Eq. (5). Hamilton's data may therefore be used as a test of Eq. (5).

Figure 1 shows Hamilton's published measurements of density and porosity for many types of marine sediment, ranging from very fine silts and clays to coarse sands. (Table I is the key to all the symbols, representing published data collected from the literature, which appear in Figs. 1 to 11). Also shown in Fig. 1 is a plot of the linear relationship between density and porosity in Eq. (5), evaluated with $\rho_w = 1005\,\mathrm{kg/m^3}$ and $\rho_g = 2730\,\mathrm{kg/m^3}$. Most of the data points fall on or very close to the line, indicating that the correlation represented by Eq. (5) holds accurately in practice, making it unnecessary to examine the functional dependence of the wave properties on both the porosity and the bulk density. Porosity alone is considered below.

III. EVALUATION OF γ_{po}, γ_{so}, AND n

As already discussed, in addition to the macroscopic variables (porosity, grain size, and depth in the sediment), the G-S dispersion relationships [Eqs. (1)–(4)] involve the three

TABLE I. Key to the data symbols in Figs. 1–11.

Symbol	Color	Source
○	Red	Hamilton (Ref. 5) (1972), Table 1
○	Green	Hamilton (Ref. 5) (1972), Table 2
○	Blue	Hamilton (Ref. 5) (1972), Table 3
○	Magenta	Hamilton (Ref. 3) (1970), Table 2
○	Cyan	Hamilton (Ref. 8) (1987), Tables 1 & 2
*	Blue	Hamilton (Ref. 3) (1970), Table 1
*	Cyan	Hamilton (Ref. 8) (1987), Tables 3 & 4
+	Red	Hamilton (Ref. 6) (1980), Table 1
+	Green	Hamilton *et al.* (Ref. 2) (1970), Table 2
+	Blue	Hamilton (Ref. 43) (1963), Table 1
⊕	Red	Hamilton (Ref. 44) (1956), Tables 1a & 1b
⊕	Green	Wood & Weston (Ref. 45) (1964),
○	Black	Buckingham & Richardson (Ref. 23) (2002), Tables 1a & 1b
+	Black	Richardson (Ref. 12) (2002), Fig. 1 & Richardson (Ref. 33) (2000) Fig. 2
⊕—⊕	Black	Richardson *et al.* (Ref. 25) (2001), Fig. 15
—	Black	Richardson (Ref. 10) (1991), Table 1 & Fig. 3
+	Cyan	Richardson (Ref. 46) (1986), Table 1
*	Magenta	Richardson *et al.* (Ref. 9) (1991), Tables 43.1 and 44.2
×	Red	Richardson & Briggs (Ref. 22) (1996), Figs. 3, 5, 6
×	Blue	Richardson (Ref. 11) (1997), Table 1, La Spezia & Adriatic Sea
×	Green	Richardson (Ref. 11) (1997), Table 1, Panama City & Boca Raton
×	Magenta	Richardson (Ref. 11) (1997), Table 1, Eckernfoerde Bay, Germany
×	Cyan	Richardson (Ref. 11) (1997), Table 1, Eel River & Orcas Bay
⊗	Green	Brunson & Johnson (Ref. 47) (1980), Table II
⊕	Magenta	Simpson *et al.* (Ref. 15) (2003), Fig. 6
⊗	Blue	Brunson (Ref. 48) (1991), Tables 1 & 2
*	Black	Richardson (Ref. 10) (1991), Table 1 & Fig. 3
+	Magenta	Richardson & Briggs (Ref. 49) (1993), Table 1
×	Black	McLeroy & DeLoach (Ref. 50) (1968), Table II
*	Red	McCann & McCann (Ref. 51) (1969), Table 3
*	Green	Richardson (Ref. 12) (2002), Fig. 3

TABLE II. Parameters appearing in the G-S dispersion relations.

Material parameter	Symbol	Value	Comments
Density of pore fluid (kg/m^3)	ρ_f	1005	Evaluated from density versus porosity fit to data, Fig. 1
Bulk modulus of pore fluid (Pa)	K_f	2.374×10^9	From Kaye & Laby (Ref. 52), p. 35
Density of grains (kg/m^3)	ρ_s	2730	Evaluated from density porosity fit to data, Fig. 1
Bulk modulus of grains (Pa)	K_t	3.6×10^{10}	From Richardson *et al.* (Ref. 53)
Compressional coefficient (Pa)	γ_{po}	3.888×10^8	Evaluated from spot-frequency compressional and shear wave data
Shear coefficient (Pa)	γ_{so}	4.588×10^7	Evaluated from spot-frequency compressional and shear wave data
Strain-hardening index	n	0.0851	Evaluated from spot-frequency compressional wave data
rms grain roughness (μm)	Δ	1	Evaluated from porosity versus grain size distribution, Fig. 7(b)
Reference grain diameter (μm)	u_o	1000	Arbitrary choice
Reference depth in sediment (m)	d_o	0.3	Arbitrary choice
Reference porosity	N_o	0.377	Arbitrary choice
Porosity	N	Variable	
Mean grain diameter (μm)	u_g	Variable	
Depth in sediment (m)	d	Variable	

unknown constants, γ_{po}, γ_{so}, and n, which characterize the microscopic interactions occurring at grain boundaries during the passage of compressional and shear waves. At present, no theory exists from which these constants could be computed, so instead they are evaluated here from spot-frequency measurements of three wave properties: the compressional speed, the compressional attenuation, and the shear speed.

The two compressional-wave observations are taken from Table 1a in Buckingham and Richardson[23] for the SAX99 medium-sand sediment [$u_g=414.7\,\mu$m ($\phi=1.27$), $N=0.377$]: at depth $d=0.3$ m and a frequency of 100 kHz, the measured sound speed (cf) and attenuation (αf) are 1787 m/s and 30.93 dB/m, respectively. The shear-speed measurement, taken from Fig. 1 in Richardson,[12] is for a well-sorted, fine-sand sediment [$u_g=238.16\,\mu$m ($\phi=2.07$), $N=0.358$] at a North Sea site designated C1: at depth $d=0.28$ m and a frequency of 1 kHz, the (average) shear speed is 131.4 m/s.

When the measurements cited above are substituted into the compressional and shear dispersion relations [Eqs. (1)–(4)], some elementary algebra yields the values of γ_{po}, γ_{so}, and n listed in Table II. In the following comparisons of the G-S theory with numerous data sets, these values are held fixed for *all* sediments. Under this constraint, it is of some interest to examine how well the G-S dispersion relationships represent the wave properties in sediments ranging from coarse sands to very fine clays. This is achieved by considering the dependencies of the wave speeds and attenuations on porosity and grain size. The depth dependence of the wave properties is also compared with data from various types of sediment. As will become evident, the fixed values of γ_{po}, γ_{so}, and n cited in Table II lead to theoretical wave-property dependencies on porosity, grain size, and depth that align with essentially all the data. Before discussing the physical properties, however, the theoretical dependencies of the wave speeds and attenuations on frequency are compared with observations.

As mentioned earlier, most of the data used in the following comparative discussions are from *in situ* measurements made in silicilastic (quartz sand) sediments. Carbonate sediments are excluded, since they differ from silicilastic materials in several ways,[24] including having open, hollow grains, which, with all else equal, gives rise to higher porosities. Hamilton and Bachman[7] found that ''hollow tests (shells) of Foraminifera act as solid particles in transmitting sound,'' implying that the intraparticle porosity had little if any effect on the wave properties. A preliminary comparison between wave data from carbonate sediments and the G-S theory, evaluated using the interparticle (rather than total) porosity, appears to support Hamilton and Bachman's conclusion but requires more detailed examination before being discussed further.

IV. FREQUENCY DEPENDENCE

A. Compressional wave

Figure 2 shows *in situ* measurements of the frequency dependence, from 25 to 100 kHz, of the compressional wave speed and attenuation at the SAX99, medium-sand experiment site off Fort Walton Beach in the northern Gulf of Mexico.[25] The smooth curves, representing the G-S theory, were evaluated from Eqs. (1) and (2). These theoretical curves pass precisely through the higher of each pair of data points at 100 kHz as expected, since both data points were used in evaluating the compressional and shear coefficients γ_{po}, γ_{so} as well as the strain-hardening index, n.

The sound-speed data from the SAX99 site [Fig. 2(a)] exhibit weak logarithmic dispersion at a level of approximately 1% per decade of frequency. A similar logarithmic gradient is predicted by the G-S theory. Although there is some scatter in the data (of the order of ±10 m/s), it is clear that the measured sound speeds cluster around the theoretical line. The observed SAX99 dispersion is consistent with recently reported *in situ* measurements by Simpson *et al.*,[15]

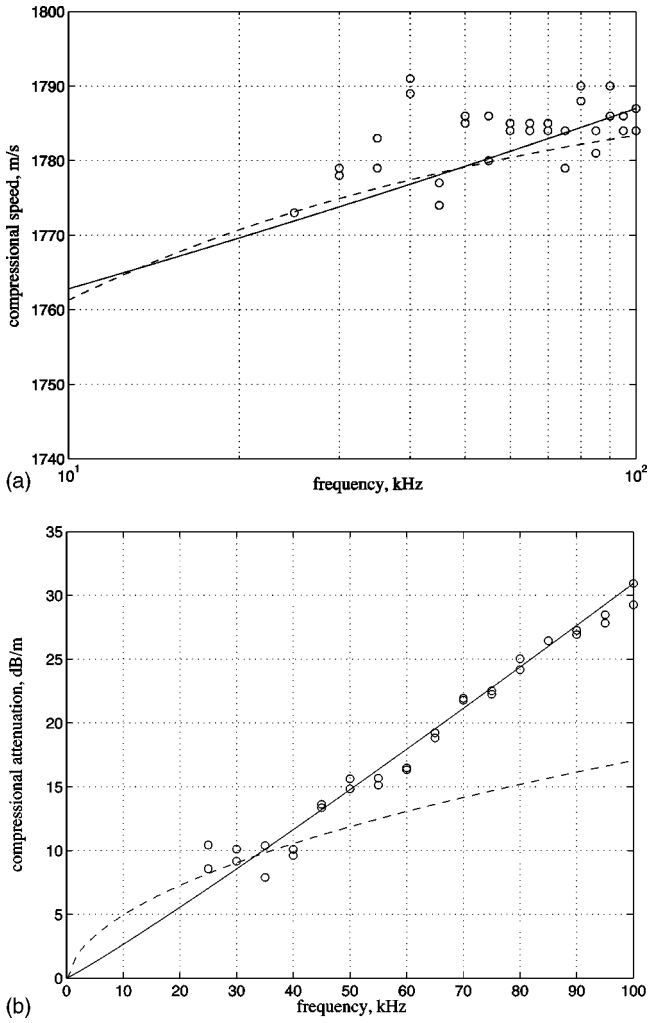


FIG. 2. Measured frequency dependence of the compressional (a) speed (semilogarithmic axes) and (b) attenuation (linear axes) for the medium-sand SAX99 site. The smooth, solid curves represent the predictions of the G-S theory, computed from Eqs. (1) and (2) using the reported values (Ref. 25) of the porosity, grain size, and depth [$N=0.377$, $u_g=414.7\ \mu$m ($\phi=1.27$), $d=0.3$ m]. The dashed curves are from William's EDF approximation (Ref. 29) to the Biot theory, with $\kappa=2.5\cdot10^{-11}$ m^2 and $A=1.35$.

which were made in a medium-sand sediment at St. Andrews Bay, Florida. Over the frequency band from 3 to 80 kHz, they also observed logarithmic dispersion in the region of 1% per decade. Such a weak gradient is quite difficult to detect *in situ*, and earlier investigators not uncommonly failed to observe any variation of sound speed with frequency. Hamilton,[4] for example, reported that ''there is negligible dependence or no dependence of velocity on frequency,'' on the basis of which he developed his empirical elastic model of wave propagation in sediments.

Visual inspection of the data in Fig. 2(b) reveals that, between 25 and 100 kHz, the compressional attenuation from the SAX99 site scales almost exactly as the first power of frequency. In fact, a straight line may be drawn through the data points which, when extrapolated, passes precisely through the origin. However, it has long been known that a pure linear scaling of attenuation with frequency is unphysical in that it violates causality.[26–28] But, no such difficulty arises with the G-S theory, which yields an attenuation that is almost but not quite linear in frequency, as exemplified in Fig. 2(b). This near-linear curve, which rises marginally faster than the first power of frequency, passes through the origin and accurately through the data points.

The question of whether the attenuation in marine sediments varies (nearly) linearly with frequency has been a controversial issue for many years. Hamilton[8] consistently maintained, on the basis of extensive measurements made by himself and others, that the attenuation exhibits a linear scaling with frequency. Stoll,[21] on the other hand, has long argued on the basis of the Biot theory[13,14] that the attenuation deviates significantly from a first-power dependence on frequency, varying as the square root of frequency over the bandwidth of the data shown in Fig. 2(b).

Biot's dispersion curves, evaluated from Williams'[29] ''effective density fluid'' (EDF) approximation to Biot's full theory, are included as the dashed lines in Fig. 2. To compute the EDF sound speed and attenuation curves, the values of two physical parameters peculiar to the Biot theory, the permeability, $\kappa=2.5\times10^{-11}$ m^2 and tortuosity $A=1.35$, were taken from Table 1 in Williams *et al.*,[30] which lists properties of the SAX99 sediment. In Fig. 2(a), it can be seen that the EDF sound speed shows weak dispersion, aligning reasonably with both the data and the G-S curve, although at lower frequencies than those shown the two theoretical predictions diverge. The EDF attenuation in Fig. 2(b), however, fails to match the data over most of the frequency range. Whereas the attenuation data scale essentially linearly with frequency, the EDF attenuation curve exhibits the classic, high-frequency $f^{1/2}$ dependence of a viscous fluid.

Since viscous dissipation of the pore fluid is the only loss mechanism in the Biot theory,[13,14] it is not surprising that the frequency dependence of the Biot attenuation is exactly the same as that of a true viscous fluid.[31] Although the details of Biot's attenuation curve may vary with material parameters such as permeability and tortuosity, the functional form is always the same: at low frequencies the attenuation scales as the square of frequency, transitioning at about 1 kHz (depending on permeability and viscosity) to a square-root scaling with frequency. The discrepancy, illustrated in Fig. 2(b), between such behavior and the near-linear scaling of the attenuation data with frequency is a strong indicator that the viscous dissipation mechanism of the Biot theory may be insignificant in marine sediments.

Evidence to support such a conclusion is found not only in the SAX99 attenuation data. From 3 to 80 kHz, Simpson *et al.*[15] have observed a (near) linear dependence of attenuation on frequency (their Fig. 6) in a medium-sand sediment in St. Andrews Bay, Florida; and earlier laboratory experiments in fine sand by Simpson *et al.*[32] returned a similar (near) linear scaling of the attenuation with frequency between 4 and 100 kHz (their Fig. 7). As illustrated in Fig. 2(b), Biot's theory cannot account for a linear scaling of attenuation with frequency; but, such behavior is consistent with Hamilton's observations and also the predictions of the G-S theory.

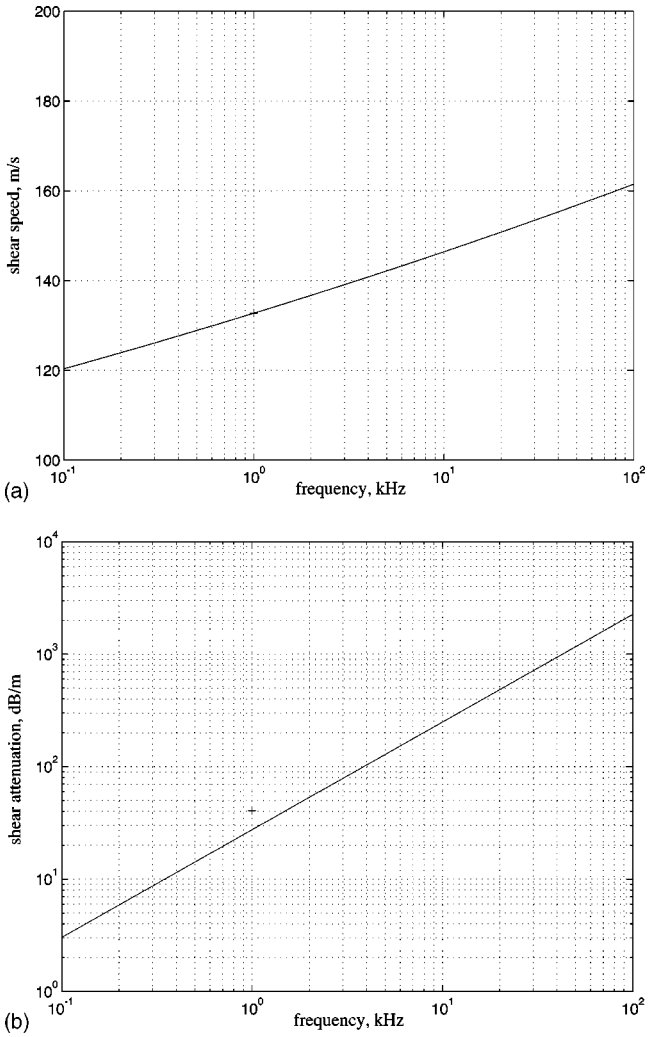


FIG. 3. Shear-wave (a) speed and (b) attenuation versus frequency from Richardson's North Sea site C1, well-sorted fine sand. The solid lines were evaluated from Eqs. (3) and (4) using the reported values (Ref. 33) of the porosity, grain size, and depth for the site [$N=0.358$, $u_g=238.2\ \mu$m ($\phi=2.07$), $d=0.28$ m].

B. Shear wave

In situ measurements of shear-wave properties as functions of frequency are rarer than those of compressional waves. In fact, no published, *in situ* measurements of shear-wave dispersion and attenuation over extended frequency ranges are known to the author.

The shear-wave speed and attenuation predicted from the G-S dispersion relations in Eqs. (3) and (4) are shown in Fig. 3. The curves were computed using values of the porosity, grain size, and depth in the sediment appropriate to Richardson's[12] North Sea site C1. The theoretical curve for the shear speed in Fig. 3(a) shows dispersion at a level of approximately 10% per decade of frequency, which is stronger than the dispersion in the compressional wave by a factor of 10 or so. It is evident from the uniform gradient of the curve in Fig. 3(a) that the predicted shear dispersion is logarithmic as expected, since c_s scales as $f^{n/2}$ in Eq. (3). The lone shear-speed datum in Fig. 3(a) falls precisely on the theoretical curve, since it was used in the evaluation of the compressional and shear coefficients, γ_{po} and γ_{so}. The predicted shear attenuation in Fig. 3(b) can be seen to exhibit a (near) linear dependence on frequency, varying as $f^{1-n/2}$. Note that the theoretical curve passes below but close to the sole reported data point[33] taken at a frequency of 1 kHz. This agreement provides a mild test of the G-S theory, since the shear attenuation was not used in computing γ_{po}, γ_{so}, or n. A single data point, of course, provides no information on the slope of the shear attenuation versus frequency curve, which is unfortunate because the gradient (as in the case of the compressional attenuation) provides an important test of theoretical models of waves in sediments. *In situ* measurements of shear dispersion and attenuation over extended frequency ranges are challenging but would be extremely useful for future validation of theoretical predictions.

V. DEPTH DEPENDENCE

A. Compressional wave

Published *in situ* data on the depth dependence of the sound speed and attenuation immediately beneath the seawater–sediment interface are scarce. Hamilton[34,35] discussed the variation of the sound speed with depth, but focused mainly on significantly greater depths than the meter or two of interest here. However, he did briefly consider the variation of the sound speed in the first 20 m of sand sediments,[34] for which he developed the following empirical power-law relationship, based on a curve fit to laboratory data:

$$c_{p\mathrm{H}}=1806d^{0.015}. \tag{13}$$

According to this expression, the sound speed is zero at the seawater–sediment interface and increases rapidly in the first meter or so, after which the rate of change slows significantly. No allowance for any variation of the sound speed with porosity is made in Eq. (13).

Hamilton[36] also considered the attenuation of sound as a function of depth. Based on laboratory measurements[37,38] of the variation of attenuation with confining pressure, he proposed an empirical inverse power-law relationship for the attenuation versus depth in sand

$$\alpha_{p\mathrm{H}}=0.45d^{-1/6}, \tag{14}$$

where the units of $\alpha_{p\mathrm{H}}$ are dB/m/kHz. Again, porosity is absent from this empirical representation. According to Eq. (14), the attenuation is infinite at the seawater–sediment interface but decreases very rapidly in the first few centimeters. It seems from curve B in Hamilton's[36] Fig. 3 that he considered Eq. (14) to be valid down to a depth of 150 m. Within the first few meters of the interface, however, his data are too poorly resolved in depth to say whether they support the strong negative gradient exhibited by Eq. (14).

Since Hamilton's[8] final review of the acoustic properties of sediments was published in 1987, few *in situ* measurements of near-interface sound-speed and sound-attenuation profiles in sand sediments have appeared in the literature. Acoustic profiles from cores, however, are available, some of the most recent being from Richardson *et al.*[25] for the SAX99 site. They measured sound speed and attenuation as functions of depth in a large number of cores at a frequency

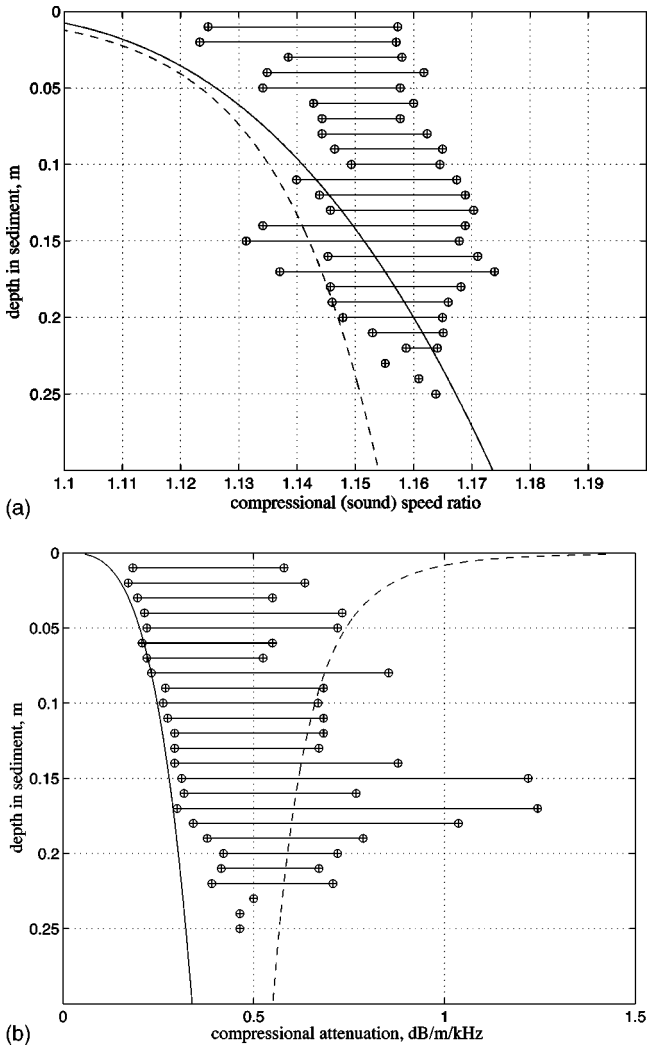


FIG. 4. Sound-speed (a) ratio and (b) attenuation versus depth for the medium-sand SAX99 site. The measurements were made on diver-collected cores at 400 kHz and each horizontal line represents the spread of the data at a given depth. The smooth solid lines, representing the G-S theory, were evaluated from Eqs. (1) and (2) using the reported values (Ref. 25) of the porosity, grain size, and measurement frequency [$N=0.377$, $u_g=414.7\ \mu$m ($\phi=1.27$), $f=400$ kHz]. The dashed lines, depicting Hamilton's empirical profiles, were evaluated from Eqs. (13) and (14).

of 400 kHz. The results are shown in Fig. 4, along with the G-S theory (smooth solid lines), evaluated taking the porosity as independent of depth.[39] Hamilton's empirical expressions are also included (dashed lines) for comparison. It should be noted that the G-S attenuation was computed for a frequency of 400 kHz and then divided by 400, to obtain the attenuation profile shown in Fig. 4(b), having units of dB/m/kHz. This procedure is consistent with that applied to the data. For the purpose of making the comparison between theory and data in Fig. 4(b), it would have been inappropriate to compute the attenuation directly for a frequency of 1 kHz, since this would have underestimated the theoretical result (relative to the data) due to the slight nonlinear dependence of the G-S attenuation on frequency [Fig. 2(b)].

At each depth increment, the sound speed and attenuation data show some spread, as indicated by the horizontal bars in Fig. 4. Note, however, that the variations in the sound speed are less than $\pm 2\%$, whereas those in the attenuation are of the order of $\pm 40\%$. As discussed below, this relatively high variability in the attenuation data may be attributable to various factors including scattering from inclusions such as shell fragments in the sediment as well as random coupling losses, which are expected whenever a sensor is inserted into a granular sedimentary medium.

Below a depth of 10 cm, the G-S theory yields a sound-speed profile that falls within the limits of the data in Fig. 4(a). At shallower depths, where the gradient in the sound speed is very high, the G-S theory falls a little below the lower limits of the data. This small discrepancy between the theory and the data may be due to the difficulty of measuring the sound speed near the interface, where the gradient of the upward-refracting profile is particularly steep: diving waves may reach the receiver first, giving rise to an overestimate of the sound speed. Note that the trends of the G-S sound-speed profile and Hamilton's empirical relationship are similar, with both showing a rapid increase within a few centimeters of the interface. In the case of the G-S profile, this high, near-interface gradient is an effect of the overburden pressure, which gives rise to a high rate of increase of the compressional coefficient [Eq. (11)] immediately below the boundary. At the boundary itself ($d=0$), the G-S profile takes Wood's value, c_o, whereas Hamilton's expression goes to zero.

Turning to the attenuation profiles in Fig. 4(b), it appears that the negative gradient of Hamilton's empirical curve is not consistent with the data (even allowing for the fact that the data show considerable spread). On the other hand, the G-S attenuation profile accurately tracks the lower-limiting values of the measured attenuation. As will become apparent, the G-S theory systematically delineates the lower limits of all the (compressional and shear) attenuation data sets examined in this paper, that is, the G-S theory traces the lower bound to the envelope of the measured attenuation values.

A plausible interpretation of this observation is that the G-S theory predicts the *intrinsic* attenuation due to intergranular interactions, which convert acoustic energy into heat, whereas the measurements represent the *effective* attenuation, that is, the total attenuation experienced by a sound or shear wave in propagating through the granular material. Thus, in addition to the intrinsic attenuation, the effective attenuation includes all other losses, notably those due to scattering from inhomogeneities such as shell fragments in the medium. Naturally, since shells and other scatterers are likely to be distributed randomly throughout the sediment, the effective attenuation is expected to be highly variable, but with a lower limiting value equal to the intrinsic attenuation. According to this argument, the G-S theory is expected to coincide with the lower bound of the data, as indeed it does in Fig. 4(b).

B. Shear wave

Figure 5(a) shows Richardson's[12] *in situ* measurements of the shear speed versus depth at 1 kHz in well-sorted fine sand at the North Sea C1 site. At each depth, the data point in Fig. 5(a) represents the mean of the several measured values reported by Richardson.[12] The data display a distinct positive gradient, with the shear speed approximately doubling be-

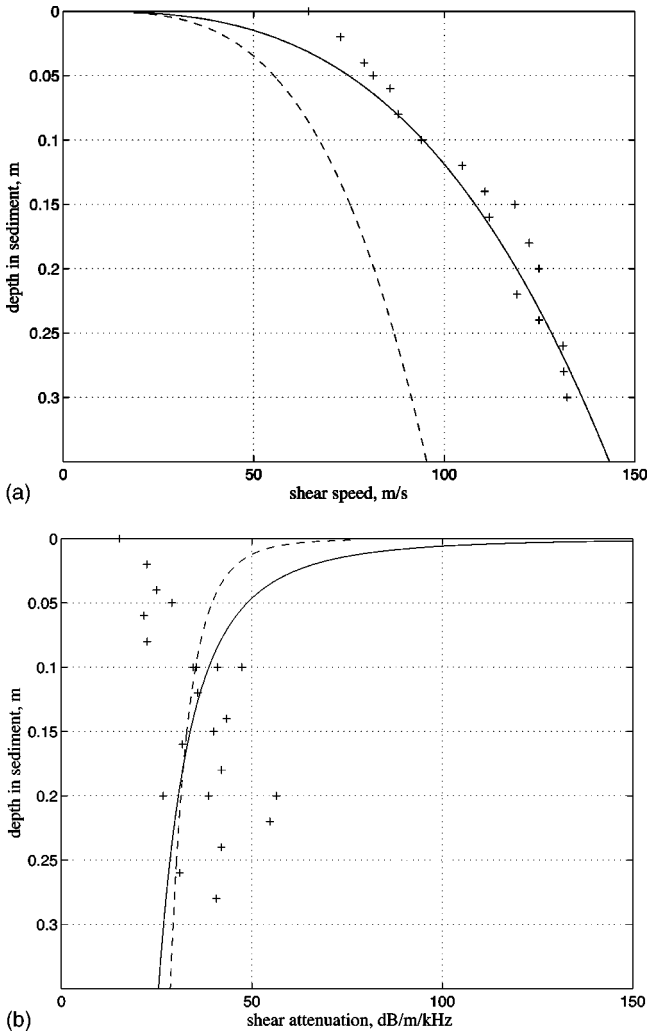


FIG. 5. Shear-wave (a) speed and (b) attenuation versus depth for the well-sorted fine sand at North Sea site C1. The smooth solid lines, representing the G-S theory, were evaluated from Eqs. (3) and (4) along with Eq. (12) using the reported values (Ref. 33) of the porosity, grain size, and measurement frequency [$N=0.358$, $u_g=238.2\ \mu$m ($\phi=2.07$), $f=1$ kHz]. The dashed lines are Hamilton's empirical profiles, as given by Eqs. (15) and (16).

tween depths of 5 and 30 cm. Similar behavior is predicted by the G-S theory [Eqs. (3) and (12)], as shown by the smooth solid line in Fig. 5(a), which increases as the cube root of depth in the sediment. In evaluating the G-S line in Fig. 5(a), the porosity was taken to be independent of depth, consistent with the porosity profiles measured on cores[25] and *in situ*[39] during the SAX99 experiment. It is clear that the theoretical curve in Fig. 5(a) closely matches the experimentally determined shear-speed profile.

The cube root of depth power law from the G-S theory is similar in form to an empirical shear-speed profile for sand sediments proposed by Hamilton[8]

$$c_{s\mathrm{H}}=128d^{0.28}. \tag{15}$$

This expression is plotted in Fig. 5(a) as the dashed line. In fact, Hamilton[8] discussed several empirical expressions having the form of Eq. (15) but with slightly different exponents, ranging from 0.25 to 0.312. Over the limited depth range of Richardson's data,[12] these exponents are almost indistinguishable, not only from one another but also from the G-S exponent of 1/3. However, Hamilton's (fixed) coefficient of 128 in Eq. (15) significantly underestimates Richardson's data in Fig. 5(a), by approximately 30% at a depth of 30 cm. It should be noted that, according to Eq. (12), the corresponding coefficient in the G-S theory is a function of porosity and grain size, and thus may differ from one sediment to another. Equation (15) obviously does not accommodate such behavior.

The shear attenuation measured by Richardson[33] at the North Sea C1 site is shown in Fig. 5(b). Included in the figure is the G-S theoretical prediction evaluated from Eqs. (4) and (12), again taking the porosity as independent of depth.[25,39] It is evident from these equations that the predicted attenuation scales as the reciprocal of the cube root of depth and that the coefficient of proportionality is a function of porosity and grain size. Below a depth of 10 cm, the G-S theory follows the lower bound of the attenuation data, which is consistent with the earlier argument that the G-S curve depicts intrinsic attenuation, as opposed to the effective attenuation represented by the data. At shallower depths, the G-S theory overestimates the measured attenuation, but Richardson[33] has suggested that these near-interface data points may in fact be too low as a result of ducting (diving waves) produced by the steep shear-speed gradient within the upper few centimeters of the sediment.

Hamilton[8] has proposed the following empirical expression for the shear attenuation profile:

$$\alpha_{s\mathrm{H}}=24d^{-1/6}, \tag{16}$$

which is shown as the dashed line in Fig. 5(b). Note that Hamilton's inverse-fractional-power-law scalings for the compressional attenuation [Eq. (14)] and shear attenuation [Eq. (16)] are exactly the same, both having an exponent of $-1/6$. Like the G-S shear attenuation profile, Eq. (16) shows a steep negative gradient immediately below the interface, whereas the gradient of Richardson's data is weakly positive. As mentioned above, this discrepancy may be the result of ducting just below the interface.

Figure 6 shows the shear-speed profiles of three finer-grained sediments, measured *in situ* by Richardson *et al.*[10] at Boa Dragaggio (fine sands and silt clays), Viareggio (silt clay), and Portovenere (silt and clay). (No shear-attenuation data were reported for these sites.) In all three cases, the shear speed is significantly less than that at the fine-sand, North Sea C1 site shown in Fig. 5(a). Notice the subtle differences between the Boa Dragaggio, Viareggio, and Portovenere profiles, and the fact that all three profiles are accurately reproduced by the G-S theory, evaluated from Eqs. (3) and (12) using the reported porosities and grain sizes.

The difference between the observed shear-speed profiles at Boa Dragaggio and Vareggio [Figs. 6(a) and 6(b)] is particularly interesting, since the reported porosities of these two sediments are almost the same, at 0.57 and 0.58, respectively. It follows that the difference in the shear speeds at the two sites, amounting to about 10 m/s at a depth of 2 m, must be due to the difference between the mean grain sizes. Boa Dragaggio is the coarser of the two sediments, with a reported mean grain diameter[10] of 3.4 μm ($\phi=8.2$), compared

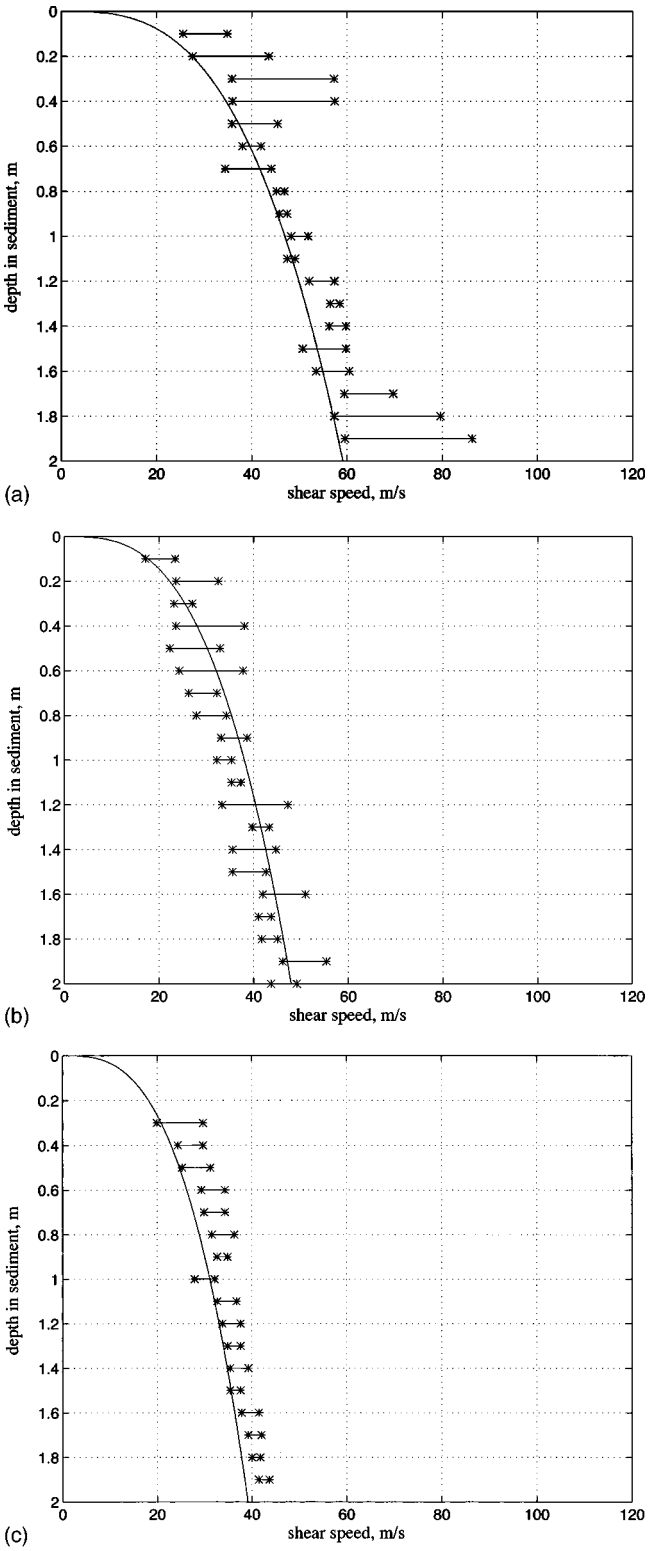


FIG. 6. Shear-speed profiles from (a) Boa Draggagio [$N=0.57$, $u_g=3.4$ micron ($\phi=8.2$), $f=1$ kHz]; (b) Viareggio [$N=0.58$, $u_g=1.82$ micron ($\phi=9.1$), $f=1$ kHz]; and (c) Portovenere [$N=0.63$, $u_g=1.05$ micron ($\phi=9.9$), $f=300$ Hz]. The smooth curves, representing the G-S theory, were evaluated from Eqs. (3) along with Eq. (12) using the reported values (Ref. 10) of the porosity, grain size, and measurement frequency for each of the sites, as listed here in square brackets.

with 1.82 μm ($\phi=9.1$) at Viareggio. In the G-S theory, the grain size affects the shear speed only through the shear modulus, γ_s, given by Eq. (12). According to this expression, the shear modulus for Boa Dragaggio is greater than that for Viareggio by a factor of 1.5. This is sufficient to give rise to slightly dissimilar theoretical shear-speed profiles for the two sites, in excellent agreement with the data, as shown in Figs. 6(a) and 6(b).

At the Portovenere site [Fig. 6(c)], the porosity is marginally higher and the grain size a little lower than at Boa Dragaggio and Viareggio. The enhanced porosity affects the bulk density, ρ_o in Eq. (5), and also the shear modulus, γ_s in Eq. (12), while the smaller grain size influences only γ_s. The net effect, as predicted by Eq. (3), is a slightly slower shear-speed profile than that at either of the other two sites. Again, it can be seen that the theoretical curve and the data are in very good agreement.

VI. RELATIONSHIPS BETWEEN PHYSICAL PROPERTIES

Although the porosity, grain size, and bulk density appear independently in the G-S dispersion expressions, these quantities are not themselves independent: finer-grain sediments tend to have lower densities and higher porosities than coarser materials. In order to evaluate the G-S dispersion relations as a function of any one of the three physical properties, porosity, grain size, and density, it is necessary first to establish the relationships that exist between them.

Of these inter-relationships, that between bulk density and porosity is well known and has already been introduced as the linear combination of the two constituent densities in Eq. (5). This expression accurately matches the data, as exemplified in Fig. 1. It follows that, provided the densities of the fluid and mineral phases are known, the bulk density can be determined directly from measurements of the porosity and vice versa.

The functional dependence of porosity on grain size is more difficult to treat, not least because porosity is not uniquely determined by grain size: sediments with identical porosities may exhibit mean grain sizes that differ from one another, as exemplified by the sediments investigated by Richardson *et al.*[10] at Boa Draggagio and Viareggio [see Figs. 6(a) and (b)]. Interestingly, Hamilton's data on porosity versus grain size show a significantly smaller scatter than the corresponding data collected from numerous sites by Richardson, a difference that becomes clear on comparing Figs. 7(a) and 7(b). However, a trend common to both data sets is that finer-grain sediments tend to exhibit higher porosities.

Any variation of the porosity with grain size represents a departure from the way smooth, uniform spheres pack together. Hamilton[8] has attributed the observed variation of porosity with grain size, illustrated in Fig. 7, to several factors including nonuniformity in the size and the shape of the grains.

Amongst unconsolidated marine sediments, the lowest porosities are found in the coarse sands, almost always taking values close to 0.37. As it happens, 0.37 is also the porosity of a random ''close'' packing of smooth, uniform

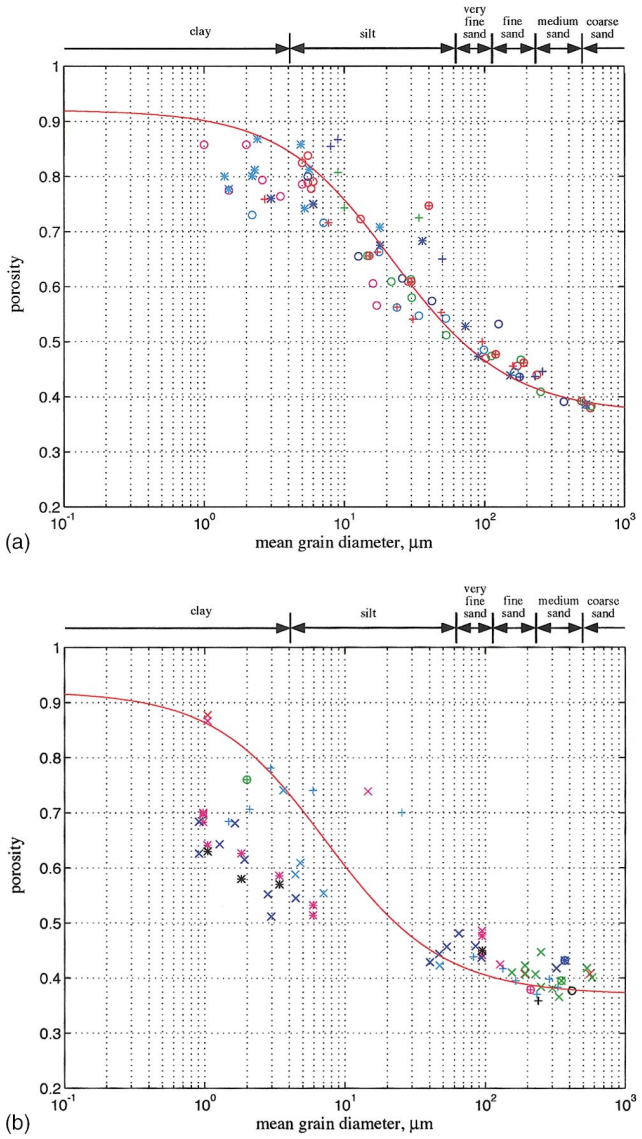


FIG. 7. Porosity versus grain size. (a) Comparison of Hamilton's data with Eq. (17), evaluated taking $\Delta=3$ μm (red curve). (b) Comparison of Richardson's data with Eq. (17), taking $\Delta=1$ μm (red curve).

spheres,[40,41] which suggests that, in the coarser sediments, grain shape (or roughness) effects represent a negligible departure from sphericity, and thus the packing is much like a random packing of smooth spheres. In the finer-grained sediments, on the other hand, surface roughness may be comparable with or much greater than the mean grain diameter, in which case close contact between adjacent grains is prevented, thus allowing pore water to percolate between grains, which results in an increase in the porosity. Grain ''shape'' and ''roughness'' in this context cover a multitude of nonspherical conditions, encompassing smooth, potato-like grains, high-aspect-ratio platelets, and very spiky, hedgehog-like particles.

Based on these ideas (i.e., a random packing of rough spheres), Buckingham[42] developed the following relationship between porosity and grain size:

$$N=1-P_s\left\{\frac{u_g+2\Delta}{u_g+4\Delta}\right\}^3, \tag{17}$$

where $P_s=0.63$ is the packing factor of a random arrangement of smooth spheres and Δ is the rms roughness measured about the mean (equivalent volume sphere) surface of the grains. The inverse of Eq. (17) gives the grain size as a function of the porosity

$$u_g=\frac{2\Delta(2B-1)}{1-B}, \tag{18}$$

where

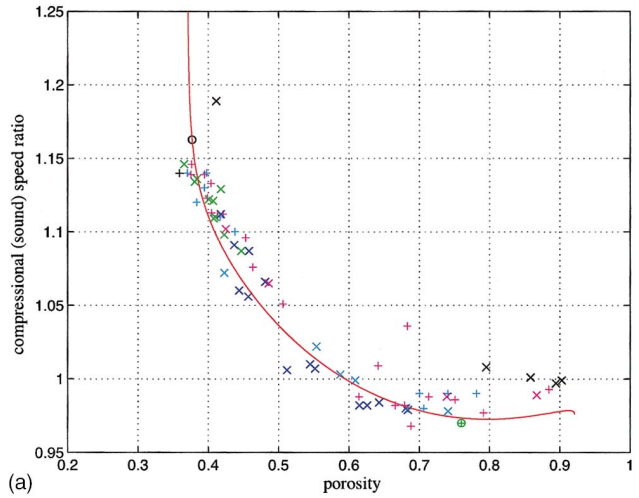
$$B=\left\{\frac{1-N}{1-N_{\min}}\right\}^{1/3}. \tag{19}$$

According to Eq. (17), when the grain size is very much greater than the rms roughness, the porosity approaches its minimum value, $N_{\min}=1-P_s=0.37$, appropriate to the coarse-sand regime of Fig. 7. At the other extreme, when the roughness is very much greater than the grain diameter, as with the high-aspect-ratio clay and silt particles, the porosity takes its maximum value, $N_{\max}=1-(P_s/8)=0.92$, which conforms with the fine-grain, high-porosity data in Fig. 7.

It is clear from the presence of Δ, the rms roughness parameter, in Eq. (17) that sediments of nominally the same mean grain size may exhibit different porosities. A sharp sand having very rough grains (high Δ) may be considerably more porous than an otherwise similar smooth-grained sand (low Δ). Although the rms roughness of the grains is not a parameter that is normally reported, the value of $\Delta=3$ μm yields a relationship between porosity and grain size from Eq. (17) that follows the average trend of Hamilton's data in Fig. 7(a), whereas a somewhat lower value of $\Delta=1$ μm is more appropriate to the average of Richardson's data, as illustrated in Fig. 7(b).

It is not clear why Hamilton's and Richardson's data sets in Figs. 7(a) and 7(b), respectively, should differ so markedly. The differences appear not only in the average trends but also in the scatter of the data about the mean porosity, especially for the finer-grained materials, a spread which is noticeably greater in Richardson's data. Perhaps the reason for the disparities between the two data sets is nothing more than coincidence in that the majority of the sediments examined by Richardson just happened to have smoother grains, and hence lower porosities, than those analyzed by Hamilton.

In order to make comparisons between the theoretical and measured wave properties as functions of the porosity and the grain size, an ''optimum'' value of Δ must be selected for substitution into Eq. (17). Since most of the data in the following comparisons stem from Richardson's measurements, the value of $\Delta=1$ μm is adopted for the rms roughness parameter, consistent with the comparison between data and theory in Fig. 7(b). With a fixed value of Δ in Eq. (17), the theoretical predictions of the wave properties from the G-S dispersion relations are, of course, single valued in the porosity and in the grain size, but the resultant curves are useful for comparison with the average trends of the multivalued experimental data.



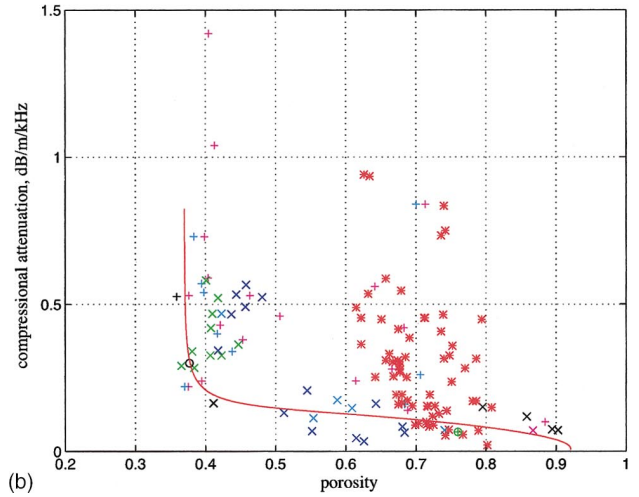
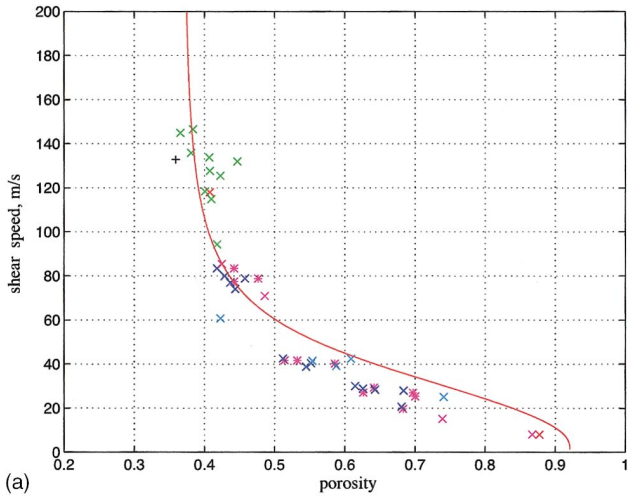


FIG. 8. (a) Sound-speed ratio versus porosity and (b) sound attenuation versus porosity. The red curves were evaluated from the G-S theory [Eq. (1)] with $f=38$ kHz and $d=0.3$ m, and $\Delta=1$ μm in Eq. (17). To obtain the attenuation in dB/m/kHz, for comparison with the data, the value computed at 38 kHz was divided by 38.



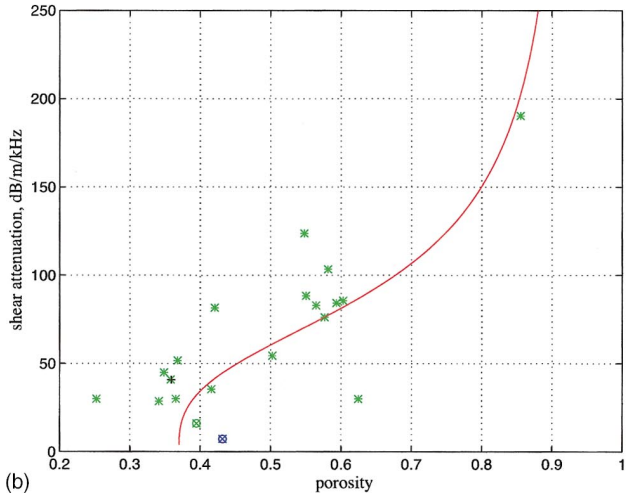


FIG. 9. (a) Shear-wave speed versus porosity and (b) shear-wave attenuation versus porosity. The red curves were evaluated from the G-S theory [Eqs. (3) and (4)] with $f=1$ kHz and $d=0.3$ m, and $\Delta=1$ μm in Eq. (17). [Note: The shear wave data reported in several of Hamilton's papers were mostly determined indirectly from measurements of interface waves. As they appear to be distinctly different in character from more recent direct measurements of shear wave properties, his data are not included in these plots.]

VII. POROSITY DEPENDENCE

A. Compressional wave

The ratio of the sound speed in the sediment divided by that in the water column is much less sensitive to temperature variations amongst sediments than the sound speed itself. Figure 8(a) shows a plot of the sound-speed ratio versus porosity computed from the G-S theory [Eq. (1)], along with data from a large number of sediments. The abundance of data points in Fig. 8(a) reveals a clear trend over the porosity range from $N\approx0.37$ to $N\approx0.92$, covering coarse sands to clays. Throughout this range, the theoretical curve accurately follows the trend of the data.

An interesting feature of both the data and the theoretical curve in Fig. 8(a) is the extremely steep gradient of the sound speed at porosities corresponding to the coarser sands ($0.37<N<0.4$). As the porosity rises marginally above its lower limit of 0.37, the sound speed plummets, eventually passing through a broad minimum around $N\approx0.8$. In view of the sensitivity to the porosity in the coarser materials, it is essential to have a high-precision measure of N available if the sound speed for a sand sediment is to be predicted accurately from the G-S theory.

The compressional attenuation as a function of porosity is shown in Fig. 8(b), where the smooth curve was computed by evaluating the G-S attenuation [Eq. (2)] for a frequency of 38 kHz (i.e., the frequency at which most of the data were collected) and dividing the result by 38 to obtain the predicted attenuation in dB/m/kHz. It is evident that the scatter is much higher in the attenuation data [Fig. 8(b)] than in the sound-speed data [Fig. 8(a)]. As with the depth profiles of attenuation in Fig. 4(b) (compressional) and Fig. 5(b) (shear), the G-S theoretical curve tracks the lower bound of the widely spread attenuation data in Fig. 8(b) (if allowance is made for relatively large measurement errors in the lowest attenuations). Such behavior is consistent with the fact that the theory predicts the intrinsic attenuation, due to the irreversible conversion of wave energy into heat, whereas the measurements represent the randomly distributed effective attenuation. As discussed earlier, in addition to the intrinsic attenuation, the effective attenuation includes all other types

of loss, due for instance to scattering from inhomogeneities such as shell fragments in the medium.

Like the sound speed, the theoretical intrinsic attenuation decays extremely rapidly as the porosity rises incrementally above the minimum value of 0.37, with a similar rapid decay exhibited by the data. It follows that, in order to predict accurately the intrinsic attenuation in a sand sediment, a high-precision estimate of the porosity is required.

B. Shear wave

Figure 9(a) shows the speed of the shear wave as a function of porosity, with the smooth curve representing the G-S theory [Eq. (3)]. The clear downward trend in the data is reproduced well by the theoretical curve. The finer, high-porosity sediments exhibit the slowest shear speeds, with values as low as 10 m/s in the highest porosity materials, the clays. At the opposite extreme, the coarse sands with porosities around 0.4 show shear speeds in excess of 100 m/s. As with the compressional speed and attenuation, the shear-wave speed decays extremely rapidly in the coarse materials as the porosity increases slightly above its lowest value of 0.37. This behavior can be clearly seen in both the theoretical curve and the data. Because of the very high gradient at porosities in the vicinity of 0.4, a high-precision measurement of porosity would be required in order to predict accurately the speed of the shear wave in a sand sediment.

Figure 9(b) shows the shear attenuation as a function of porosity, with the smooth curve representing the prediction of the G-S theory [Eq. (4)]. Relatively few data points appear in Fig. 9(b), reflecting the difficulty of making *in situ* shear-attenuation measurements. Nevertheless, sufficient data are present to identify a lower bound to the effective attenuation, a boundary which is accurately traced by the intrinsic attenuation curve from the G-S theory. Apart from one errant point at $N=0.63$, the data lie on or above the theoretical line, again consistent with the idea that scattering and other loss mechanisms may add to the intrinsic attenuation predicted by the theory to yield the effective attenuation of the measurements.

VIII. GRAIN-SIZE DEPENDENCE

A. Compressional wave

Figure 10(a) shows the sound-speed ratio as a function of grain size, with the smooth curve representing the G-S theory [Eq. (1)]. The data are well distributed throughout the full range of grain sizes, from clays to coarse sands, and show a distinct upward trend with increasing grain size. The G-S theory follows the average trend of the data very satisfactorily.

The compressional attenuation as a function of the grain size is shown in Fig. 10(b), where the smooth curve represents the G-S theory [Eq. (2)], which was evaluated for a frequency of 38 kHz (i.e., the measurement frequency for much of the data) and divided by 38 to obtain dB/m/kHz. Throughout the range of grain sizes, from clays to coarse sands, the data points show a high degree of scatter as expected, since the data represent the effective attenuation. The lower boundary of the envelope occupied by the data is accurately traced by the theoretical curve computed from the G-S theory for the intrinsic attenuation.

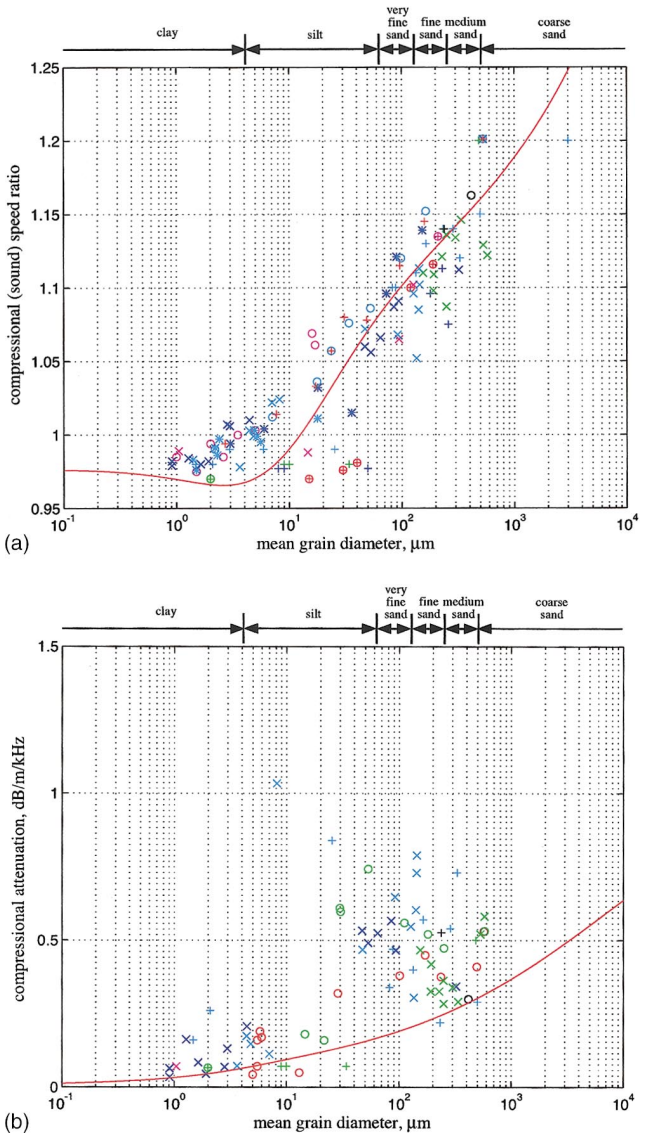


FIG. 10. (a) Sound-speed ratio versus mean grain diameter and (b) sound attenuation versus mean grain diameter. The red curves were evaluated from the G-S theory [Eqs. (1) and (2)] with $f=38$ kHz and $d=0.3$ m, and $\Delta=1$ μm in Eq. (17). To obtain the attenuation in dB/m/kHz, for comparison with the data, the value computed at 38 kHz was divided by 38. [Note: Hamilton's data have been included in these plots because the G-S theory and the experimental data indicate an insensitivity to the grain roughness parameter, Δ that is, Hamilton's and Richardson's data follow much the same trend.]

B. Shear wave

Data on shear-wave properties as a function of grain size are less abundant than those on the compressional wave. The shear speed versus mean grain diameter is plotted in Fig. 11(a), where the smooth curve represents the G-S theory [Eq. (3)] evaluated at a frequency of 1 kHz. Although a gap appears in the distribution of the data points between 15 and 40 μm, a strong upward trend in the measured shear speed with increasing grain size is still easy to distinguish. Similar behavior is exhibited by the G-S theoretical curve.

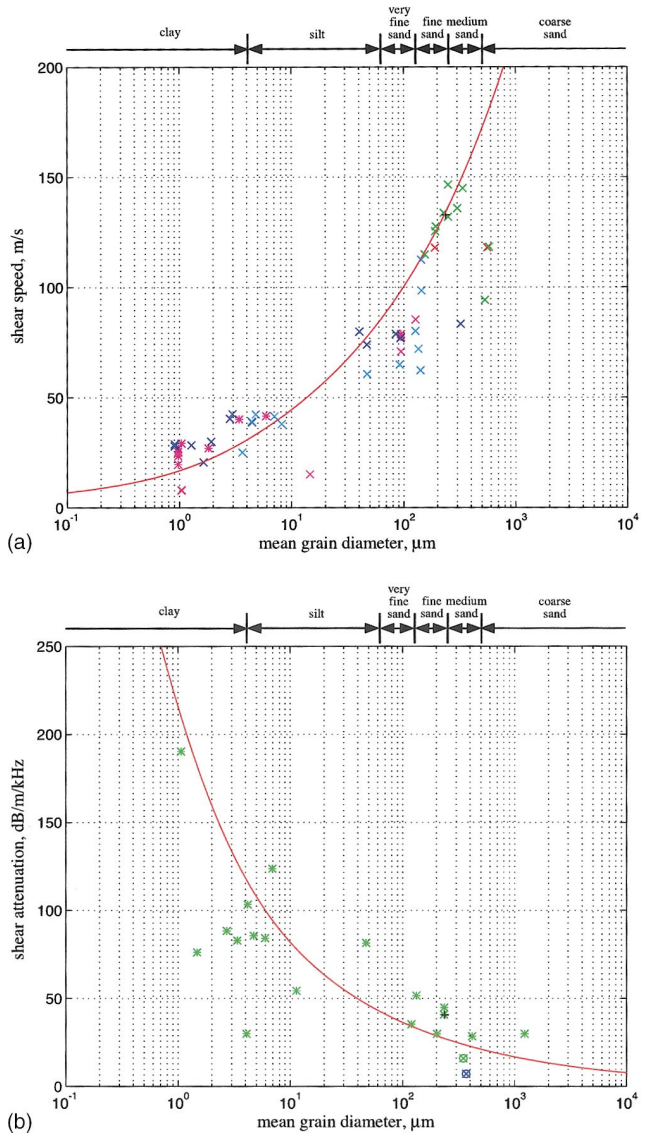


FIG. 11. (a) Shear-wave speed versus mean grain diameter and (b) shear-wave attenuation versus mean grain diameter. The red curves were evaluated from the G-S theory [Eqs. (3) and (4)] with $f=1$ kHz and $d=0.3$ m, and $\Delta=1$ μm in Eq. (17). [Note: The comment on Hamilton's shear wave data in the legend to Fig. 9 also applies here.]

As pointed out by Richardson,[12] shear attenuation has been measured at far fewer sites than shear speed. Those measurements of shear attenuation that are available are plotted in Fig. 11(b) against the mean grain diameter. The smooth curve in the diagram represents the G-S theory evaluated for a frequency of 1 kHz. Many of the data points fall close to or above the theoretical curve, again consistent with the idea that the data represent effective attenuation, the lower limiting value of which is the intrinsic attenuation yielded by the theory.

However, a group of data points in Fig. 11(b), representing very-fine-grained materials, silts and clays, with mean grain diameters in the range between 1 and 10 μm, falls noticeably below the theoretical curve. It is not clear why these particular data points are seemingly too low (or the theoretical curve too high), especially as the very same measured values of shear attenuation lie above the theoretical curve in the plot of shear attenuation versus porosity in Fig. 9(b). The data points in question in Fig. 9(b) fall in the porosity range from approximately 0.55 to 0.6. In view of the extreme difficulty of making precision *in situ* measurements of shear attenuation in very-fine-grained sediments (shear speeds less than 50 m/s), the explanation for the apparently low data points in Fig. 11(b) may simply be a high level of uncertainty in these measurements. Since no error bars were reported with the original data set, it is difficult to ascertain whether this interpretation is plausible but the anomaly should be resolved, one way or the other, as more *in situ* shear-attenuation data become available.

IX. A SHEAR-WAVE INVARIANT

If the product of the G-S expressions for the shear speed [Eq. (3)] and intrinsic attenuation [Eq. (4)] is formed, the resultant expression is

$$c_s\alpha_s = 2\pi f \tan\left(\frac{n\pi}{4}\right) = 0.42f, \tag{20}$$

where f is frequency (Hz) and the value of n listed in Table II has been used to evaluate the scaling constant (0.42) on the right. According to Eq. (20), the product of the shear speed and attenuation is directly proportional to the frequency. The scaling constant is a function only of the strain-hardening coefficient n, being independent of all the macroscopic material properties of the sediment, that is, the porosity, the grain size, the density, and the overburden pressure. Since n represents microscopic processes occurring at grain contacts,[17] including strain hardening in the molecularly thin layer of fluid separating asperities, it is expected to be essentially constant for all sediments composed of similar materials, for instance, quartz grains and seawater. Thus, the constant 0.42 in Eq. (20) should hold for all silicilastic marine sediments

In principle, the predicted invariance of the shear speed times the shear attenuation provides a good test of the G-S theory. In practice, however, the number of sediment sites for which both shear-wave properties are available is almost vanishingly small, making such a test impracticable at present.

X. CONCLUDING REMARKS

The properties of the phase speed and attenuation of compressional and shear waves in marine sediments have been examined in this article. Detailed comparisons have been made between measured wave properties, taken from the literature, and the predictions of a recently developed grain-shearing (G-S) theory of wave propagation in saturated porous media. The theory takes the form of two dispersion pairs, that is, four expressions, representing the phase speed and attenuation of the compressional wave and the shear wave. In addition to frequency, the four G-S dispersion relationships are functions of porosity, density, and grain size, and also overburden pressure, which translates into depth in the sediment.

Besides the material properties, the theoretical expressions involve three real parameters, representing microscopic shearing processes that are postulated to occur at grain con-

tacts during the passage of a wave. Two of these parameters, the compressional and shear coefficients, are analogous to the Lamé coefficients of elasticity theory, and the third is a numerical index representing the phenomenon of strain hardening in the molecularly thin layer of pore fluid sandwiched between grain contacts. Each of the three grain-shearing parameters has been assigned a value, which was held fixed in all the comparisons with data. Thus, the only available variables in the theoretical dispersion relationships are the material properties (porosity, grain size, and overburden pressure) and the frequency.

The compressional and shear-wave speeds and attenuations have been plotted as functions of frequency, depth in the sediment, porosity, and grain size, in each case with all other variables held constant. In every comparison, the theory accurately matches the average trend of the data set. It is to be emphasized that no adjustable parameters were available in the theory to help achieve these multiple fits to data.

The high quality of the match to data even holds for the attenuation of both the compressional and the shear wave: the theory reliably traces out the *lower bound* to the highly variable attenuation values returned by measurements. A straightforward interpretation of this observation is that the theory predicts the *intrinsic* attenuation, which arises from the irreversible conversion of wave energy into heat, whereas the measurements yield the *effective* attenuation, that is, the intrinsic attenuation plus any additional sources of loss, such as scattering from inhomogeneities in the granular medium. Since inclusions such as shell fragments tend to be randomly distributed in the sediment, the effective (i.e., measured) attenuation shows large fluctuations, taking values that are bounded from below by the intrinsic attenuation, which is stable and well predicted by the theory. In the plots of the compressional attenuation as a function of depth, for example, the theory accurately traces out the profile of the lowest attenuation values observed at each depth [Fig. 4(b)].

It is concluded that the G-S dispersion relationships in Eqs. (1) to (4), with the three fixed parameters γ_{po}, γ_{so}, and n taking the values listed in Table I, are useful as a practical tool for making estimates of the wave and physical properties of a sediment (excluding the effective attenuation, of course, but including the intrinsic attenuation). For instance, if the sound speed were known, the shear speed, both (intrinsic) attenuations, the porosity, and the density could be evaluated immediately and uniquely from the G-S theoretical expressions. Similarly, if the porosity were available from, say, a core sample, the wave speeds and (intrinsic) attenuations could be evaluated from the G-S dispersion relations. Such utility has potential as the basis of a variety of inversion techniques for estimating the geoacoustic parameters of marine sediments.

ACKNOWLEDGMENTS

Throughout the evolution of this work, Dr. Michael Richardson, NRL Stennis, has been generous with his sediment data as well as providing valuable guidance on the material properties of sediments. The research was supported by Dr. Ellen Livingston, Ocean Acoustics Code, the Office of Naval Research, under Grant Number N00014-04-1-0063.

[1] E. L. Hamilton, G. Shumway, H. W. Menard, and C. J. Shipek, "Acoustic and other physical properties of shallow-water sediments off San Diego," J. Acoust. Soc. Am. **28**, 1–15 (1956).
[2] E. L. Hamilton, H. P. Bucker, D. L. Keir, and J. A. Whitney, "Velocities of compressional and shear waves in marine sediments determined *in situ* from a research submersible," J. Geophys. Res. **75**, 4039–4049 (1970).
[3] E. L. Hamilton, "Sound velocity and related properties of marine sediments, North Pacific," J. Geophys. Res. **75**, 4423–4446 (1970).
[4] E. L. Hamilton, "Elastic properties of marine sediments," J. Geophys. Res. **76**, 579–604 (1971).
[5] E. L. Hamilton, "Compressional-wave attenuation in marine sediments," Geophysics **37**, 620–646 (1972).
[6] E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1336 (1980).
[7] E. L. Hamilton and R. T. Bachman, "Sound velocity and related properties of marine sediments," J. Acoust. Soc. Am. **72**, 1891–1904 (1982).
[8] E. L. Hamilton, "Acoustic properties of sediments," in *Acoustics and the Ocean Bottom*, edited by A. Lara-Saenz, C. Ranz Cuierra, and C. Carbo-Fité (Consejo Superior de Investigacions Cientificas, Madrid, 1987), pp. 3–58.
[9] M. D. Richardson, E. Muzi, L. Troiano, and B. Miaschi, "Sediment shear waves: A comparison of *in situ* and laboratory measurements," in *Microstructure of Fine-Grained Sediments*, edited by R. H. Bennet, W. R. Bryant, and M. H. Hulbert (Springer, New York, 1991), pp. 403–415.
[10] M. D. Richardson, E. Muzi, B. Miashi, and F. Turgutcan, "Shear wave velocity gradients in near-surface marine sediment," in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht, 1991), pp. 295–304.
[11] M. D. Richardson, "*In-situ*, shallow-water sediment geoacoustic properties," in *Shallow-Water Acoustics*, edited by R. Zhang and J. Zhou (China Ocean, Beijing, China, 1997), pp. 163–170.
[12] M. D. Richardson, "Variability of shear wave speed and attenuation in surficial marine sediments," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, La Spezia, 2002), pp. 107–114.
[13] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[14] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[15] H. J. Simpson, B. H. Houston, S. W. Liskey, *et al.*, "At-sea measurements of sound penetration into sediments using a buried vertical synthetic array," J. Acoust. Soc. Am. **114**, 1281–1290 (2003).
[16] E. L. Hamilton, "Sound velocity, elasticity, and related properties of marine sediments, North Pacific," Naval Undersea Research and Development Center, Report No. NUC TP 143 (1969).
[17] M. J. Buckingham, "Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am. **108**, 2796–2815 (2000).
[18] A. B. Wood, *A Textbook of Sound*, 3rd ed. (Bell, London, 1964).
[19] S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 3rd ed. (McGraw-Hill, New York, 1970).
[20] E. L. Hamilton, "Shear-wave velocity versus depth in marine sediments: A review," Geophysics **41**, 985–996 (1976).
[21] R. D. Stoll, *Sediment Acoustics* (Springer, Berlin, 1989).
[22] M. D. Richardson and K. B. Briggs, "*In situ* and laboratory geoacoustic measurements in soft mud and hard-packed sand sediments: Implications for high-frequency acoustic propagation and scattering," Geo-Mar. Lett. **16**, 196–203 (1996).
[23] M. J. Buckingham and M. D. Richardson, "On tone-burst measurements of sound speed and attenuation in sandy marine sediments," IEEE J. Ocean. Eng. **27**, 429–453 (2002).
[24] M. D. Richardson, D. L. Lavoie, and K. B. Briggs, "Geoacoustic and physical properties of carbonate sediments of the Lower Florida Keys," Geo-Mar. Lett. **17**, 316–324 (1997).
[25] M. D. Richardson, K. B. Briggs, D. L. Bibee *et al.*, "Overview of SAX99: Environmental considerations," IEEE J. Ocean. Eng. **26**, 26–53 (2001).
[26] F. Collins and C. C. Lee, "Seismic wave attenuation characteristics," Geophysics **21**, 16–40 (1956).
[27] L. Knopoff and G. J. F. MacDonald, "The attenuation of small amplitude stress waves in solids," Rev. Mod. Phys. **30**, 1178–1192 (1958).
[28] W. I. Futterman, "Dispersive body waves," J. Geophys. Res. **67**, 5279–5291 (1962).

[29] K. L. Williams, ''An effective density fluid model for acoustic propagation in sediments derived from Biot theory,'' J. Acoust. Soc. Am. **110**, 2276–2281 (2001).
[30] K. L. Williams, D. R. Jackson, E. I. Thorsos, , ''Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media,'' IEEE J. Ocean. Eng. **27**, 413–428 (2002).
[31] M. J. Buckingham, ''A three-parameter dispersion relationship for Biot's fast compressional wave in a marine sediment,'' J. Acoust. Soc. Am. **116**, 769–776 (2004).
[32] H. J. Simpson and B. H. Houston, ''Synthetic array measurements of acoustical waves propagating into a water-saturated sandy bottom for a smoothed and a roughened interface,'' J. Acoust. Soc. Am. **107**, 2329–2337 (2000).
[33] Michael D. Richardson, personal communication, Feb. 2000.
[34] E. L. Hamilton, ''Sound velocity gradients in marine sediments,'' J. Acoust. Soc. Am. **65**, 909–922 (1979).
[35] E. L. Hamilton, ''Sound velocity as a function of depth in marine sediments,'' J. Acoust. Soc. Am. **78**, 1348–1355 (1985).
[36] E. L. Hamilton, ''Sound attenuation as a function of depth in the sea floor,'' J. Acoust. Soc. Am. **59**, 528–535 (1976).
[37] G. H. F. Gardner, M. R. J. Wyllie, and D. M. Droschak, ''Effects of pressure and fluid saturation on the attenuation of elastic waves in sands,'' J. Pet. Technol. **16**, 189–198 (1964).
[38] A. N. Hunter, R. Legge, and E. Matsukawa, ''Measurements of acoustic attenuation and velocity in sand,'' Acustica **11**, 26–31 (1961).
[39] R. A. Wheatcroft, ''*In situ* measurements of near-surface porosity in shallow-water marine sands,'' IEEE J. Ocean. Eng. **27**, 561–570 (2002).
[40] O. K. Rice, ''On the statistical mechanics of liquids, and the gas of hard elastic spheres,'' J. Chem. Phys. **12**, 1–18 (1944).
[41] M. R. Wyllie, A. R. Gregory, and L. W. Gardner, ''Elastic wave velocities in heterogeneous and porous media,'' Geophysics **21**, 41–70 (1956).
[42] M. J. Buckingham, ''Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments,'' J. Acoust. Soc. Am. **102**, 2579–2596 (1997).
[43] E. L. Hamilton, ''Sediment sound velocity measurement made *in-situ* from bathyscaph *Trieste*,'' J. Geophys. Res. **68**, 5991–5998 (1963).
[44] E. L. Hamilton, ''Low sound velocities in high-porosity sediments,'' J. Acoust. Soc. Am. **28**, 16–19 (1956).
[45] A. B. Wood and D. E. Weston, ''The propagation of sound in mud,'' Acustica **14**, 156–162 (1964).
[46] M. D. Richardson, ''Spatial variability of surficial shallow water sediment geoacoustic properties,'' in *Ocean-Seismo Acoustics: Low-Frequency Underwater Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York, 1986), pp. 527–536.
[47] B. A. Brunson and R. K. Johnson, ''Laboratory measurements of shear wave attenuation in saturated sand,'' J. Acoust. Soc. Am. **68**, 1371–1375 (1980).
[48] B. A. Brunson, ''Shear wave attenuation in unconsolidated laboratory sediments,'' in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht, 1991), pp. 141–147.
[49] M. D. Richardson and K. B. Briggs, ''On the use of acoustic impedance values to determine sediment properties,'' in *Acoustic Classification and Mapping of the Seabed*, edited by N. G. Pace and D. N. Langhorne (University of Bath, Bath, 1993), Vol. 15, pp. 15–24.
[50] E. G. McLeroy and A. DeLoach, ''Sound speed and attenuation, from 15 to 1500 kHz, measured in natural sea-floor sediments,'' J. Acoust. Soc. Am. **44**, 1148–1150 (1968).
[51] C. McCann and D. M. McCann, ''The attenuation of compressional waves in marine sediments,'' Geophysics **34**(6), 882–892 (1969).
[52] G. W. C. Kaye and T. H. Laby, *Tables of Physical and Chemical Constants and Some Mathematical Functions*, 12th ed. (Longmans Green and Co., London, 1959).
[53] M. D. Richardson, K. L. Williams, K. B. Briggs, and E. I. Thorsos, ''Dynamic measurement of sediment grain compressibility at atmospheric pressure: Acoustic applications,'' IEEE J. Ocean. Eng. **27**, 593–601 (2002).

Acoustic measurements of the sound-speed profile in the bubbly wake formed by a small motor boat

Svein Vagle
Institute of Ocean Sciences, PO Box 6000, Sidney, British Columbia, V8L 4B2 Canada

Holly Burch
821 Plymouth St. SW, Olympia, Washington 98502

(Received 12 December 2003; revised 8 September 2004; accepted 18 September 2004)

In situ measurements of the bubble field within wakes generated by a small motorboat show that the bubble field, shortly after the initial turbulent generation period, consists mainly of bubbles with radii between 20 and 200 μm. The subsequent dispersion of the wake field can be described using a model that includes bubble buoyancy and dissolution only, and the air volume fraction within the wakes decay exponentially with an e-folding time of between 40 and 60 s. Simultaneous measurements of sound propagating through the bubbly wake exhibit spectral banding due to waveguide propagation. Inversions using the inverse-square theory developed by Buckingham [Philos. Trans. R. Soc. London, Ser. A **335**, 513–555 (1991)] show that this acoustic inversion technique provide a viable means of estimating the low-frequency sound-speed profile in an upward refractive bubble layer when dispersion can be neglected. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1819502]

PACS numbers: 43.30.Es, 43.30.Nb, 43.30.Pc [WMC] Pages: 153–163

I. INTRODUCTION

The presence of large numbers of microbubbles in the upper ocean will have a profound influence on sound propagation by changing the speed of sound. Breaking waves naturally generates bubbles, but ship and boat hulls and propeller cavitations will also generate bubbles close to the water surface. Because the source of the bubbles is at, or close to, the surface, and also due to the buoyancy and dissolution of the bubbles, the highest air volume fractions and therefore the most significant sound-speed changes will in general occur close to the surface and decrease with depth. The sound speed will also be dispersive, resulting in sound speeds higher than the bubble-free sound speed at acoustical frequencies above the resonance frequency of the dominant bubble radius (typically 50 kHz) and reduced sound speeds at frequencies below this frequency.

The low-frequency asymptotic value of the speed is dependent only on the fluid compressibility due to bubbles and the density. Following Wood (1941), the low-frequency sound speed, c can be related to the air volume fraction, β (volume of air to the water–air mixture), the densities of air and water, ρ_a and ρ_w, and the speed of sound in air and bubble-free water, c_a and c_w, as

$$\frac{1}{c}=\frac{1}{c_w}\sqrt{\left[\beta\left(\frac{c_w}{c_a}\right)^2 m+\frac{(1-\beta)}{m}\right]\left[\frac{\beta}{m}+(1-\beta)m\right]}, \quad (1)$$

where

$$m=\sqrt{\frac{\rho_w}{\rho_a}}. \quad (2)$$

The effect of bubbles on the sound speed can be quite large. Even at a volume fraction of air as low as 10^{-4} the sound speed is reduced from 1500 m/s for pure seawater down to approximately 800 m/s for the mixture.

Direct methods have been developed to measure the near-surface bubble field and the sound-speed profile (Medwin, 1974; Medwin *et al.*, 1975; Lamarre and Melville, 1992, 1994; Farmer *et al.*, 1998; Terrill and Melville 1997, 2000). However, it is clearly a challenge to develop instrumentation that can withstand the hostile environment close to the ocean surface, and to keep sensors at fixed depths relative to the free surface.

This paper presents the results of a study into the evolution of the bubble field and the sound-speed profile in the wake of a small motorboat. Microbubbles are also strong acoustic scatterers, and this property has been used in the past in the studies of ship wakes (NDRC, 1945; Trevorrow *et al.*, 1994; Marmarino and Trump, 1996).

The local propagation path will influence sound propagation in the upper ocean (Wille and Geyer, 1984), and Farmer and Vagle (1989) suggested that the regular spectral structure they observed in ocean ambient noise data could be attributed to waveguide propagation effects through the near-surface bubble layer. They found that the spectrum of the ambient noise exhibited peaks that were approximately stationary over a single breaking event and generally consistent from one breaking event to the next, but that the position of the peaks changed significantly between different storms. In the present study the sound-speed profiles are estimated from an acoustic inversion technique for estimating near-surface sound speed and air-fraction profiles from the spectral modulations caused by the waveguide (Burch, 1998). The technique, based on the theory developed by Buckingham (1991), requires only simple acoustic measurements of the sound field at two points near the surface, and from these, theoretical inversions are made to estimate the sound speed. In addition to being very simple, this inversion technique provides a complementary measurement to *in situ* measurements. The approach relies on having coherent spectral struc-

ture over which to invert. Therefore, the bubble layer must show some level of homogeneity for protracted waveguide propagation to occur.

II. INVERSE-SQUARE SOUND-SPEED PROFILE

The near-surface sound-speed profile has been represented by exponential (Farmer and Vagle, 1989; Lamarre and Melville, 1994; Ye, 1997), power law (-2 to -4) (Terrill and Melville, 1997), and inverse-square (Buckingham, 1991) functional forms. Both the exponential and inverse-square forms lend themselves to exact analytical solutions for the near-surface sound field. However, we have chosen to focus on the inverse-square theory (even though references will be made to the exponential) rather than considering the exponential theory because the acoustic data we collected exhibit an acoustic signature which is dependent on receiver depth, and only the inverse-square theory can fully represent this feature (Buckingham, 1993).

Buckingham (1991) has developed a complete, exact solution for the sound file in a semi-infinite ocean with an inverse-square profile, and his theory is based on a simplified one-dimensional steady-state model of equilibrium conditions where turbulent diffusion of bubbles downward into the bubble layer is exactly balanced by the buoyant rise of bubbles. The model suggests that there is a simple relationship between the normalized surface gradient of the air-fraction profile, the turbulent diffusion coefficient, and the bubble rise speed, and predicts a regular spectral banding structure that is the result of the interference of a sum of normal modes in the upward-refracting waveguide.

The inverse-square sound-speed profile can be expressed as

$$\frac{1}{c^2}=\frac{1}{c_w^2}\left\{1+\frac{z_1^2}{(z+z_s)^2}\right\}. \tag{3}$$

In this expression, z is depth and positive downward, z_s is an offset in the depth coordinate to allow for a nonzero sound speed at the surface, c_w is the sound speed as $z\to\infty$, and z_1 is a measure of the curvature of the profile. The general shape of this function is illustrated in Fig. 1 for parameters obtained in one of the motorboat wakes discussed below.

The parameter z_1 can be considered the effective depth of the upward-refracting waveguide, and it gives an indication of the number of modes that can be supported. It follows that waveguides with larger values of z_1, associated with more pronounced profile curvature, support more modes. The surface is positioned at a depth of z_s in the model coordinate system, allowing for a nonzero sound-speed value at the upper, pressure-release boundary surface. Thus, three parameters: c_w, z_1, and z_s, completely describe the inverse-square sound-speed profile. The receiver and source positions within the coordinate system and the boundary conditions are also shown in Fig. 1. Note that the variable z is used to represent the shifted receiver depth $z+z_s$, and z' represents the shifted source depth $z'+z_s$. The cylindrical wave equation for the harmonic velocity potential, G at range r, depth z, and with unit source strength is written as

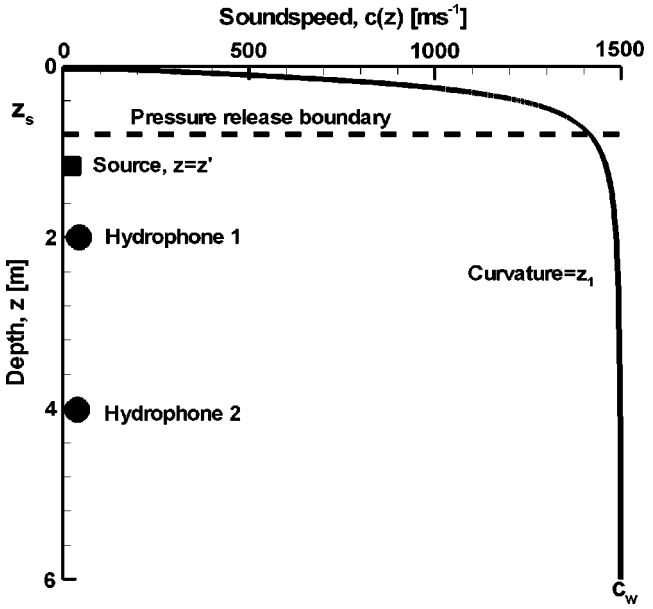


FIG. 1. Schematic diagram of the upward-refractive sound-speed profile. Particular profile shown has $z_1=0.28$ m, $z_s=0.8$ m, and $c_w=1500$ ms^{-1}.

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial G}{\partial r}\right)+\frac{\partial^2 G}{\partial z^2}+\frac{\omega^2}{c^2(z)}G=-\frac{1}{2\pi r}\delta(z-z'), \tag{4}$$

where ω is radial frequency and δ is the Dirac delta function. For a detailed description of the solution, refer to Buckingham (1991, 1993). Using Buckingham's work [(1993) Eqs. (5)–(7)], the mode functions for the solution to (4) are

$$F(\sigma_m\xi)=\begin{cases}-2\sqrt{\dfrac{2}{\pi}}(\nu^2-\sigma_m^2\xi^2)^{-1/4}\cos\left\{\nu\ln\left[\dfrac{\nu+\sqrt{\nu^2-\sigma_m^2\xi^2}}{\sigma_m\xi}\right]-\sqrt{\nu^2-\sigma_m^2\xi^2}-\dfrac{\pi}{4}\right\}(I)\\ -2\left(\dfrac{2}{\sigma_m\xi}\right)^{1/3}Ai\left[(\sigma_m\xi-\nu)\left(\dfrac{2}{\sigma_m\xi}\right)^{1/3}\right](II)\end{cases}, \tag{5}$$

where $\xi=z/z_s$. The expressions in (5) are matched at a point in the vicinity of the extinction depth, ξ_{ext}, which is defined as

$$\xi_{\text{ext}}=\frac{\nu}{\sigma_m}. \tag{6}$$

The Airy function of the first kind, Ai, contained in the expansion, characterizes the mode shape as a function of depth. For negative arguments, the Airy function is oscillatory, and then it passes into decay for positive arguments. This transition from an oscillating region to a decaying region is an

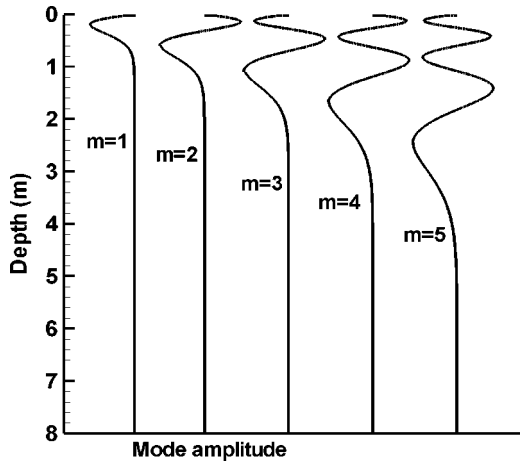


FIG. 2. Five inverse-square model modes ($m=1-5$) at a frequency of 10 kHz using $z_1=0.28$ m, $z_s=0.8$ m, and $c_w=1500$ ms^{-1}.

important feature of the inverse-square theory mode shapes because it indicates that energy is trapped near the surface in the oscillating region (Buckingham, 1991, 1993). The first five modes versus depth for a frequency of 10 kHz have been plotted in Fig. 2 using the sound-speed profile parameters shown in Fig. 1. At the extinction depth, a particular mode will pass from its oscillatory region into a region of decay. This extinction depth is a function of the mode number and frequency, and for a fixed frequency the extinction depth increases for increasing mode number. For a particular mode number the extinction depth decreases for increasing frequency (Fig. 3). The result is that the mode is compressed into a shallower region of the ocean as the frequency increases. When a receiver is placed below the extinction depth for a particular mode, that mode will not be a significant contributor to the field at that depth because it has passed into evanescence.

The normal mode component, G_{nm}, of the velocity potential from (4) [Buckingham, 1993, Eq. (5)] has been plotted in Fig. 4 for fixed source depth of 1.5 m and different receiver depths. The horizontal spacing between source and receiver was 12 m. Four spectra are plotted for the representative wake sound-speed profile used above. The far-field component of the analytical solution shows two characteristic periodicities in its frequency spectrum. This structure is produced by the interference of the modes comprising the solution. The first periodicity is analogous to a carrier frequency in the time domain, and the second is analogous to a modulation frequency.

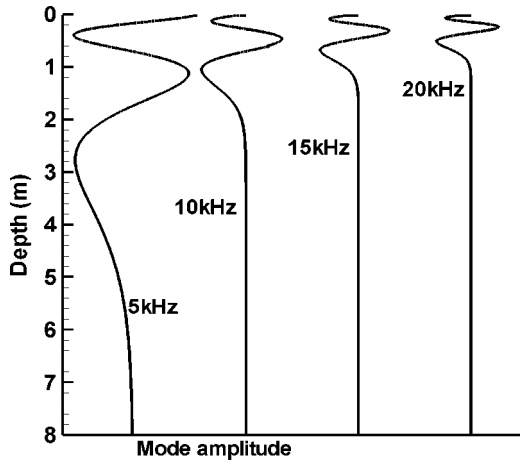


FIG. 3. Inverse-square model mode 3 at four different acoustical frequencies between 5 and 20 kHz using $z_1=0.28$ m, $z_s=0.8$ m, and $c_w=1500$ ms^{-1}.

Buckingham found simple expressions that approximate these two periodicities. They are dependent on the inverse-square sound-speed profile parameters and the receiver depth z for the carrier component Δf_0 and the source depth z' for the modulation component $\Delta f_0'$

$$\Delta f_0 \approx \frac{c_\infty}{2z_1 \ln(z/z_s)}, \tag{7}$$

and

$$\Delta f_0' \approx \frac{c_\infty}{2z_1 \ln(z'/z_s)}. \tag{8}$$

Figure 4 illustrates how these periodicities change as a function of receiver depth. As the receiver depth increases from 5 to 10, 15, and 20 m we see that the carrier frequency Δf_0 becomes more closely spaced. In contrast, the modulation component $\Delta f_0'$ remains constant because it depends on the source depth, which is the same for each spectrum. For comparison, the exponential sound-speed profile can be expressed as

$$c = c_\infty - \Delta c \exp(-z/d_e), \tag{9}$$

where Δc is the surface sound-speed anomaly and d_e is the e-folding depth. The far-field component of the exponential theory solution is a finite sum of normal modes, which sum to give a spectrum with regular peaks, which are separated by

$$\Delta f_e = \frac{c_\infty}{4d_e}\sqrt{\frac{c_\infty}{2\Delta c}}. \tag{10}$$

It is clear that the peak separation is only dependent upon the exponential sound-speed profile parameters and not the source or receiver depths or the horizontal range between the source and receiver as is the case for the inverse-square theory.

The expressions in (7) and (8) provide a means for determining the profile parameters z_1 and z_s from acoustic measurements at known receiver depths. Data from a vertical line array of receivers will display different periodicities as a function of depth. In a configuration where the horizontal range and the source depth stays the same for several hydrophones, there is a redundancy of measurements that allows for the determination of the unknowns z_s and z_1. The bubble-free water sound speed, c_w, can be measured separately, e.g., by obtaining a CTD profile in the vicinity of the sampling area.

III. THE MOTOR BOAT WAKE EXPERIMENTS

The study took place in Saanich Inlet off the West Coast of Canada in August 1995. During this time of the year precipitation is limited, the winds are light, and the inlet is strongly stratified. The instrumentation included three re-

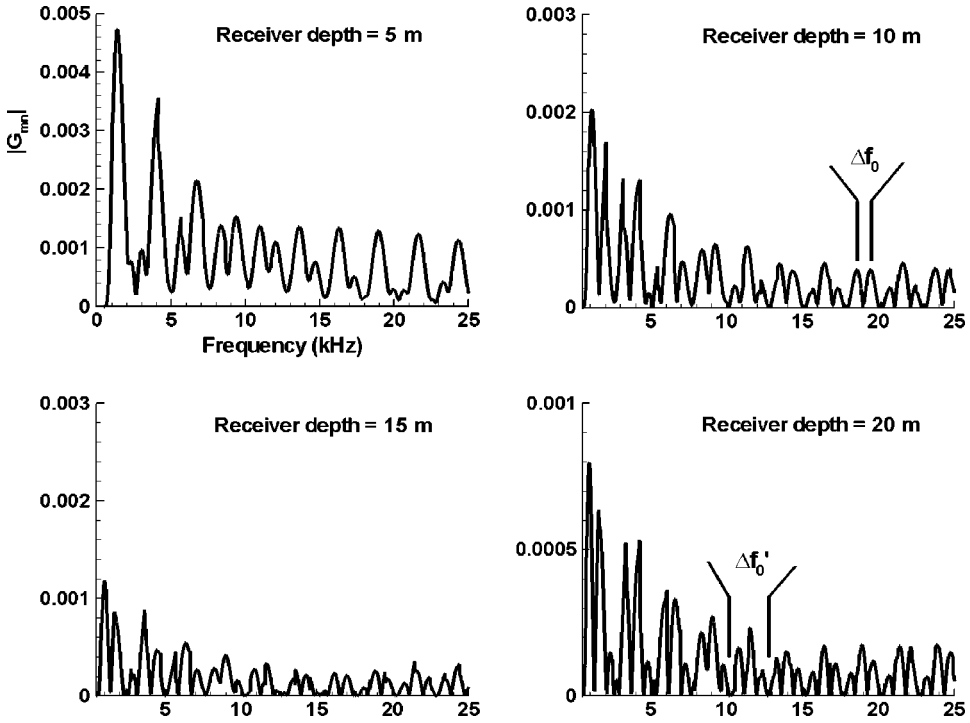


FIG. 4. Inverse-square spectra for a fixed source depth of 1.5 m and several different receiver depths between 5 and 20 m using $z_1 = 0.28$ m, $z_s = 0.8$ m, $c_w = 1500\ \mathrm{ms}^{-1}$, and a source to receiver spacing of 12 m. The carrier frequency increment Δf_0 becomes smaller with increasing depth. The modulation component $\Delta f_0'$ is dependent on the source depth rather than the receiver depth, so it remains constant as the receiver depth changes.

ceive hydrophones, a sound source, and four sensors for *in situ* bubble measurements. The broadband (0–20-kHz) acoustic source (ITC-1001) and a reference hydrophone (AN/SSQ-57A) were suspended at 1.1 and 1.4 m, respectively, at one end of a 12-m wooden beam (Fig. 5). The electronics were designed to produce a broadband noise spectrum between 50 Hz and 20 kHz that could be operated in continuous or pulsed mode. A Motorola MM5437 digital noise source IC chip was used to generate a pseudorandom noise signal, which was then low-pass filtered to a cutoff frequency of 20 kHz. Next, the signal was equalized using a Rane FPE 13 parametric equalizer to create an inverted form of the transducer's transmitting voltage response so that the acoustic signal from the source would be essentially flat across the spectrum. The transmitted sound level was approximately 180 dB *re*: 1 μPa at 18 kHz. The pulsed mode was used here to allow acoustical resonators (see below) to acquire interference-free signals during periods when the source was not transmitting. The duty cycle used during the experiment was 11-s transmission periods followed by 64 s of quiet time to give a cycling period of 75 s. This cycling allowed for an 11-s acoustic record of the evolving wake field every 1.25 min.

At the other end of the wooden beam was a 6-m-long instrument platform, which rigidly supported four acoustical resonators at depths of 0.5, 1.2, 2.5, and 4.6 m. These devices were used to measure the *in situ* bubble-size distribution at up to 40 bubble radii between 17 and 500 μm once a second (Farmer *et al.*, 1998; Vagle and Farmer, 1998) using the inversion approach developed by Commander and McDonald (1991).

Two modified sonobuoys (AN/SSQ-57A) were also mounted on the frame at depths of 2 and 4 m. Each sonobuoy system consisted of a hydrophone with a receive bandwidth from 10 Hz to 20 kHz, an rf transmitter with its own factory-set transmitting frequency, and an rf receiver on shore tuned to that sonobuoy's transmitting frequency. The rf signals were received on shore at a distance of about 4 km and were recorded on an 8-channel DAT recorder to Hi-8 tape. Later, these data were played back and the three hydrophone-channels were digitized simultaneously at 48 kHz.

Individual calibration of the sonobuoys used here was not feasible. However, repeatable calibrations of identical

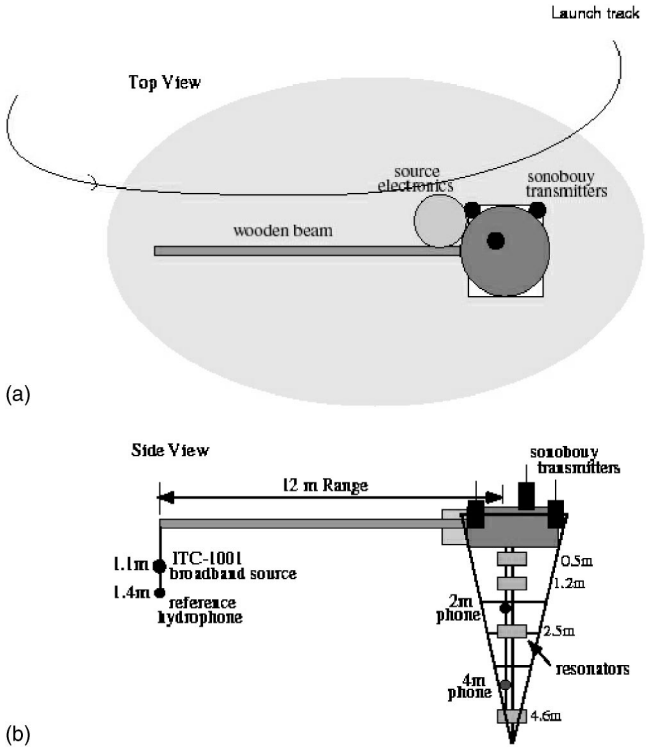


FIG. 5. Top view (a) and side view (b) of the experimental setup during the boat wake experiments.

TABLE I. Motor boat specifications.

Length at waterline	6.6 m
Beam (width of vessel at widest point)	3 m
Draft (water depth at bottom of vessel)	0.2 m
Propeller depth	0.53 m
Propeller diameter	0.38 m
Propeller type	Twin, 5-blade
Nominal propeller RPM	5000

systems have resulted in published receive response curves which fit within ± 2 dB of a standard curve for frequencies between 10 Hz and 10 kHz and increase to a maximum of ± 7 dB at 20 kHz (Horsley, 1989). Nevertheless, no absolute sound-level measurements are presented here, and only the differences between the two different systems have been used in the analysis. Analysis of the reference hydrophone data and measurements from periods without bubbles confirmed the flatness of the source spectrum and smooth frequency-dependent receive response of the hydrophones.

The experiment was conducted in a region of Saanich Inlet where the water depth is greater than 100 m. The bottom-reflected path is expected to be about 30 dB lower than the direct path due to spherical spreading and attenuation by the sediment upon reflection from the bottom, and it is reasonable to consider any bottom interactions negligible. During the experiment 11 wake fields were generated and to generate longer lasting, more spatially homogeneous bubble fields, several wakes were made adjacent to one another in a localized area by a small motorboat at high speeds. The boat was an 8-m flat-bottomed launch with two outboard engines (Table I). Several tracks were driven closely along both sides of the wooden beam as well as downwind of the beam to generate the wake fields. In contrast to larger vessels, the primary bubble injection mechanism for the small vessels we utilized in our experiments was propeller cavitation. The resulting wake fields, which had a greater horizontal than vertical extent, were characterized by a higher aspect ratio than single-pass wakes.

Even though the beam drifted relative to the water column due to some light wind forcing, it continued to remain within the wake field for periods of up to 10 min because the bubbly region was created downwind. The larger bubbles of the wake quickly left the field due to buoyancy effects, and the remaining field slowly evolved with time. The extent of the wake field was visible due to the damping of the short surface gravity and capillary waves.

IV. EVOLUTION OF A MOTORBOAT-GENERATED BUBBLE FIELD

The motorboat was driven at full speed towards one end of the instrument on the upwind side, along the wooden beam and then away (see Fig. 5). Immediately afterward the engines were stopped to avoid any further engine noise interference. This whole procedure was repeated approximately every 10–15 min. Figure 6(a) shows the total air fraction as a function of depth, and the depth-integrated air volume fraction for a period of approximately 2 h with eight such events. The air-volume fractions, β, are calculated from

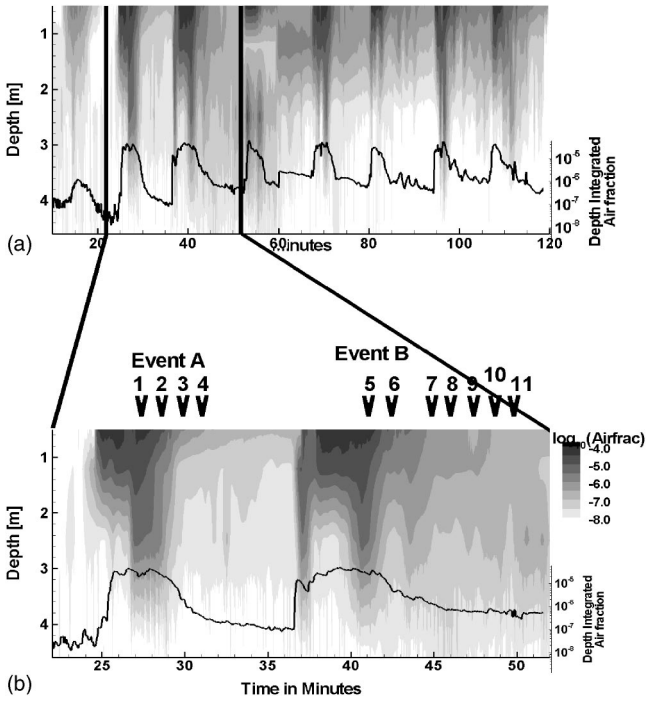


FIG. 6. Images showing the air-volume fraction calculated from the acoustical resonators using linear interpolation between the four depths for a 2-h period (a) and a 30-min expanded section (b). Eight bubble injections are shown in the figure. Also shown is the depth-integrated air-volume fraction and arrows showing the two events with the 11 active acoustic transmission periods discussed in the text and shown in Figs. 7–11.

the bubble-size distributions obtained from the acoustical resonators over the bubble radius range of 15 to 500 μm using

$$\beta = \tfrac{4}{3}\pi \int_{a=15\mu\mathrm{m}}^{a=500\mu\mathrm{m}} a^3 n(a)\,da, \tag{11}$$

where a is the bubble radius and $n(a)da$ is the number of bubbles per volume within a 1-μm radius increment. Summing the data from all four *in situ* sensors and averaging over 20 s resulted in the depth-integrated air fraction. When the total number of bubbles is relatively low ($\beta < 10^{-6}$) it is possible to get the dispersive sound-speed profiles directly from the resonators by tracking the location of the resonant peaks (Farmer *et al.*, 1998). However, at higher bubble concentrations the resonant peaks are no longer well enough defined and the attenuation and therefore the bubble-size distributions have to be used indirectly via (11) to estimate the sound-speed profile. Immediately after generation the bubble density is so high that the upper two resonators are saturated, making it impossible to determine the maximum air fraction of the wakes at this time. The saturation level of the resonators for typical bubble-size distributions corresponds to air fractions of approximately 10^4. These saturated levels dominate the depth-integrated air fractions for the first part of the events, as can be seen in Fig. 6. Within about a minute the largest bubbles will have left the water column and turbulence, buoyancy, and dissolution effects control the remaining bubble field. The data presented in Fig. 6 show that even modest wakes made by a small launch would persist with high air fractions for up to 5 min. The data also show that

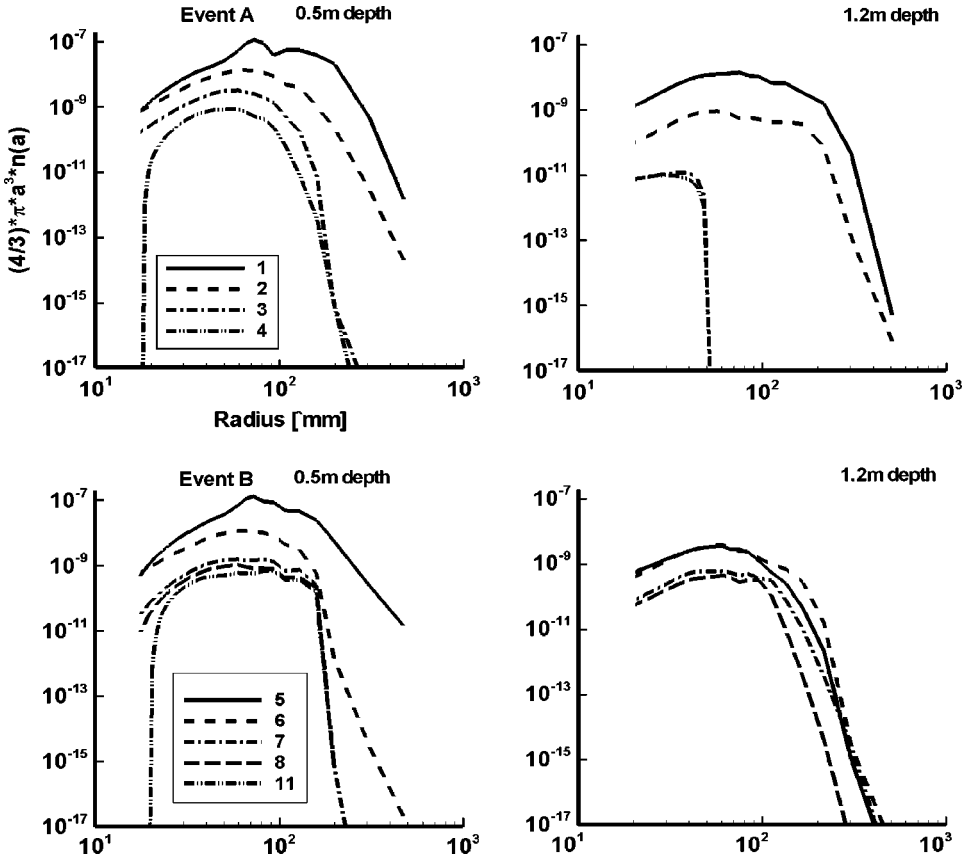


FIG. 7. Volume-scaled bubble-size distributions averaged over 11 s for the first four transmission periods during event A plotted at two depths: 0.5 and 1.2 m (upper two panels), and for periods 5–11 during event B at the same two depths (lower two panels). The different periods are indicated with arrows in Fig. 6(b).

with wake generation every 10 to 15 min the upper 1–2 m of the water column slowly gets filled by bubbles with a dominant radius of $\sim 50\ \mu$m.

Figure 7 shows a sequence of bubble-size distributions at two different depths, averaged over 11 s at the times indicated by arrows in Fig. 6(b), for two events. The size distributions have been plotted as volume-scaled distributions to emphasize the size of bubbles that dominate the air fraction. Even though all 11 runs show similar bubble fields and sound propagation characteristics, only these two events will be presented and discussed in detail here. These two runs were chosen as the ones with the least amount of radio frequency interference on the sonobuoy transmissions and the clearest signals on both upper and lower hydrophones. With time, following an injection bubbles with radii below 20 μm and bubbles with radii greater than 200 μm disappear due to dissolution and buoyancy, and one is left with a persistent shallow distribution of bubbles with radii between 20 and 100 μm.

An indication of the time evolution of the boat wake is presented in Figs. 8(a)–(c), where the air-volume fractions from the three shallowest resonators are plotted for three different wakes, the two shown in Fig. 6(b) and the one following. At the two shallowest depths the resonator signals are saturated for the first 200 s following the injection. For the next 200 s the air-volume fraction decays exponentially with an e-folding time constant of approximately 60 s at a depth of 0.5 m, and 40 s at 1.2 m. At 2.5 m the signals were not saturated, and for the first 160 s following a wake the air-volume fraction decayed exponentially with an e-folding time of approximately 40 s.

The generation of the bubble field is accompanied by intense turbulence, which quickly mixes the bubbles down to a depth of at least 2.5 m. In the present analysis we leave out the discussion of the bubble mixing because the high bubble densities saturate our sensors during this phase of the wake evolution. We therefore consider the behavior of the bubble field following the high-turbulence mixing period, taking account of buoyancy and gas dissolution. The initial size distribution of bubbles at the point of wake generation is taken to satisfy a power-law distribution of the form $a^{-2.5}$ (Deane and Stokes, 1999). The evolution of the bubble field after generation is evaluated using the approach presented by Thorpe (1982). The buoyant rise speed is determined by their radius and contamination of the bubble skin. Here, we make the simplifying assumption that bubbles are contaminated in this way at all time (Thorpe, 1982). The terminal rise speed of a bubble, w_b, depends on a balance between drag and buoyancy forces

$$C_D \frac{\pi a^2}{2} \rho w_b^2 = \rho g \frac{4\pi}{3} a^3, \tag{12}$$

where C_D is a drag coefficient (Keeling, 1993), ρ is the density of water, and g is the gravitational acceleration. For bubbles with radii greater than 560 μm the rise speed is assumed to be constant (0.3m/s) (Levich, 1962).

Calculation of the bubble motion must be followed for each individual bubble, taking due account of gas exchange and changes in hydrostatic pressure, since these affect the bubble radius and consequently its rise speed. Following Thorpe's (1982) approach, gas invasion or evasion across the bubble skin depends upon the Nusselt number, which is as-

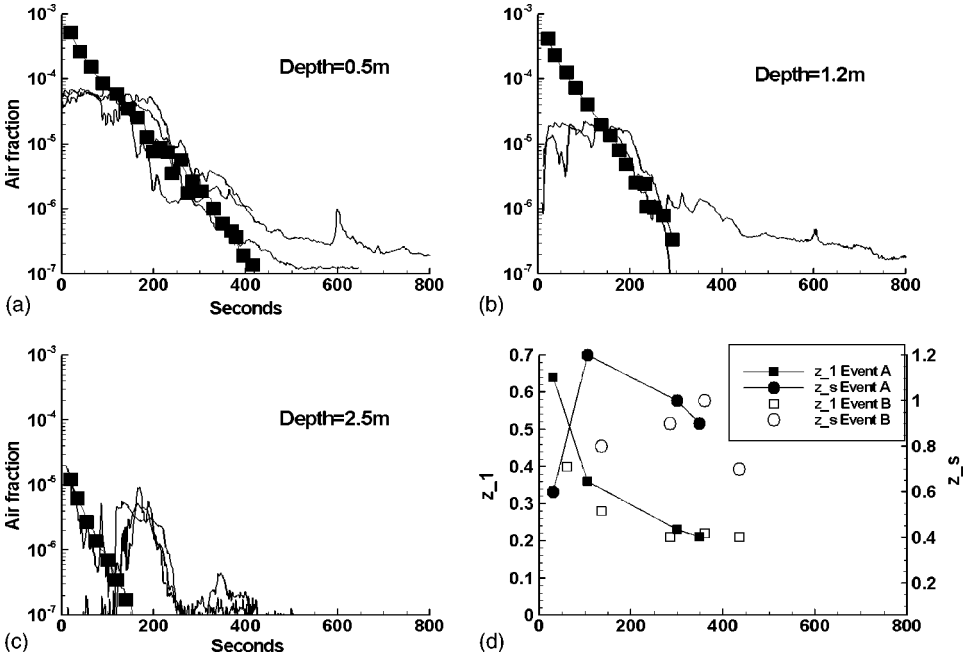


FIG. 8. Time evolution of the air fraction in three wakes (Fig. 6) at a depth of 0.5 m (a), 1.2 m (b), and 2.5 m (c), with corresponding modeled air fractions (solid squares) using a model based on buoyancy and dissolution only. The inverse-square parameters z_1 and z_s, as obtained for the two events in Figs. 6 and 7, are shown in (d).

sumed independent of fluctuations in hydrostatic pressure. At high Peclet number [Pe] and for radii greater than approximately 20 μm, Levich (1962) found that for surface-contaminated bubbles

$$N_i = \frac{Q_i}{4\pi a D_i (C_i - C_{i\infty})} = \left(\frac{2}{\pi}\right) \mathrm{Pe}_i^{1/3}, \tag{13}$$

and

$$\mathrm{Pe}_i = \frac{a w_b}{D_i}, \tag{14}$$

where $i=1$, 2 corresponds to oxygen and nitrogen respectively, D_i is the gas diffusivity in water, Q_i is the rate of transfer of gases from the bubble surface, C_i the gas concentration in the bubble, and $C_{i\infty}$ the concentration in the water far from the bubble. The concentrations can also be expressed in terms of the coefficient of gas absorption, ϕ_i, as

$$C_i = \phi_i P_i . \tag{15}$$

From (13) to (15) equations may be derived expressing the time rate of change of bubble radius in terms of gas diffusion and changing hydrostatic pressure [i.e., Eqs. (23) and (24) in Thorpe, 1982]. The calculations are carried out for both of the primary gas species and depend upon the saturation level of each gas in the water, the surface tension, ocean surface pressure, water density and gravitational acceleration. The results from this model are shown in Figs. 8(a)–(c), where the decaying air-fraction time series are indicated by solid squares. The air fraction of the initial size distribution was scaled to agree with the data at a depth of 0.5 m. The good agreement suggests that buoyancy and dissolution indeed are the dominant processes in determining the evolution of the wake. The model results suggest that the air fraction was at least 10^3 shortly after generation, implying low-frequency sound speeds of less than 300 m/s.

V. SOUND PROPAGATION THROUGH A MOTORBOAT WAKE

A. Spectral banding

Sound that has propagated through an upper ocean bubble field has a spectrum with regularly spaced peaks, and as the field evolves, the position of these peaks changes (Farmer and Vagle, 1989). The most significant shifting is seen in the first 3 min, and then the peaks move only slightly with time. The hydrophone data were processed using 1024-point FFTs and presented in spectrogram form. Because each source pulse lasted approximately 11 s, with 64 s between pulses, only the ''source-on'' periods were acquired with a second or two taken before and after each pulse length. In Fig. 9 we see spectrograms of four acoustic transmission periods for one of the generation events at 2- and 4-m depths. These periods occurred at 0.5, 1.75, 3.0, and 4.25 min after bubble field generation, and correspond to the evolution through a single-wake injection event. In the figure these periods are concatenated together, with most of the ''off'' periods cut out [The time of each of the four transmission periods has been indicated in Fig. 6(b)]. Therefore, spectrograms of the raw data in Fig. 9 are presented merely as a means to illustrate the spectral banding that results from the waveguide propagation through the bubbly wake layer as well as to show how this banding changes with time as the bubble field degasses.

These two effects are independent of the characteristics of the receiving systems, which have a smoothly varying frequency response that will not manifest itself as banding within the spectrum. There is some evidence of vertical banding in the off periods of the spectrogram, caused by the acoustical resonators switching on and off. The visible oscillations in the horizontal bands are quasiperiodic with a period of about 2.5 s, which match the dominant period of the small inlet surface waves. Lloyd's Mirror fits to hydrophone spectra obtained between bubble injections, when the *in situ*

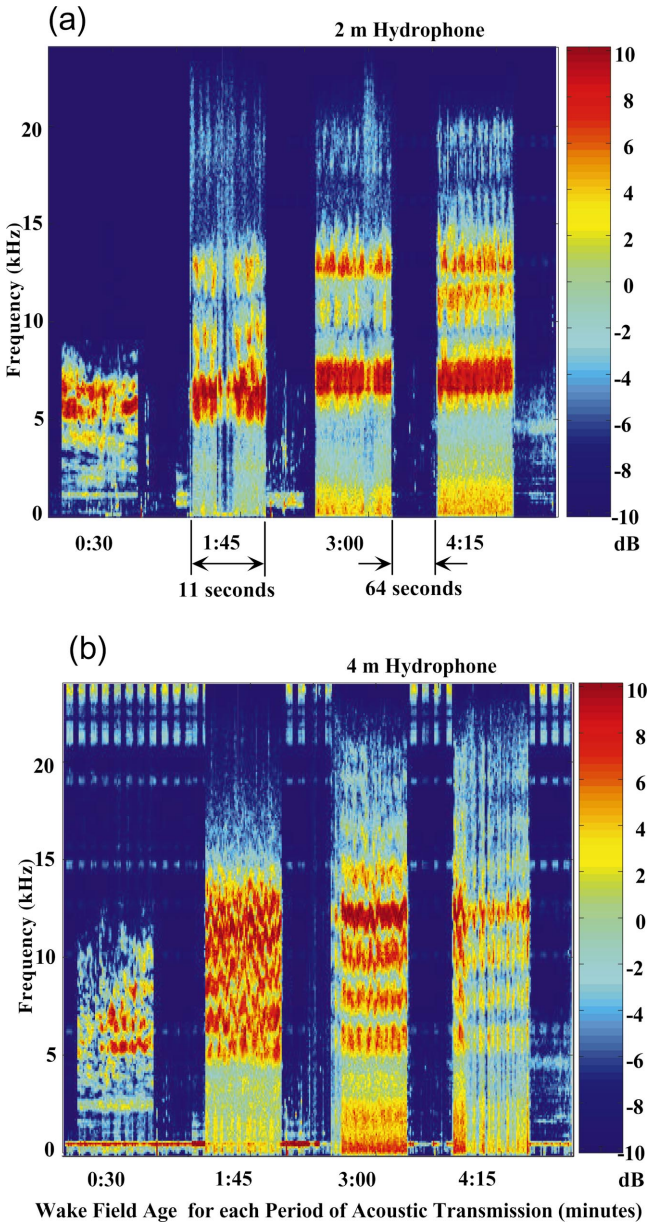


FIG. 9. Concatenated spectrograms of the first four acoustic transmissions in Fig. 6(b) from event A of the wake field at a depth of 2 m (a) and 4 m (b). Each transmission lasts about 11 s, with 64 s between each. The intensities are not calibrated.

sensors show isovelocity conditions, indicate that the oscillations are consistent with the hydrophone movement of ± 12.5 cm about some mean depth. This is in good agreement with the surface waves present during the experiments. The Lloyd's Mirror banding pattern is created by the constructive and destructive interference of the direct and reflected path from the source to the receiver in conditions without any gradients in the sound-speed profile (Burch *et al.*, 1996). Variations in the receiver depth will modulate the carrier frequency component Δf_0 (7). With the sound profile parameters used in Figs. 1–4 Δf_0 is reduced by 200 Hz if the receiver depth changes from 2 to 2.2 m and increased by 150 Hz with a depth change from 2 to 1.8 m. Small changes in the source depth, although less likely with the present geometry, will change modulation frequency component $\Delta f_0'$ (8) in a similar fashion. Any changes in the overall distance between source and receiver will only result in variations in the overall spectral level, and these changes are immaterial for the present inversion because they are not included in the inverse square model.

For both the 2- and 4-m hydrophones, we see evolution in the banding as the wake field evolves with time. In the first three transmission periods the spacing between the horizontal bands expands with time in both hydrophone records. This is expected because of the decreasing air fraction as the wake field evolves. In the first 4 min or so of the wake's evolution, rapid changes occur because the larger bubbles are leaving the water column under buoyancy forces and the very small bubbles are undergoing dissolution. Afterward, only a relatively stable bubble layer with low air-volume fraction values remains. Similar banding and spectral fluctuations are also seen in the data from the second period, labeled event B in Fig. 6(b).

A feature seen in the 2- and 4-m spectrograms is the absence of acoustic energy at frequencies above 10 kHz for the first transmission period. This absence is caused by excess frequency-dependent attenuation due to the high air-fraction values within the wake field shortly after its generation. From the resonator at a depth of 2.5 m, the bubble attenuation at 10 kHz 30 s after wake generation was calculated to be close to 0.8 dB/m. For a propagation path of 12 m, the attenuation loss due to bubbles is about 10 dB. The corresponding bubble attenuation 1 min 45 s after wake generation is only 1 dB.

B. Sound-speed profiles from spectral peaks

The inverse-square theory is an attractive approach to investigate bubble field properties because the spectral structure it predicts is related to the inverse-square sound-speed profile parameters by the two simple expressions of Eqs. (7) and (8), which lend themselves nicely to inversions. We also observe depth dependence in the banding of the wake spectra that can be represented by the inverse-square model (Fig. 9). The inverse-square theory predicts a spectral banding structure with spacing between bands that is inversely proportional to the logarithm of the receiver depth. This is the aspect of the acoustic behavior that distinguishes the inverse-square profile from the exponential and power-law profiles. Comparing the spectrograms at the 2- and 4-m sensors, we see that the horizontal banding is narrower for the deeper hydrophone (Fig. 9).

The voltage spectrum output $V(f)$ of a receiver measuring the pressure field can be considered as the combination of a source component $S(f)$, a propagation component $G(f)$, and the receive response of the system $R(f)$. The voltage output at two sensors (1 and 2) at the same range from the source and at different depths can be written as

$$V_1 = S_1(f) G_1(f) R_1(f), \tag{16}$$

and

$$V_2 = S_2(f) G_2(f) R_2(f). \tag{17}$$

When a single source insonifies the bubble field to produce the pressure field at the two hydrophones, i.e., $S_1(f) = S_2(f)$, and by assuming that both hydrophones have the

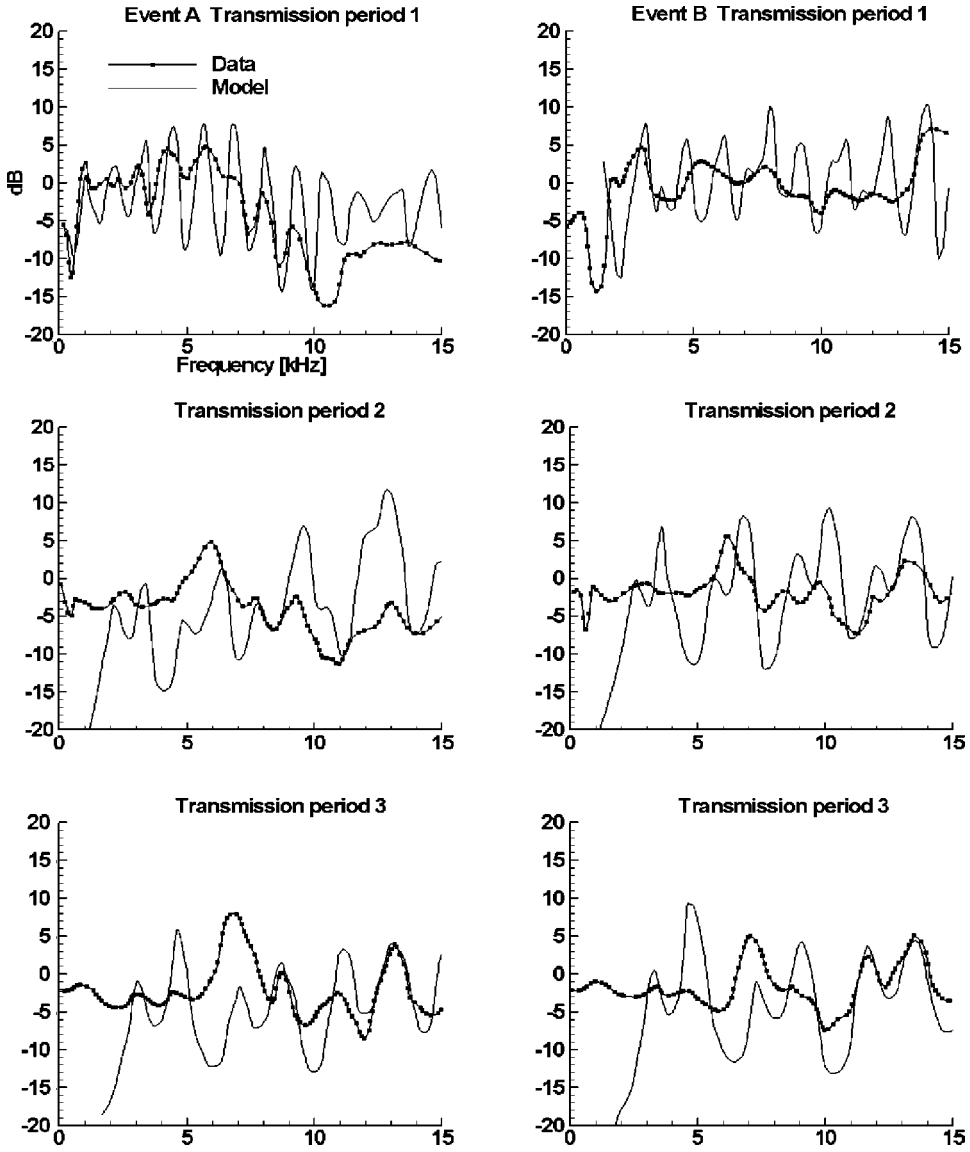


FIG. 10. Ratios of the 2- and 4-m hydrophone data (dotted lines) and best-fit inverse-square spectral ratios (solid lines). The results are shown for event A spectral periods in the left three panels and for event B periods in the right three panels.

same response $[R_1(f)=R_2(f)]$, the ratio of the two sensors becomes

$$\frac{V_1(f)}{V_2(f)}=\frac{G_1(f)}{G_2(f)}. \tag{18}$$

The spectra predicted by inverse-square theory only represent the propagation component $G(f)$. Therefore, by taking the ratio of the data at the 2- and 4-m hydrophones, no prior knowledge about the source spectrum and the receiver response is required. This ratio can be inverted using Eqs. (7) and (8) to obtain the sound-speed profile.

Inverse-square inversions were performed on the 11 transmission periods of events A and B (Fig. 6) by choosing a relatively stable portion of the banded regions in which clear data were received on both the upper and lower hydrophones simultaneously. The best inverse-square fits were determined by performing a global search over two of the inverse-square profile parameters, z_1 and z_s, and the third parameter c_w was set to 1500 m/s. The best-fit model spectrum was taken as the one that most closely matched the peaks of the data ratio. Values for the inverse-square profile parameters for the two events described here are shown in Fig. 8(d). The curvature parameter z_1 decreases close to exponentially with time, while the data suggest an increase in the offset parameter z_s with time. However, its time dependence is not that clear.

In Fig. 10, the inversions for the first three transmission periods in each event are presented. The spectral data ratio shows reduced levels at frequencies above about 10 kHz in Fig. 10 (upper left panel), indicating the increased bubble attenuation on the shallower hydrophone signal. In general there is good agreement between estimated and measured location of spectral bands.

The agreement is not so good with regard to the relative band intensity. The discrepancies may be attributed to the fact that during the evolution of the bubble field the resulting sound-speed profile is not always inverse square and the bubble field is rapidly evolving. The spectral levels are also highly sensitive to the distance between the source and the ocean surface and between the receivers and the ocean surface. However, as discussed before, the actual magnitudes are of no relevance to this particular model. The locations of observed and fitted peaks for event A transmission periods

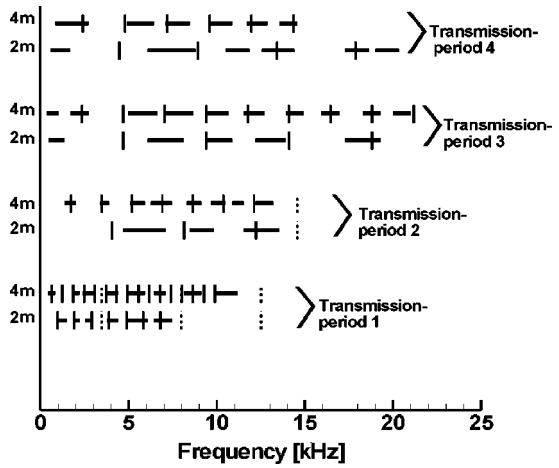


FIG. 11. Summary of the spectral banding for four event A periods (Fig. 6). Horizontal lines indicate the observed locations and widths of the different spectral peaks from the two hydrophones at 2 and 4 m. The solid vertical lines indicate locations of the modeled peaks using the inverse-square model, and the dotted vertical lines indicate the location of the corresponding fitted peaks if the exponential profile model was used.

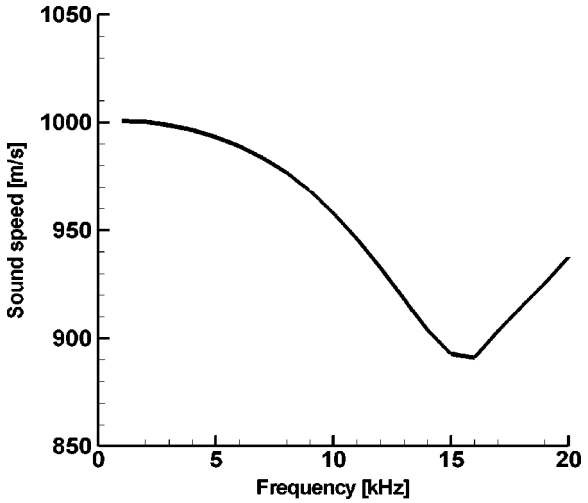


FIG. 13. Dispersion relationship for the first event A transmission period at a depth of 0.5 m [Fig. 6(a)].

have been summarized in Fig. 11. In addition to the best-fit inverse-square sound-speed profiles, the equivalent exponential sound-speed model locations have also been included for comparison.

There is generally good agreement between observed and fitted spectral bands, and the depth dependence predicted by the inverse-square model is clearly observed in the figure.

Using these inversions and Eqs. (7) and (8), best-fit inverse-square sound-speed profiles were obtained for the transmission periods in both events (Fig. 12). Similarly, Eqs. (1), (2), and (11) were used to calculate sound-speed profiles from the acoustical resonator data, averaged over the same periods. These have been plotted as symbols on the inverse-square sound-speed profiles in Fig. 12 for six of the periods with active transmission. The differences between the two estimates are shown in Figs. 12(b) and (d) for the sound-speed data in Figs. 12(a) and (c), respectively. The error is less than 2% for most of the values, and this is within the estimated error of the resonator inversions, the only exception being the values at a depth of 0.5 m in the first period [Figs. 12(a) and (b)], where the error is closer to 18%. The reason for this might be the significant dispersion immediately after wake generation. Figure 13 shows part of the dispersion curve for the first transmission period, which was obtained 30 s after the bubble injection. During this period the speed of sound varied by as much as 100 m/s between frequencies of 8 and 15 kHz. The inverse-square model, which assumes a nondispersive medium and the validity of

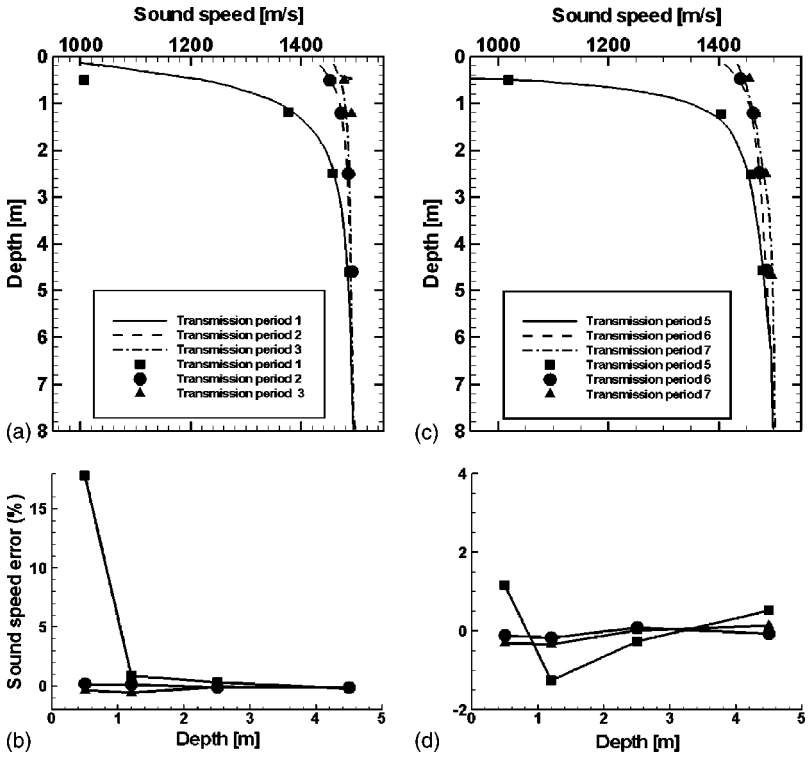


FIG. 12. Event A and event B inverse-square inversion sound-speed profiles (solid lines) and sound speeds calculated using bubble-size distributions from the acoustical resonators (symbols) [(a) and (c)]. The corresponding differences between inverse-square inversion sound-speed profiles and measured sound-speed profiles are shown in (b) and (d).

Wood's equation (1), is therefore not well suited for conditions in which the bubble concentrations are high. A reasonable threshold would be for air-volume fractions less than 10^{-5}. The agreement is therefore best at the deeper sensors where the bubble concentration is less.

VI. CONCLUDING REMARKS

Indirect and *in situ* measurements of the bubble field and associated sound-speed profile in motorboat wakes show that following the initial turbulent injection phase with very high air-volume fractions ($>10^{-3}$) the wake evolution can be explained by bubble buoyancy and dissolution only. Even though one has to be careful with direct comparisons with measurements in wakes from much larger ships, the air-fraction e-folding time constant observed here (40–60 s) compares favorably with measurements by Trevorrow *et al.* (1994), who reported a time constant of 80 s for measurements in wakes from a 40-m-long research vessel. Ezerskii *et al.* (1989) also predicted a dissolution time constant of 85 s using a bubble dispersion model similar to the one used here, including buoyancy and dissolution only. Following the initial turbulent period when the number density of bubbles is too high to be measured by our present instrumentation, the bubble-size distribution in the wakes quickly loses the larger bubbles ($a>300$ μm) and the dominant air-fraction radius decreases from about 100 μm to less than 40 μm in a few minutes. We would expect further evolution of this persistent background layer to occur on a very slow time scale, with dissolution being the important mechanism.

The wake bubble field and its associated sound-speed profile have a significant impact on the sound field for periods up to 4 min after the wake generation. Evolution was seen in the spectral structure of sound that had propagated through the waveguide, with the most significant changes observed within about the first 3 min after generation. Comparisons between the *in situ* measurements of the inverse-square theory used on sound propagating through the bubbly layer show that this theory represent a complementary technique to direct methods for measuring sound speed in the near surface. In contrast to more direct techniques, the inversion method tends to emphasize the sound speed in the background bubble layer rather than that in the high air-fraction bubble plumes beneath ocean breaking waves.

A dedicated broadband acoustic source at a known depth and range from the receivers removed one aspect of uncertainty in the inversion. However, the results give us a foundation for applying this inversion technique to the sound field within the complex, time- and spatially varying bubble field beneath ocean breaking waves, either with a controllable sound source, or ultimately by using the ambient noise from the breaking waves themselves.

ACKNOWLEDGMENTS

The authors are thankful to Dr. M. J. Buckingham and Dr. D. M. Farmer for making these experiments possible and for many interesting discussions. We would also like to thank Reece Hasanen for helping us to deploy and recover the instrumentation and for helping to run the boat.

Buckingham, M. J. (**1991**). ''On acoustic transmissions in ocean-surface waveguides,'' Philos. Trans. R. Soc. London, Ser. A **335**, 513–555.

Buckingham, M. J. (**1993**). ''Ocean acoustic propagation and ambient noise in a surface duct,'' in *Natural Physical Sources of Underwater Sound*, edited by B. R. Kerman (Kluwer Academic Publishers, Dordrecht), pp. 1–15.

Burch, H. A. (**1998**). ''Development of an Acoustic Inversion Technique for Estimating Near-Surface Sound Speed,'' Ph.D. dissertation, Scripps Institution of Oceanography.

Burch, H. A., Vagle, S., Buckingham, M. J., and Farmer, D. M. (**1996**). ''Sound propagation thorough the near-surface bubble layer,'' in Proceedings of the 3rd European Conference on Underwater Acoustics, edited by J. S. Papadakis, pp. 1161–1166.

Commander, K. C., and McDonald, R. J. (**1991**). ''Finite-element solution of the inverse problem in bubble swarm acoustics,'' J. Acoust. Soc. Am. **89**, 592–597.

Deane, G. B., and Stokes, M. D. (**1999**). ''Air entrainment processes and bubble size distributions in the surf zone,'' J. Phys. Oceanogr. **29**(7), 1393–1403.

Ezerskii, A., Sandler, B., and Selivanovskii, D. (**1989**). ''Echo-ranging observations of gas bubbles near the sea surface,'' Sov. Phys. Acoust. **35**, 483–485.

Farmer, D. M., and Vagle, S. (**1989**). ''Waveguide propagation of ambient sound in the ocean-surface bubble layer,'' J. Acoust. Soc. Am. **86**(5), 1897–1908.

Farmer, D. M., Vagle, S., and Booth, A. D. (**1998**). ''A free flooding acoustical resonator for measurement of bubble size distributions,'' J. Atmos. Ocean. Technol. **15**, 1132–1146.

Horsley, L. V. E. (**1989**). ''Modification and deployment techniques for hand-deployed arctic long-life sonobuoys,'' IEEE J. Ocean. Eng. **14**(2), 211–220.

Keeling, R. F. (**1993**). ''On the role of large bubbles in air–sea gas exchange and supersaturation in the ocean,'' J. Mar. Res. **51**, 237–271.

Lamarre, E., and Melville, W. K. (**1992**). ''Instrumentation for the measurement of void fraction in breaking waves: Laboratory and field results,'' IEEE J. Ocean. Eng. **17**(2), 204–215.

Lamarre, E., and Melville, W. K. (**1994**). ''Sound-speed measurements near the ocean surface,'' J. Acoust. Soc. Am. **96**(6), 3605–3616.

Levich, V. G. (**1962**). *Physico-chemical Hydrodynamics* (Prentice Hall, New York).

Marmarino, G. O., and Trump, C. L. (**1996**). ''Preliminary side-scan ADCP measurements across a ship's wake,'' J. Atmos. Ocean. Technol. **13**, 507–513.

Medwin, H. (**1974**). ''Acoustic fluctuations due to microbubbles in the near-surface ocean,'' J. Acoust. Soc. Am. **56**, 1100–11004.

Medwin, H., Fitzgerald, J., and Rautmann, G. (**1975**). ''Acoustic miniprobing for ocean microstructure and bubbles,'' J. Geophys. Res. **80**, 405–413.

NDRC (**1945**). ''Acoustic properties of wakes,'' in *Physics of Sound in the Sea* ed. D. 6. S. T. R. V. 8. National Defense Research Comittee reprinted 1969 as NAVMAT Rep. p-9675 (Peninsula, Los Altos, CA).

Terrill, E., and Melville, W. K. (**1997**). ''Sound speed measurements in the surface-wave layer,'' J. Acoust. Soc. Am. **102**(5), 2607–2625.

Terrill, E., and Melville, W. K. (**2000**). ''A broadband acoustic technique for measuring bubble size distributions: Laboratory and shallow water measurements,'' J. Atmos. Ocean. Technol. **17**, 220–239.

Thorpe, S. A. (**1982**). ''On the clouds of bubbles formed by breaking wind-waves in deep water, and their role in air–sea gas transfer,'' Philos. Trans. R. Soc. London, Ser. A **304**, 155–210.

Trevorrow, M. V., Vagle, S., and Farmer, D. M. (**1994**). ''Acoustical measurements of microbubbles within ship wakes,'' J. Acoust. Soc. Am. **95**(4), 1922–1930.

Vagle, S., and Farmer, D. M. (**1998**). ''A comparison of four methods for bubble size and void fraction measurements,'' IEEE J. Ocean. Eng. **23**, 211–222.

Wille, P. C., and Geyer, D. (**1984**). ''Measurements on the origin of the wind-dependent ambient noise variability in shallow water,'' J. Acoust. Soc. Am. **75**(1), 173–185.

Wood, A. B. (**1941**). *A Text Book of Sound* (Macmillan, New York).

Ye, Z. (**1997**). ''Further consideration of the waveguide propagation of ambient sound in the ocean-surface bubble layer,'' J. Acoust. Soc. Am. **102**(2), 788–797.

Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide[a)]

Karim G. Sabra, Philippe Roux, and W. A. Kuperman[b)]
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92103-0238

(Received 23 February 2004; revised 29 October 2004; accepted 29 October 2004)

Coherent deterministic arrival times can be extracted from the derivative of the time-averaged ambient noise cross-correlation function between two receivers. These coherent arrival times are related to those of the time-domain Green's function between these two receivers and have been observed experimentally in various environments and frequency range of interest (e.g., in ultrasonics, seismology, or underwater acoustics). This nonintuitive result can be demonstrated based on a simple time-domain image formulation of the noise cross-correlation function, for a uniform distribution of noise sources in a Pekeris waveguide. This image formulation determines the influence of the noise-source distribution (in range and depth) as well as the dependence on the receiver bandwidth for the arrival-time structure of the derivative of the cross-correlation function. These results are compared with previously derived formulations of the ambient noise cross-correlation function. Practical implications of these results for sea experiments are also discussed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1835507]

PACS numbers: 43.30.Nb, 43.30.Wi, 43.20.Bi [DRD] Pages: 164–174

I. INTRODUCTION

The acoustic time-domain Green's function (TDGF) fully determines the acoustic propagation timing between two points. The TDGF is usually characterized by broadcasting a known signal shape from one point and determining its modifications (in the frequency- or the time domain), after propagation through the medium, from the recorded signal at the other point. However, it has recently been experimentally demonstrated[1,2] that by cross correlating ambient noise recordings from two distinct locations, it is possible to build an estimate of the TDGF between these two receivers. Earlier conjectures of this result were also formulated.[3,4] Successful extraction of TDGF estimates between two points has recently been achieved experimentally using acoustic ambient noise measurements *only* for specific environments and measured frequency. For instance (1) ultrasonic, using measurements of diffuse noise fields;[1,2] (2) seismology, using fully diffuse ambient seismic noise;[5] (3) helioseismology[4] using passive solar Dopplergrams recorded by the Michelson Doppler imager; or (4) underwater acoustics using recording of sea surface generated ambient noise.[6] In addition, different techniques for constructing TDGF estimates, still based on cross correlations, have also been proposed. These other techniques do not directly use ambient noise measurements but instead distributed sources in the propagating medium in order to create a ''sufficiently'' diffuse acoustic field in this medium. Examples are (1) an ultrasonic scattering medium using several controlled sources;[7,8] (2) in seismology[9] using the diffuse coda of identifiable seismic events; or (3) in an underwater waveguide using controlled secondary sound sources distributed in the water column.[10]

Even if the ambient noise cross-correlation technique is applied in a different type of environment, the underlying physics relies on the same fact: when using ambient noise measurements, the deterministic coherent wavefronts extracted from time-averaged cross correlations correspond to residual coherent noise between the receivers. This small coherent component of the ambient noise field is buried in the spatially and temporally incoherent field produced by the whole noise source distribution, but emerges after time averaging from those correlations that contain noise sources whose acoustic path passes through both receivers. Thus, these coherent wavefronts yield an estimate of the arrival-time structure of the TDGF. A sufficiently long time-averaging interval (as long as environmental changes do not modify the acoustic propagation paths) and a spatially homogeneous noise distribution help in estimating the arrival-time structure of the TDGF from this correlation process.[6] The differences between the various TDGF estimates (or coherent wavefronts) obtained experimentally using the ambient noise cross-correlations technique in specific environments are linked to the spatio-temporal statistics of the noise sources. Indeed, the amplitudes of the individual extracted coherent wavefronts are shaded by the directionality of the noise sources and their spatial distribution,[6,7] and more frequent random noise events accelerate the emergence of the TDGF estimates from the ambient noise cross correlations.

In a complex environment with boundaries (e.g., oceanic waveguides, ultrasonic cavities), the emergence of an estimate of the TDGF from the long-term correlation of ambient noise recordings was demonstrated experimentally[1–6] but appears at first glance to be a nonintuitive conjecture. This result has been demonstrated theoretically in a cavity using a normal-mode formulation.[1,2] Furthermore, a stationary phase derivation has been proposed to explain the extraction of the ballistic Green's function in seismology from the correlation of multiply scattered waves in a free-space medium with embedded scatterers uniformly distributed.[11] Relationships

[a)]Portions of this work were presented at the 147th ASA meeting, Austin, TX, Nov. 2003.

[b)]Electronic mail: ksabra@mpl.ucsd.edu

between the cross correlation of the transmission response with the reflection response were also derived for an inhomogeneous elastic media.[12] The central result of this article is a demonstration of this conjecture using a straightforward application (though geometrically intricate) of the method of images for a uniform distribution of noise sources in an oceanic waveguide. This method can also be applied to other environments with plane boundaries (for instance in seismology[11]). The main goal of this article is to relate the arrival times (τ) of the time-averaged noise cross-correlation function (NCF) $C_{1,2}(\tau)$ between the two receivers (i.e., the asymptotic limit of the NCF after a long time-averaging interval) to the arrival times of the TDGF between these two receivers by using a stationary phase condition. In addition, this article provides an explanation for the amplitude shading of the individual extracted coherent wavefronts, as predicted by previous normal-mode formulations[13–15] and also observed experimentally in the ocean:[6] this shading is related to the particular geometric distribution of sea-surface noise sources (which induce a dipole effect when the noise sources are located close to the ocean surface).

Previous detailed analysis of the sources of ocean noise[16–18] (natural and manmade) indicates that high-frequency ocean noise (above several hundred hertz) is not influenced by shipping noise, and is well approximated by a homogeneous spatial distribution of random surface noise sources, with known statistics.[13,19] Thus, this article will focus on high-frequency oceanic ambient noise. In this paper, the case of a simple Pekeris waveguide[20] is studied, for which a simple formulation of the TDGF exists using the image theory: the TDGF between two points can be expressed as a sum of free-space travel times from each of the images of one of these points to the other one.[20] This image formulation is an exact solution of the wave equation for an isospeed and range-independent environment, yet it is analytically simple. It also easily includes the near-field propagation effects for noise sources close to the receivers. The resulting image formulation for the NCF $C_{1,2}(\tau)$ can be reduced to previous normal-mode formulations[13–15] or ray formulations[21] of the NCF for range-independent environments. From this simple model it follows: (1) that only noise sources located in the endfire direction of the two receivers contribute to the NCF; (2) which among these endfire noise sources actually contributes to a given arrival-time τ of $C_{1,2}(\tau)$; and (3) how the arrival times of the NCF depend on the depth of these noise sources. This simple image formulation of the NCF explains the basic physics of this noise cross-correlation process, and should provide an upper-bound estimate for the performance of this noise cross-correlation process in the case of more complex oceanic waveguides.

This paper is divided into three sections. In Sec. II, a simple model of the sea-surface-generated ambient noise is discussed and closed-form analytical results for the NCF, $C_{1,2}(\tau)$, are developed for this model. Here, the effects of the noise source depths and the influence of the bandwidth of the two receiving transducers on the time-delay structure of the measured NCF are also investigated. In Sec. III, the issues associated with a field experiment are discussed, notably the choice of the length of the correlation time window. Section IV summarizes the findings and conclusions drawn from this study.

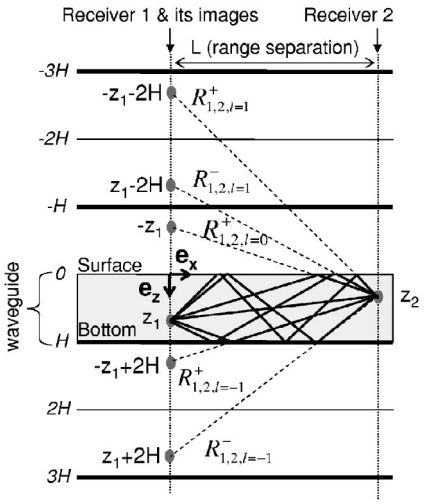


FIG. 1. Superposition of free-space propagations paths ($R^{-}_{1,2,l}$ and $R^{+}_{1,2,l}$, displayed in dashed lines) for the first five image sources solutions ($|l| \leqslant 1$) in an ideal waveguide (depth H), between a source located at $(0,0,z_1)$ and a receiver located at $(L,0,z_2)$, as defined in Eq. (1) and Eq. (2). The corresponding physical propagation paths are indicated in plain lines in the waveguide. The various depths of each image and of the two receivers are also displayed.

II. TIME-DOMAIN FORMULATION OF THE AMBIENT NOISE CROSS CORRELATION

A. Surface noise model

At frequencies above several hundred hertz, ambient noise is mostly generated at the ocean's surface.[12,17,22–24] A simple model for the surface-generated ambient noise field[13,19,25–27] is a random distribution, in space and time, of individual point noise sources located in a plane lying beneath the sea surface at a constant depth z_s. The signal recorded at the two receivers locations from a single random noise source is fully determined by the TDGF for the oceanic waveguide of interest, between this random source and the receivers. The superposition of these random pulses from many noise sources, at a receiver location, produces the measured surface-generated ambient noise.

B. Image formulation of the ambient noise cross correlation

We start with the time-domain representation of the acoustic pressure field from an impulse point source, based on the method of images,[28] in a isovelocity 3D oceanic waveguide (with a speed of sound c_0), of depth H, bounded above by a pressure-release surface and below by a semi-infinite bottom layer which reduces the pressure amplitude by a reflection coefficient V_b (see Fig. 1). V_b is a function of the ray grazing angle and can be determined by the geoacoustic properties of the bottom.[28] Furthermore, volume attenuation is included by adding an imaginary component to the speed of sound[20] ($c = c_0 - ic_i$). For the Cartesian coordinate system ($\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$) defined in Fig. 1, the depth axes, $\mathbf{e}_z$, starts from the surface and points downward. The radial vec-

tor in the horizontal plane is $\mathbf{r}=x\mathbf{e}_x+y\mathbf{e}_y$, and the horizontal axis originates at the first receiver location ($x_1=0$, $y_1=0$, z_1). The second receiver is located at ($x_2=L$, $y_2=0$, z_2), where $L>0$ by hypothesis. The finite frequency bandwidth of the recordings is $[\omega_1,\omega_2]$. To simplify notation, the negative part of the frequency spectrum of the TDGF is ignored. The bandlimited TDGF between the two sensors, $G(\mathbf{r}_1,z_1,\mathbf{r}_2,z_2;t)$, can then be expressed as[28]

$$G(\mathbf{r}_1,z_1,\mathbf{r}_2,z_2;t) = \int_{\omega_1}^{\omega_2}\frac{d\omega}{2\pi}\sum_{l=-\infty}^{l=+\infty}(-1)^{|l|}V_b^{|l|}\left(\frac{e^{i\omega(t-R^-_{1,2,l}/c_0)}\cdot e^{-\omega R^-_{1,2,l}c_i/c_0^2}}{R^-_{1,2,l}} - \frac{e^{i\omega(t-R^+_{1,2,l}/c_0)}\cdot e^{-\omega R^+_{1,2,l}c_i/c_0^2}}{R^+_{1,2,l}}\right), \tag{1}$$

where $R^-_{1,2,l}$ and $R^+_{1,2,l}$ are the free-space distances between the lth-order image (the third index l is an integer) of receiver 1 (represented by the first index) and receiver 2 (represented by the second index)

$$\begin{aligned} R^-_{1,2,l} &= \sqrt{(|\mathbf{r}_2-\mathbf{r}_1|)^2+(z_2-z_1+2lH)^2} \\ &= \sqrt{L^2+(z_2-z_1+2lH)^2}, \\ R^+_{1,2,l} &= \sqrt{(|\mathbf{r}_2-\mathbf{r}_1|)^2+(z_2+z_1+2lH)^2} \\ &= \sqrt{L^2+(z_2+z_1+2lH)^2}. \end{aligned} \tag{2}$$

This notation for the free-space distances R associated with the image theorem (here, $R^-_{1,2,l}$ and $R^+_{1,2,l}$) will be used consistently throughout the remainder of this paper: the first index (here 1) corresponds to the source location, the second index (here 2) to the receiver location, and the third index (here l) to the source-image order. The exponent ($+$ or $-$) indicates, respectively, that the sum (z_2+z_1) or difference (z_2-z_1) of source/receiver depths is used. Figure 1 represents the geometric interpretation of the image theorem for the two receivers. Thus, in Eq. (1), the arrival times of the TDGF correspond to the free-space travel times from each individual images of receiver 1 to receiver 2. For infinite bandwidth (i.e., $[\omega_1,\omega_2]=[-\infty,+\infty]$) and in the limit of no volume attenuation (i.e., $c_i=0$), the previous time-domain formulation reduces to

$$G(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1;t)^{inf} = \sum_{l=-\infty}^{l=+\infty}(-1)^{|l|}V_b^{|l|}\left(\frac{\delta(t-R^-_{1,2,l}/c_0)}{R^-_{1,2,l}} - \frac{\delta(t-R^+_{1,2,l}/c_0)}{R^+_{1,2,l}}\right). \tag{3}$$

In this 3D oceanic waveguide, the noise source distribution is modeled as an infinite 2D plane sheet of monopole impulse sources located at ($\mathbf{r}_s=x_s\mathbf{e}_x+y_s\mathbf{e}_y,z_s$), where z_s is a constant. In the absence of attenuation in the ocean, the resulting power of the recording ambient noise field would be infinite. In practice, the contribution of the noise sources is limited in range due to bottom absorption and volume attenuation. In this model, each noise source broadcasts, at a random time t_s, an impulse of amplitude $S(\mathbf{r}_s;t_s)$. The signal recorded at the two receivers, locations, from a single random source, represented by a delta function in space and time, is fully determined by the TDGF for the oceanic waveguide of interest, between this source and the receivers.[13,19,25] The time-domain recorded signals at the two receivers, $P(\mathbf{r}_s,z_s,\mathbf{r}_k,z_k;t)$, $k=1, 2$, are obtained by summing over all the impulse noise sources contributions,[28] using Eq. (1) for the Green's function between each noise source and the receivers $G(\mathbf{r}_s,z_s,\mathbf{r}_k,z_k;t-t_s)$ for $k=1, 2$

$$P(\mathbf{r}_s,z_s,\mathbf{r}_k,z_k;t) = \int_{-\infty}^{+\infty}d\mathbf{r}_s\int_{-\infty}^{+\infty}dt_s\,S(\mathbf{r}_s;t_s)G \times(\mathbf{r}_s,z_s,\mathbf{r}_k,z_k;t-t_s). \tag{4}$$

In a stationary medium, the temporal NCF, $C_{1,2}(\tau)$, between the signals recorded by both receivers is defined as

$$C_{1,2}(\tau) = \int_{-\infty}^{+\infty}dt\,P(\mathbf{r}_s,z_s,\mathbf{r}_1,z_1;t)P \times(\mathbf{r}_s,z_s,\mathbf{r}_2,z_2;t+\tau). \tag{5}$$

These random impulse noise sources have a white frequency spectrum, which does not represent the physical frequency spectrum of the sound pulse generated from breaking waves.[22,23,25] However, if the pulses induced by the breaking events are assumed to be identical on average, the source spectrum can be factored out of the spectrum of the received signal. Thus, the NCF is independent of the source spectrum when normalized by the energy of the recorded signals at each receiver location.[26]

In practice, the NCF is constructed from ensemble averages, denoted by $\langle\ \rangle$, over realizations of the noise source amplitude $S(\mathbf{r}_s;t_s)$. We assume that the random impulse noise sources have the same amplitude Q (Pa $m^{3/2}$ s), have a creation rate[29] ν ($m^{-2}\,s^{-1}$) per unit time per surface area, and are temporally and spatially incoherent[13] in the limit of infinite recording time and infinite bandwidth (due to the impulsive nature of the sources); then

$$\langle S(\mathbf{r}_s;t_s)S(\mathbf{r}_{s'};t_{s'})\rangle = 2\nu Q^2\delta(\mathbf{r}_{s'}-\mathbf{r}_s)\cdot\delta(t_{s'}-t_s). \tag{6}$$

Empirical relations between the surface wind speed and the spatial and temporal statistics of breaking waves are available,[30] and can be used to provide estimates of the parameters ν and Q for given sea-state conditions. Since the noise sources are assumed to be uncorrelated after ensemble averaging [see Eq. (6)], only the correlations between the arrival times produced by the images of the same noise source will contribute to the arrival times of the NCF. Thus, using Eqs. (1), (2), (4)–(6), the NCF can be expressed—after simplification—as a sum of four cross terms

$$C_{1,2}(\tau) = \Delta^{-,-}_{1,2} - \Delta^{+,-}_{1,2} - \Delta^{-,+}_{1,2} + \Delta^{+,+}_{1,2}, \tag{7}$$

where

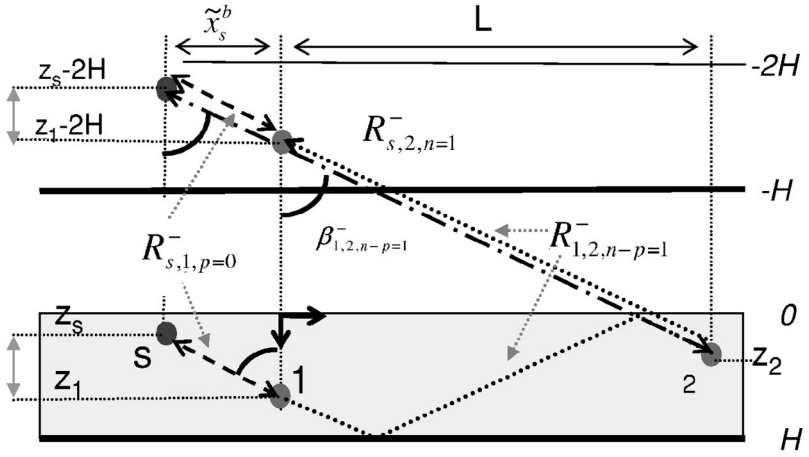


FIG. 2. Geometric construction of the location of the noise source corresponding to the first solution of the stationary phase conditions $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{b}=(\tilde{x}_s^b,\tilde{y}_s^b=0)$, for $n=1$ and $p=0$ [see Eq. (12) and Eq. (13)]. Note the alignment between the noise source image [at depth $z_s-2(n-p)H=z_s-2H$], the image of receiver 1 [at depth $z_1-2(n-p)H=z_1-2H$], and receiver 2. Hence, the difference between the free-space propagation path $R^-_{s,2,n=1}$ (dash-dotted line) and $R^-_{s,1,p=0}$ (dashed line) is equal to the free-space distance $R^-_{1,2,n-p=1}$ (dotted line), i.e., the distance between the image of receiver 1 and receiver 2 itself.

$$\Delta_{1,2}^{\sigma,\zeta} = \nu Q^2 \int_{\omega_1}^{\omega_2} \frac{d\omega}{\pi} \int_{-\infty}^{+\infty} d\mathbf{r}_s \sum_{p=-\infty}^{p=+\infty} \sum_{n=-\infty}^{n=+\infty} (-1)^{|p|+|n|} V_b^{|p|+|n|} \times \left(\frac{e^{i\omega(\tau+(R^\sigma_{s,1,p}-R^\zeta_{s,2,n})/c_0)} \cdot e^{-\omega(R^\sigma_{s,1,p}+R^\zeta_{s,2,n})c_i/c_0^2}}{R^\sigma_{s,1,p} R^\zeta_{s,2,n}} \right), \quad (8)$$

where the sign exponents symbols are [using the notation of Eq. (2)] $\sigma=\pm$ and $\zeta=\pm$. The double sum over image index p and n is associated with the particular images of the noise sources recorded by receiver 1 and receiver 2, respectively. The arrival times of the NCF, τ, are thus set by the difference of travel times $R^\sigma_{s,1,p}/c_0$ (from the noise source image of order p to receiver 1) and $R^\zeta_{s,2,n}/c_0$ (from the noise source image of order n to receiver 2). On the other hand, the volume attenuation for the amplitude associated with these arrival times depends on the sum of those travel times $-\omega(R^\sigma_{s,1,p}+R^\zeta_{s,2,n})c_i/c_0^2$ and on the accumulated amplitude damping due to bottom reflection, $V_b^{|p|+|n|}$. Thus, the contributions to the NCF of distant noise sources as well as the contributions resulting from high-order images (which follow steep propagation paths) will be attenuated quicker.

The last step in the image formulation of the NCF involves performing the horizontal spatial integration over the whole distribution of noise sources. Section II C presents the mathematical derivations for the first term only, $\Delta_{1,2}^{-,-}$ (i.e., $\sigma=\zeta=-$), of the NCF [see Eq. (7)]. The values of the three other terms can be directly computed by simple substitution, in Eq. (8), of a different pair $(R^\sigma_{s,1,p},R^\zeta_{s,2,n})$, for $\sigma=\pm$ and $\zeta=\pm$.

C. Evaluation of the spatial integration of the NCF

1. The stationary phase approximation

In the high-frequency regime, the spatial integration in Eq. (8) over the distribution of noise sources can be estimated using a stationary phase approximation[20] where the phase $\Phi_{1,2}^{-,-}(\mathbf{r}_s)$ [associated with term $\Delta_{1,2}^{-,-}$; see Eq. (8)] corresponds to

$$\Phi_{1,2}^{-,-}(\mathbf{r}_s)=\omega(\tau+(R^-_{s,1,p}-R^-_{s,2,n})/c_0). \quad (9)$$

For this method, the spatial integration can be estimated by first finding the noise source locations where $\Phi_{1,2}^{-,-}(\mathbf{r}_s)$ has vanishing derivatives [i.e., the extrema of $\Phi_{1,2}^{-,-}(\mathbf{r}_s)$], then evaluating the spatial integral in the vicinity of each of these locations by using a Taylor series expansion, and finally summing these contributions.

The particular noise source locations $[\tilde{\mathbf{r}}_\mathbf{s}=(\tilde{x}_s,\tilde{y}_s)]$ corresponding to the extrema of the phase function $\Phi_{1,2}^{-,-}(\tilde{\mathbf{r}}_\mathbf{s})$ are determined by the two spatial conditions, for a given pair of indices (n,p)

$$\tilde{y}_s=0, \quad (10)$$

$$\frac{\tilde{x}_s}{\tilde{x}_s^2+(z_1-z_s+2pH)^2}=\frac{\tilde{x}_s-L}{(\tilde{x}_s-L)^2+(z_2-z_s+2nH)^2}. \quad (11)$$

The first condition, in Eq. (10), implies that noise sources contributing to the NCF are located in the endfire direction, with respect to the two receivers. The second condition in Eq. (11) specifies a discrete set of noise sources located on the endfire line due to the boundary conditions at the interfaces of the waveguide. The remainder of this section will show that these noise sources create a time delay τ in the NCF between receiver 1 and receiver 2 related to the arrival time of the TDGF [see Eq. (1)] between these two receivers.

2. Geometric interpretation of the arrival times of the NCF derived from the stationary phase conditions

Figure 2 gives a geometric interpretation of the arrival times of the NCF derived from the stationary phase conditions [see Eq. (10) and Eq. (11)], for a given pair of image indices (n,p). The distance between the noise source and receiver 1 can be expressed as $R^-_{s,1,p}=\sqrt{|\tilde{\mathbf{r}}_\mathbf{s}|^2+((z_1+2(p-n)H)-(z_s-2nH))^2}$. Hence, $R^-_{s,1,p}$ is also the distance between the $(-n)$th image of the noise source and the $(p-n)$th image of receiver 1. Consider the noise source located at $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{b}=(\tilde{x}_s^b,\tilde{y}_s^b=0)$, such that the $(-n)$th image of this noise source, the $(p-n)$th image of receiver 1, and receiver 2 itself are all aligned. The resulting arrival time $\tau_G^{-,-}$ of the NCF is then

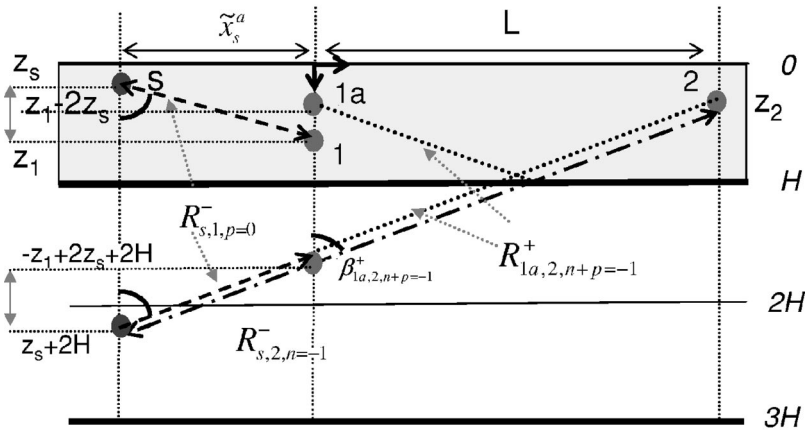


FIG. 3. Geometric construction of the location of the noise source corresponding to the second solution of the stationary phase conditions $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{a}=(\tilde{x}_s^a,\tilde{y}_s^a=0)$, for $n=-1$ and $p=0$ [see Eq. (14) and Eq. (15)]. The noise source image orders considered are $n=-1$ and $p=0$. Note the alignment between the noise source image [at depth $z_s-2(n-p)H=z_s+2H$], the image of receiver 1 [at depth $-(z_1-2z_s+2(n+p)H)=-z_1+2z_s+2H$], and receiver 2. Hence, the difference between the free-space propagation paths $R^-_{s,2,n=-1}$ (dash-dotted line) and $R^-_{s,1,p=0}$ (dashed line) is equal to the free-space distance $R^+_{1a,2,n+p=-1}$ (dotted line), i.e., the distance between the image located at a depth $-[z1-2z_s+2(n+p)H]$ and receiver 2 itself.

$$\tau_G^{-,-}=\frac{R^-_{s,2,n}-R^-_{s,1,p}}{c_0}=\pm\frac{R^-_{1,2,n-p}}{c_0}=\pm\frac{\sqrt{L^2+(z2-z_1+2(n-p)H)^2}}{c_0}. \quad (12)$$

Hence, $\tau_G^{-,-}$ is also the arrival time of the TDGF between the $(p-n)$th image of receiver 1 and receiver 2. Using this geometric construction, a closed-form expression for the noise source location $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{b}=(\tilde{x}_s^b,\tilde{y}_s^b)$ is then

$$\tilde{x}_s^b=-\operatorname{sign}(\tau)(z_1-z_s+2pH)\cdot\tan(\beta^-_{1,2,n-p});\quad \tilde{y}_s^b=0, \quad (13)$$

$$\tan(\beta^-_{1,2,n-p})=\frac{L}{z_2-z_1+2(n-p)H},$$

where $\beta^-_{1,2,n-p}$ is the angle made by the line joining the $(-n)$th image of this noise source, the $(p-n)$th image of receiver 1, and receiver 2 itself with respect to the vertical axis (see Fig. 2). Furthermore, the corresponding phase $\Phi^{-,-}_{1,2}(\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{b})$ [see Eq. (9)] reduces to $\Phi^{-,-}_{1,2}(\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{b})=\omega(\tau\pm\tau_G^{-,-})$, which is *independent* of the noise source location $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{b}$. This geometric condition on the noise source location gives an interpretation of the stationary phase conditions expressed by Eq. (10) and Eq. (11).

Similarly, Fig. 3 shows a geometric construction for a second noise source location $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{a}=(\tilde{x}_s^a,\tilde{y}_s^a=0)$ which satisfies the same stationary phase conditions for the pair of image indices (n,p). Another way to express the distance between the noise source and receiver 1 is $R^-_{s,1,p}=\sqrt{|\tilde{\mathbf{r}}_\mathbf{s}|^2+((-z_1+2z_s-2(p+n)H)-(z_s-2nH))^2}$. Introduce an extra receiver $1a$ located at $(x_1=0,\ y_1=0,\ z_1-2z_s)$, which is shifted in depth by a distance -2_zs with respect to receiver 1. In this case, $R^-_{s,1,p}$ corresponds the distance between the $(-n)$th image of the noise source and the $(p+n)$th image of receiver $1a$. If we now consider the noise source located at $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{a}$, so that the $(-n)$th image of this noise source, the $(p-n)$th image of receiver $1a$, and receiver 2 itself are all aligned, the resulting arrival time $\tau_s^{-,-}$ of the NCF is then

$$\tau_s^{-,-}=\frac{R^-_{s,2,n}-R^-_{s,1,p}}{c_0}=\pm\frac{R^+_{1a,2,n+p}}{c_0}=\pm\frac{\sqrt{L^2+(z2+(z_1-2z_s)+2(n-p)H)^2}}{c_0}. \quad (14)$$

Hence, $\tau_s^{-,-}$ is also the arrival time of the TDGF between the $(p+n)$th image of receiver $1a$ (located at a shifted depth z_1-2z_s) and receiver 2. Note that $\tau_s^{-,-}$ now depends on the value of the noise source depth z_s. Using this geometric construction, a closed-form expression for the noise source location $\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{a}=(\tilde{x}_s^a,\tilde{y}_s^a)$ is

$$\tilde{x}_s^a=-\operatorname{sign}(\tau)(z_1-z_s+2pH)\cdot\tan(\beta^+_{1a,2,n+p});\quad \tilde{y}_s^a=0, \quad (15)$$

$$\tan(\beta^+_{1,2,n+p})=\frac{L}{z_2+(z_1-2z_s)+2(n+p)H},$$

where $\beta^+_{1,2,n+p}$ is the angle made by the line joining the $(-n)$th image of this noise source, the $(p+n)$th image of receiver $1a$, and receiver 2 itself with respect to the vertical axis (see Fig. 3). The stationary phase condition now yields $\Phi^{-,-}_{1,2}(\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{a})=\omega(\tau\pm\tau_s^{-,-})$, which is *independent* of the horizontal coordinates of the noise source $(\tilde{x}_s^a,\tilde{y}_s^a)$.

Thus, the noise sources located at $(\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{b})$ and $(\tilde{\mathbf{r}}_\mathbf{s}^\mathbf{a})$, satisfy the stationary phase conditions for $\Phi^{-,-}_{1,2}$ [see Eq. (10) and Eq. (11)] and are in the endfire direction with respect to the

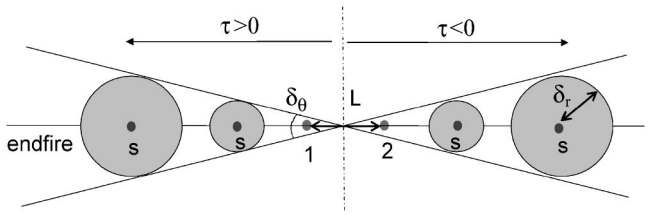


FIG. 4. Top view of the ocean surface. The noise sources making a significant contribution to the image of receiver 1 of order $n\pm p$ in $C_{1,2}(\tau)$ are located within a disk of radius δ_r [see Eq. (17)] from the noise source locations (s) determined by the stationary phase conditions [see Eqs. (10), (11)], which are on the endfire direction. δ_θ is the aperture of the cone in which all of the noise sources contributing to the image of receiver 1 of order $n\pm p$ lie. By definition of the cross-correlation function $C_{1,2}(\tau)$ [see Eq. (5)], the noise sources associated with positive arrival times are located at a range $x_s<0$, while the noise sources associated with negative arrival times are located at a range $x_s>L$.

two receivers. These noise sources contribute to two specific arrival times of the NCF: (1) $\pm\tau_G^{-,-}$, which is also an arrival time of the TDGF between receiver 1 and 2; and (2) $\pm\tau_s^{-,-}$, which is also an arrival time of the TDGF between receiver $1a$ and 2 and depends on the noise source depth z_s. The sign of these arrival times depends on the relative depth position of the noise source and the two receivers (see the Appendix for a complete discussion). Positive arrival times τ correspond to the correlation of noise sources having a range $x_s<0$ (i.e., being on the left of receiver 1 in the endfire direction). On the other hand, negative arrival times τ result from the correlation of noise sources having a range $x_s>L$ (i.e., being on the right of receiver 2, in the endfire direction) (see Fig. 4). Since the noise sources are distributed symmetrically around the two receivers' locations, the NCF, $C_{1,2}(\tau)$, for $\tau>0$ will be a mirror image of the NCF, $C_{1,2}(\tau)$, for $\tau<0$, with respect to the time origin $\tau=0$.

3. Spatial resolution of the NCF

Since the phase $\Phi_{1,2}^{-,-}(\mathbf{r}_s)$, has vanishing derivatives in the vicinity of noise source location $\widetilde{\mathbf{r}_s^b}$, it can be expressed as a second-order Taylor series (and similarly for $\widetilde{\mathbf{r}_s^a}$)

$$\Phi_{1,2}^{-,-}(\mathbf{r}_s)=\Phi_{1,2}^{-,-}(\widetilde{\mathbf{r}_s^b})+\tfrac{1}{2}(\mathbf{r}_s-\widetilde{\mathbf{r}_s^b})\cdot[(\mathbf{r}_s-\widetilde{\mathbf{r}_s^b})\cdot\nabla_{\mathbf{r}_s}(\nabla_{\mathbf{r}_s}\Phi_{1,2}^{-,-}(\mathbf{r}_s))]_{\mathbf{r}_s=\widetilde{\mathbf{r}_s^b}}. \tag{16}$$

Equation (16) can be used to evaluate the spatial integration for the partial cross correlation between the two receivers $\Delta_{1,2}^{-,-}(\tau)$ [see Eq. (8)] using the stationary phase approximation. Additionally, Eq. (16) shows that the spatial area, centered on the noise source location, $\widetilde{\mathbf{r}_s^b}$ or ($\widetilde{\mathbf{r}_s^a}$) (in the horizontal plane of the noise sources), which makes a significant contribution to $\Delta_{1,2}^{-,-}(\tau)$ is of the order of $[\nabla_{\mathbf{r}_s}(\nabla_{\mathbf{r}_s}\Phi_{1,2}^{-,-}(\mathbf{r}_s))]^{-1}$, for $\mathbf{r}_s=\widetilde{\mathbf{r}_s^a}$ or $\widetilde{\mathbf{r}_s^b}$. This quantity corresponds to a disk area centered on the noise source location $\widetilde{\mathbf{r}_s^b}$ of radius δ_r (see Fig. 4)

$$\delta_r=|\widetilde{x_s^b}|\cdot\delta_\theta,$$
$$\delta_\theta=\left(\sqrt{\frac{\omega R_{1,2,n-p}^{-}\sin(\beta_{1,2,n-p}^{-})}{c_0}}\right)^{-1}\cdot\sqrt{\frac{|\widetilde{x_s^b}-L|}{|\widetilde{x_s^b}|}}, \tag{17}$$

where δ_θ is the aperture of the cone in which the noise sources associated with the arrival times $\tau_G^{-,-}$ or $\tau_s^{-,-}$ of the NCF lie [see Eq. (12) and Eq. (14)]. When using a coordinate system centered on the location $(L/2,0,0)$ (i.e., at equal range from receiver 1 and receiver 2) instead of receiver 1, the new range coordinate for the source location $\widetilde{\mathbf{r}_s^b}$ is then shifted to $|\widetilde{x_s^b}-L/2|$ and this coordinate system is symmetrical (with respect to the two receivers) $|\widetilde{x_s^b}|=|\widetilde{x_s^b}-L|$. Similar results can be derived for the location $\widetilde{\mathbf{r}_s^a}$. Thus, in this symmetrical coordinate system, the expression of the angular aperture δ_θ is then simply inversely proportional to

$$\sqrt{\omega R_{1,2,n-p}^{-}\sin(\beta_{1,2,n-p}^{-})/c_0}$$
$$(\text{or }\sqrt{\omega R_{1a,2,n+p}^{+}\sin(\beta_{1a,2,n+p}^{+})/c_0}).$$

The angular aperture δ_θ determines the amount of noise sources contributing to the arrival times $\tau_G^{-,-}$ or $\tau_s^{-,-}$ of the NCF. The parameter δ_θ decreases with the image order $n\pm p$ for the noise source images considered due to the factor $\sin(\beta_{1,2,n-p}^{-})$ [or $\sin(\beta_{1a,2,n+p}^{+})$], and is inversely proportional to the product of the wave number (ω/c_0) times the free-space distance $R_{1,2,n-p}^{-}$ (or $R_{1a,2,n+p}^{+}$). The parameter δ_θ can be interpreted as a directivity pattern of the NCF, which acts as a beamforming process on the endfire direction for the homogeneous horizontal noise source distributions. Equation (17) then appears as a generalization of previous free-space results[6] to the case of an oceanic waveguide, and includes now the dependency on the image order $n\pm p$ and waveguide geometry.

4. Final formulation of the NCF and its time derivative

Using the results of Secs. II C 1 and II C 2 and noting that $d(e^{i\omega\tau})/d\tau=i\omega e^{i\omega\tau}$, the derivative of the partial cross correlation between the two receivers $\Delta_{1,2}^{-,-}(\tau)$, [see Eq. (2) and Eq. (8)] can then be simplified to

$$\frac{d\Delta_{1,2}^{-,-}}{d\tau}=\nu Q^2 c_0\int_{\omega_1}^{\omega_2}d\omega\sum_{p=-\infty}^{p=+\infty}\sum_{n=-\infty}^{n=+\infty}(-1)^{|p|+|n|}V_b^{|p|+|n|}\left(\frac{e^{i\omega(\tau\pm R_{1,2,n-p}^{-}/c_0)}\cdot e^{-\omega(|\widetilde{x_s^b}|+|\widetilde{x_s^b}-L|)/\sin(\beta_{1,2,n-p}^{-})\cdot c_i/c_0^2}}{R_{1,2,n-p}^{-}}\right.$$
$$\left.+\frac{e^{i\omega(\tau\pm R_{1a,2,n+p}^{+}/c_0)}\cdot e^{-\omega(|\widetilde{x_s^a}|+|\widetilde{x_s^a}-L|)/\sin(\beta_{1a,2,n+p}^{+})\cdot c_i/c_0^2}}{R_{1a,2,n+p}^{+}}\right). \tag{18}$$

In this geometry, the $(-p)$th and $(-n)$th images of this noise source will contribute to a pair of arrival times of the NCF, determined by the stationary phase conditions [see Eq. (12) and Eq. (14)]: $\tau_G^{-,-}=\pm R_{1,2,n-p}^{-}/c_0$ ($\tau_s^{-,-}=\pm R_{1a,2,n+p}^{+}/c_0$) for the temporal derivative of the NCF $d\Delta_{1,2}^{-,-}/d\tau$. The arrival time $\tau_s^{-,-}$ depends on the noise source depth z_s and is close to the arrival times of the TDGF between receiver 1 and 2 since $R_{1a,2,n+p}^{+}/c_0\approx R_{1,2,n+p}^{+}/c_0$ when z_s is small (i.e., $z_1-2z_s\approx z_1$). Similar relationships between the time-arrival structure of the time derivative of the NCF and the TDGF have also been derived for ultrasonic cavities.[1,2]

The three other cross terms $\Delta_{1,2}^{\sigma,\zeta}$, for $\sigma=\pm$ and $\zeta=\pm$-as defined by Eq. (8), of the NCF $C_{1,2}(\tau)$ [see Eq. (5)], can be computed in a similar way to the previous derivation for $\Delta_{1,2}^{-,-}(\tau)$. Based on the same stationary phase arguments, each time derivative of these cross terms yields a pair of arrival times $(\tau_G^{\pm,\pm},\tau_s^{\pm,\pm})$. Figure 5 illustrates the resulting arrival-time structure of the NCF for a given value of the image order $n\pm p$. Thus, it appears that some of the arrival times of the NCF exactly match the arrival times of the TDGF between receiver 1 and 2: $\tau_G=\sqrt{L^2+(z2\pm z1+2(n\pm p)H)^2}/c_0$ [note that the exponent $(^-,^-)$ has been dropped since we are discussing now the complete arrival-time structure of the NCF $C_{1,2}(\tau)$ and no longer $\Delta_{1,2}^{-,-}$ alone]. But, additional times delays also exist, whose values depend on the noise source depth z_s: $\tau_s=\sqrt{L^2+(z2\pm z1\pm 2z_s+2(n\pm p)H)^2}/c_0$ [see Eq. (14)]. The amplitudes corresponding to the arrival times τ_G and τ_s have opposite phase [see Eq. (7)].

Similar results can be derived based on earlier modal formulations of the NCF (Ref. 13) for the case of the Pekeris waveguide: the dependency on the noise source depth z_s appears explicitly in the arrival times by transforming the modal sum into a infinite sum of images.[31] Furthermore, similar dependency on the noise source depth [through the term $\pm 2z_s$; see Eq. (19)] for the vertical coherence of sea-surface ambient noise has been previously reported.[26] The expression depends on various combinations depth differences $z_2\pm z_1$ and $z_2\pm z_1\pm 2z_s$.

D. Simplified image formulation of the NCF

This section simplifies this result for the case of (1) no volume attenuation; or (2) a small source depth z_s (dipole expansion). A direct interpretation of the arrival-time structure of the derivative of the time-averaged NCF then becomes apparent.

1. Case of no volume attenuation

In the limit of no volume attenuation ($c_i=0$), the expression for the time derivative of the NCF $dC_{1,2}(\tau)/d\tau$ [see Eq. (7) and Eq. (18)] can be expressed simply as, using the notation of Eq. (1) and Eq. (2)

$$\frac{dC_{1,2}(\tau)}{d\tau}=\nu Q^2 c_0\int_{\omega_1}^{\omega_2}d\omega\sum_{p=-\infty}^{p=+\infty}\sum_{n=-\infty}^{n=+\infty}(-1)^{|p|+|n|}V_b^{|p|+|n|}[2(G_{n-p}^{-}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1;\tau)-G_{n+p}^{+}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1;\tau))$$
$$-(G_{n-p}^{-}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1-2z_s;\tau)-G_{n+p}^{+}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1-2z_s;\tau))-(G_{n-p}^{-}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1+2z_s;\tau)$$
$$-G_{n+p}^{+}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1+2z_s;\tau))], \tag{19}$$

where

$$G_l^{-}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1;t)=\frac{e^{i\omega(t-R_{1,2,l}^{-}/c_0)}}{R_{1,2,l}^{-}},$$
$$G_l^{+}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1;t)=\frac{e^{i\omega(t-R_{1,2,l}^{+}/c_0)}}{R_{1,2,l}^{+}}. \tag{20}$$

The derivative of the NCF $dC_{1,2}(\tau)/d\tau$ [see Eq. (19)] can then be related to three TDGF: (1) $G_{n\pm p}^{\pm}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1;\tau)$, the TDGF between receiver 2 and receiver 1 itself which has double amplitude; (2) $G_{n\pm p}^{\pm}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1-2z_s;\tau)$, the TDGF between receiver 2 and a receiver (referred to as receiver $1a$ in Sec. II C 2) at a location $(\mathbf{r}_1,z_1-2z_s)$ shifted in depth with respect to receiver 1; and (3) $G_{n\pm p}^{\pm}(\mathbf{r}_2,z_2,\mathbf{r}_1,z_1+2z_s;\tau)$, the TDGF between receiver 2 and a receiver at a location $(\mathbf{r}_1,z_1+2z_s)$ also shifted in depth with respect to receiver 1.

2. Dipole expansion

When the noise source depth z_s is very small compared to the depth of the two receivers, the source and its surface-reflected image (at a depth $-z_s$) acts as a dipole. The dipole radiation pattern can be derived from the last expression of the noise cross correlation [see Eq. (19)] using second-order expansion in terms of z_s

$$\frac{dC_{1,2}^{\mathrm{Dipole}}(\tau)}{d\tau}=2z_s^2\nu Q^2 c_0\int_{\omega_1}^{\omega_2}\frac{d\omega}{2\pi}\sum_{p=-\infty}^{p=+\infty}\sum_{n=-\infty}^{n=+\infty}(-1)^{|p|+|n|}V_b^{|p|+|n|}\frac{\omega}{c_0}\left(\frac{e^{i\omega(\tau\pm R_{1,2,n-p}^{-}/c_0)}\cdot e^{-\omega(|\widetilde{x_s^b}|+|\widetilde{x_s^b}-L|)/\sin(\beta_{1,2,n-p}^{-})\cdot c_i/c_0^2}}{R_{1,2,n-p}^{-}}\right.$$
$$\left.\times\cos^2(\beta_{1,2,n-p}^{-})-\frac{e^{i\omega(\tau\pm R_{1,2,n+p}^{+}/c_0)}\cdot e^{-\omega(|\widetilde{x_s^a}|+|\widetilde{x_s^a}-L|)/\sin(\beta_{1,2,n+p}^{+})\cdot c_i/c_0^2}}{R_{1,2,n+p}^{+}}\cos^2(\beta_{1,2,n+p}^{+})\right), \tag{21}$$

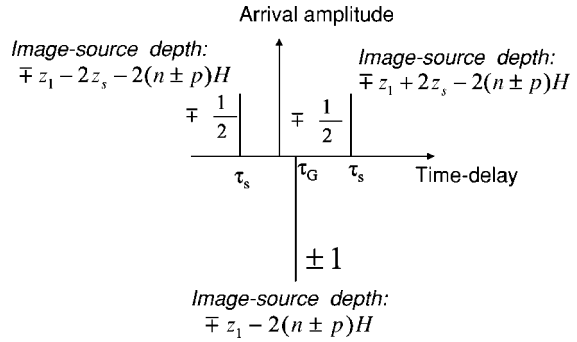


FIG. 5. Schematic of the structure of the time-delay arrivals of the noise cross correlation $C_{1,2}(\tau)$ for a given value of the image indices $n \pm p$. The three correlation arrival times can be related to the superimposed arrival-time structure of three TDGFs between sources located at (1) a depth z_1 (i.e., receiver 1, corresponding to a TDGF arrival τ_G) or (2) $z_1 \pm 2z_s$ (i.e., two distinct locations shifted in depth from the location of receiver 1, corresponding to a TDGF arrival τ_s) and a receiver located at depth z_2 (receiver 2). The depth of the images of the three equivalent sources are indicated above each of the three corresponding arrival times of the NCF. The amplitudes associated with the two arrival times τ_s are equal and are half the amplitude associated with the time delay τ_G as well as having opposite sign. Thus, for $z_s \to 0$, these three arrival times merge and yield a dipole radiation pattern.

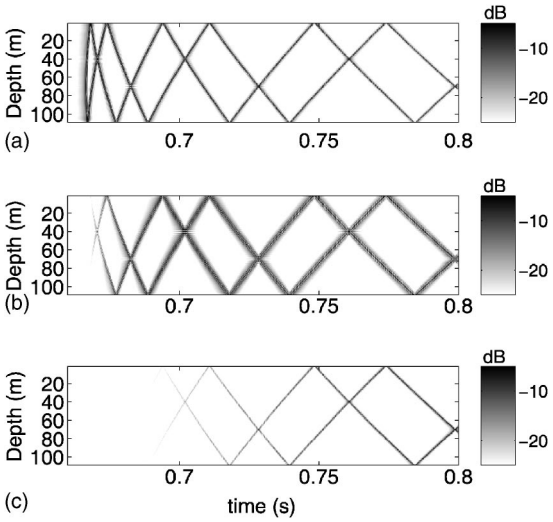


FIG. 6. Positive time-delay wavefronts of $C_{1,2}(\tau)$, for varying depth of receiver 2 in the water column, in the case of a simple oceanic waveguide ($H = 110$ m, $L = 1$ km, $z1 = 70$ m) for both $z_s = 1$ m and $z_s = 0.1$ m [(b) and (c)]. It also provides a comparison with the TDGF of this waveguide [(a)]. The bandwidth is [$\omega_1 = 1$ kHz, $\omega_2 = 3$ kHz].

where the expressions of $\beta^+_{1a,2,n+p} = \beta^+_{1,2,n+p}$ and $R^+_{1a,2,n+p} = R^+_{1,2,n+p}$ were evaluated for $z_s = 0$. The three arrival times of the NCF represented in Fig. 5 merge for $z_s \to 0$, thus creating a dipole effect. The associated dipole radiation pattern [represented by the factors $\cos^2(\beta^{n \pm p}_{1,2,n \pm p})$ in Eq. (21)] yields a stronger amplitudes for the arrival times of the NCF associated with a high image order $(n \pm p)$.

The following section discusses the influence of the source depth z_s and the effective receiver bandwidth ($[\omega_1, \omega_2]$) on the cross-arrival times of the NCF which can be effectively measured.

E. Influence of the noise source depths and the receiver's frequency bandwidth

The previous sections showed how the distribution of the various arrival times measured by the time derivative of the NCF $C_{1,2}(\tau)$ depends, among others factors, on the noise source depth z_s [see Fig. 5, Eq. (18), and Eq. (19)]. In practice, the determination of the depth z_s of the acoustically active bubbles from a breaking event is a complex problem (since it may depend on the bubble size, density, or acoustic frequency of interest) and is an ongoing research topic.[24,25] For instance, this depth was estimated to be 1.5 m, based on inversion of wave-breaking sound measured at wind speed of 10 m/s or greater in the sea-surface bubble layer.[26]

Only the arrival times, $\tau_s = \sqrt{L^2 + (z_2 \pm z_1 \pm 2z_s + 2(n \pm p)H)^2}/c_0$ of the time derivative of the cross correlation $dC_{1,2}(\tau)/d\tau$, that do not exactly match the arrival times of the TDGF [see Eq. (1)], depend on the spatial distribution of z_s (i.e., the depth dependency of the acoustically active noise bubbles). Note that the notation τ_s represent both depth shifts $\pm 2z_s$. On the other hand, the arrival times $\tau_G = \sqrt{L^2 + (z_2 \pm z_1 + 2(n \pm p)H)^2}/c_0$ of $dC_{1,2}(\tau)/d\tau$ that are independent of the value z_s for each noise source considered correspond exactly to an arrival time of the TDGF. Thus, for a given value of the image order $n \pm p$ (i.e., a given set of noise sources), the arrival times τ_s should be spread around the constant time delay τ_G if the variations of z_s are large. Furthermore, the amplitudes of the signals occurring at these arrival times τ_s should be much weaker than the amplitude of the signals occurring at the arrival times τ_G since these last ones result from the coherent sum of all the random noise source amplitudes. Consequently, if the random noise sources z_s are widely distributed in the water column, the high signal amplitude of the temporal derivative of the NCF $dC_{1,2}(\tau)/d\tau$ should occur at arrival times τ_G and thus the temporal derivative of the NCF would better match the structure of the TDGF. A similar effect was also predicted by an earlier study of the cross correlation between a pair of receivers, in the presence of a set of controlled sound sources located across the water column depth.[9]

For a given value of the noise source depth z_s and of the image indices $n \pm p$, the ability to discriminate between the arrival times τ_s and τ_G is set by the temporal resolution ΔT of the cross-correlation process [i.e., here the inverse of the frequency bandwidth $\Delta T = 2\pi/(\omega_2 - \omega_1)$]. For sufficiently large bandwidth, i.e., if $\Delta T < |\tau_G - \tau_s|$, the arrival-time structure of $dC_{1,2}(\tau)/d\tau$ can be resolved. Figure 6 shows wavefronts of $dC_{1,2}(\tau)/d\tau$, occurring at positive arrival times, for varying depth of receiver 2 in the water column, in the case of a simple oceanic waveguide ($H = 110$ m, $L = 1$ km, $z_1 = 70$ m) for both $z_s = 1$ m and $z_s = 0.1$ m [Figs. 6(b) and (c)]. It also provides a comparison with the TDGF of this waveguide [Fig. 6(a)]. Here, the frequency bandwidth is from 1 to 3 kHz; thus, $\Delta T \approx 0.5$ ms. For $|n \pm p| > 2$ and $z_s = 1$ m, the various wavefronts for $dC_{1,2}(\tau)/d\tau$ can be resolved. They are thickened compared to the wavefronts of the TDGF since they correspond to the merging of three different arrival times τ_G and τ_s [compare Figs. 6(a) and (b)].

In the opposite case, if $\Delta T > |\tau_G - \tau_s|$, the temporal resolution is then too coarse and does not allow separation of

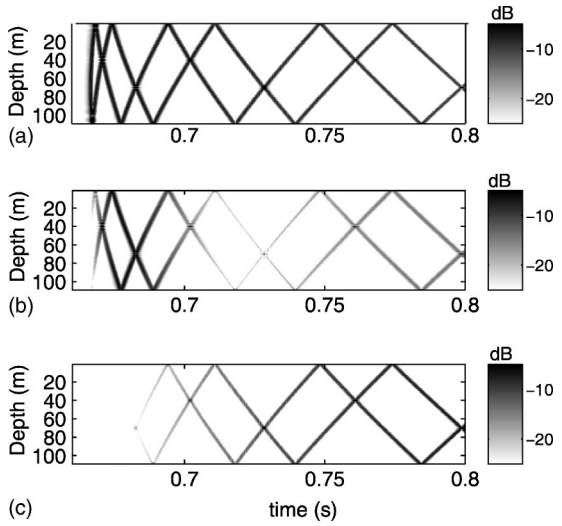


FIG. 7. Same parameters as Fig. 6 but using a previous normal-mode formulation (Ref. 13). Note the similarity of the arrival-time structure obtained with the ray formulations in Fig. 6.

the signals occurring at these time delays τ_G and τ_s. Furthermore, the signals occurring at arrival times τ_s average down with the signals occurring at τ_G since they are of opposite phase [see Eq. (19)]. In practice, the separation between τ_G and τ_s decreases with decreasing image order $n \pm p$. Thus, the amplitudes of the signals occurring at early arrival times τ_G should then be weaker compared to subsequent arrivals. This corresponds to a dipole effect as predicted by the asymptotic formulas for small values of z_s in Eq. (21) [see Fig. 6(c), for $z_s = 0.1$ m].

Figure 7 displays the wavefront structure of the TDGF and the noise cross correlation $dC_{1,2}(\tau)/d\tau$ using a previous normal-mode formulation,[13] for the same parameters as in Fig. 6. The good agreement between Fig. 6 and Fig. 7 confirms the equivalence between the ray formulation described in this article and previous normal-mode formulations. The temporal structure is identical but the discrepancies in amplitude are due to (1) the difference in modeling volume and bottom attenuation; and (2) the use of a finite number of images as well as a finite number of modes.

III. PRACTICAL CONSIDERATIONS FOR SEA EXPERIMENTS

The successful implementation of the ambient noise cross-correlations process to extract coherent wavefronts will be mainly determined by three factors: (1) hardware configuration (see below); (2) the choice of the duration of the ambient noise recordings used for the correlation process; and (3) the spatio-temporal distribution of the ambient noise sources in the environment surrounding the receivers. The following discussion suggests optimal conditions for these three factors.

First, the noise cross-correlation process requires the use of time-synchronized receivers and the absence of relative motions between them, in order to precisely determine the correlation arrival times. Moored or bottom-mounted arrays, having a common time clock for each of the array elements recordings, should provide an efficient hardware configuration. Additionally, transducers having a flat frequency response over a large bandwidth would allow one to finely separate the various arrival times present in the noise cross correlations, even for noise source depths located relatively close to the sea surface (see Sec. II E).

Second, the choice of the recording time (T_{rec}, which determines the length of the cross-correlations window) is driven by the need to record a sufficient number of uncorrelated noise events in order to extract all the different paths of the TDGF from the noise cross-correlation process. Thus, the recording time T_{rec} should be larger than both (1) the dispersion time of the sound channel (T_{disp}, which corresponds to temporal spreading of the TDGF between the two receivers and which is also referred to as the break time of the system[1]); and (2) the statistical time T_{stat}. Under the ergodic assumption, the noise events are considered as realizations of a wide sense stationary stochastic process for which corresponding time and ensemble averages of this process are equivalent. In this case, T_{stat} is then the averaging time required to converge towards a homogeneous distribution of spatially and temporally uncorrelated noise sources [see Eq. (6)]. Any two particular noise events may have significant peaks in their cross-correlation function due to sidelobes and other correlated parts of their waveforms. Thus, T_{stat} depends on the particular characteristics of the noise source process and is usually an unknown parameter, but may be estimated using previous knowledge of the environment. Finally, the formulation of the cross-correlation process, as defined by Eq. (5), is valid for a stationary environment. Thus, T_{rec} should be smaller than the fluctuation's time scale of the environment T_{fluc} (such as currents or tidal periods). Otherwise, an estimate of noise cross correlation can be derived by averaging over several small correlations windows (such that $T_{\text{rec}} < T_{\text{fluc}}$ for each correlation window). However, in the case of sea-surface motion and long averaging time, it is likely that $T_{\text{rec}} > T_{\text{fluc}}$. But, the main effect of sea-surface motion should only be to prevent the emergence of acoustic paths associated with sea-surface reflections in the estimate of the time-domain Green's function extracted from the noise cross-correlation function due to the destructive interference occurring during the long-time averaging. Overall, the practical use of the noise cross-correlation process should be optimal for the following order of the various time scales of the problem: $T_{\text{disp}} < T_{\text{stat}} < T_{\text{rec}} < T_{\text{fluc}}$.

Third, a random homogeneous distribution, in space and time, of uncorrelated noise point sources was assumed when deriving an analytical expression of NCF in Sec. II [see Eq. (6)]. The effect of the noise source depth z_s distribution can be clearly identified and recognized from the measured cross correlation (see Sec. II C.). The approximation of the noise sources being uncorrelated should hold for sufficiently long averaging times. However, the noise distribution may not be homogeneous in time or in space. For instance, temporal inhomogeneous distribution may result from some loud (compared to the average ambient noise level) noise events occurring at certain random times (e.g., a ship passing by, or an accidental burst of sound). The effectiveness of temporal and frequency equalization to remove high-energy single events (which differ from the homogeneous noise level) from

the noise recordings (previous to performing the cross-correlations) needs to be assessed. Additionally, a spatial anisotropic noise source distribution may be detrimental if there is no, or only a few, noise source(s) close to endfire direction of the two receivers. Indeed, these noise sources contribute, on average, to the arrival times of the NCF related to the arrival times of TDGF between these two receivers [see Eq. (10) and Eq. (11)]. It might also be the case that the bottom is not flat and is not equally absorbing in all spatial direction in the vicinity of a field test, thus leading to a specific directionality of the noise field. Thus, a careful study of the spatio-temporal distribution of the noise sources, in the environment of interest, is critical for explaining the performance of the noise cross-correlation process. On the other hand, when the extraction of coherent wavefronts from the noise cross-correlation process is successful, one may then attempt to invert for the spatio-temporal distributions of these noise sources.

IV. SUMMARY AND CONCLUSIONS

We have demonstrated theoretically how the arrival-time structure of the time-domain Green's function (TDGF) between two points can be estimated by cross correlating the ambient noise recorded at these two points, based on a model of surface-generated noise in the ocean. The ambient noise cross-correlation function (NCF) $C_{1,2}(\tau)$ obtained after time averaging is dominated by the correlations that contain noise sources whose acoustic waves pass through both receivers, thus allowing the extraction of coherent wavefronts and corresponding arrival times. For the case of a homogeneous noise distribution, the spatial locations of the noise sources (all located in the vicinity of the endfire direction of the two receivers) contributing to a given coherent arrival time of the time derivative of the NCF were determined. This result provides a physical explanation of the nonintuitive relationship between the NCF and the TDGF. In particular, the relation between the arrivals times of the TDGF and these coherent arrival times of the time derivative of the NCF were established given the parameters of the oceanic waveguide (water depth, bottom and volume attenuation), the receiver (horizontal and vertical) separations, the spatial and temporal noise statistics parameters (e.g., depths of the noise sources, noise events creation rate), and the frequency bandwidth of interest. This simple image formulation of the NCF also explains the amplitude shading of the individual extracted coherent wavefronts, as observed experimentally in the ocean.[6] Thus, these theoretical results should be applicable in practice for the case of simple, range-independent, shallow-water waveguides, and may provide an upper-bound estimate for the performance of this noise cross-correlation process in more complex oceanic environments and noise sources distribution.

Ambient noise cross correlations may provide a mean to characterize the oceanic environment and its fluctuations (via the relationship between the arrival-time structure of the NCF and the TDGF) with little signal processing involved. The results presented in this article should be useful for interpreting the outcome of the noise cross-correlation process for other types of ambient noise mechanisms (e.g., biological noise, shipping noise). High-frequency experiments could be used to determine more specifically the relationships between the spatial and temporal noise statistics of the ambient noise field of interest and the emergence of these coherent wavefronts from the NCF (e.g., shading of these wavefronts with respect to the actual TDGF wavefronts, convergence time towards the TDGF). Those relationships could also be used to monitor the evolution of the noise sources in the environment.

ACKNOWLEDGMENT

The research was sponsored by the Ocean Acoustics Program of the Office of Naval Research.

APPENDIX: SIGN OF THE ARRIVAL TIMES OF THE NCF DETERMINED BY THE STATIONARY PHASE CONDITIONS

For the case of a Pekeris waveguide, a geometric construction of the noise source locations ($\widetilde{\mathbf{r}}_s^{\mathbf{a}}$ and $\widetilde{\mathbf{r}}_s^{\mathbf{b}}$) satisfying the stationary phase conditions Eq. (10) and Eq. (11) was described in Secs. II C 1 and II C 2. This Appendix specifies the exact sign of the corresponding arrival times $\tau_G^{-,-} = \pm R^{-}_{1,2,n-p}/c_0$ and $\tau_s^{-,-} = \pm R^{+}_{1a,2,n+p}/c_0$ of the NCF. Their sign is a function of the relative position in depth of the noise source and the two receivers in order to satisfy Eq. (11). The stationary phase $\Phi^{-,-}_{1,2}(\mathbf{r}_s)$ at the locations ($\widetilde{\mathbf{r}}_s^{\mathbf{b}}$) or ($\widetilde{\mathbf{r}}_s^{\mathbf{a}}$) is then

$$\begin{aligned}
\Phi^{-,-}_{1,2}(\widetilde{\mathbf{r}}_s^{\mathbf{b}}) &= \omega(\tau - \mathrm{sign}(z_2 - z_1 + 2(n-p)H) \\
&\quad \cdot R^{-}_{1,2,n-p}/c_0), \quad \text{for } (z_2 - z_s + 2nH) > 0 \\
&\quad \text{and } (z_1 - z_s + 2pH) > 0; \\
\Phi^{-,-}_{1,2}(\widetilde{\mathbf{r}}_s^{\mathbf{b}}) &= \omega(\tau + \mathrm{sign}(z_2 - z_1 + 2(n-p)H) \\
&\quad \cdot R^{-}_{1,2,n-p}/c_0), \quad \text{for } (z_2 - z_s + 2nH) < 0 \\
&\quad \text{and } (z_1 - z_s + 2pH) < 0;
\end{aligned} \tag{A1}$$

and

$$\begin{aligned}
\Phi^{-,-}_{1,2}(\widetilde{\mathbf{r}}_s^{\mathbf{a}}) &= \omega(\tau + \mathrm{sign}(z_2 + z_1 - 2z_s + 2(n+p)H) \\
&\quad \cdot R^{+}_{1a,2,n+p}/c_0), \quad \text{for } (z_2 - z_s + 2nH) < 0 \\
&\quad \text{and } (z_1 - z_s + 2pH) > 0; \\
\Phi^{-,-}_{1,2}(\widetilde{\mathbf{r}}_s^{\mathbf{a}}) &= \omega(\tau - \mathrm{sign}(z_2 + z_1 - 2z_s + 2(n+p)H) \\
&\quad \cdot R^{+}_{1a,2,n+p}/c_0), \quad \text{for } (z_2 - z_s + 2nH) > 0 \\
&\quad \text{and } (z_1 - z_s + 2pH) < 0.
\end{aligned} \tag{A2}$$

[1] O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[2] R. L. Weaver and O. I. Lobkis, "Elastic wave thermal fluctuations, ultrasonic waveforms by correlation of thermal phonons," J. Acoust. Soc. Am. **113**, 2611–2621 (2003).
[3] J. Rickett and J. Claerbout, "Passive seismic imaging applied to synthetic data," SEP **92**, 83–90 (1996).
[4] J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral fac-

torization: Helioseismology and reservoir monitoring,'' The Leading Edge **18**, 957–960 (1999).
[5] N. M. Shapiro and M. Campillo, ''Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise,'' Geophys. Res. Lett. **31**, L07614 (2004).
[6] P. Roux, W. A. Kuperman, and the NPAL Group, ''Extracting coherent wavefronts from acoustic ambient noise in the ocean,'' J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[7] A. Derode, E. Larose, M. Tanter, J. Rosny, A. Tourin, M. Campillo, and M. Fink, ''Recovering the Green's function from field-field correlations in an open scattering medium,'' J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[8] A. Derode, E. Larose, M. Campillo, and M. Fink, ''How to estimate the Green's function of a heterogeneous medium between two passive sensors,'' Appl. Phys. Lett. **83**, 3054–3056 (2003).
[9] M. Campillo and A. Paul, ''Long-range correlations in the diffuse seismic coda,'' Science **299**, 547–549 (2003).
[10] P. Roux and M. Fink, ''Green's function estimation using secondary sources in a shallow water environment,'' J. Acoust. Soc. Am. **113**, 1406–1416 (2003).
[11] R. Snieder, ''Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase,'' Phys. Rev. E **69**, 046610 (2004).
[12] K. Wapenaar, J. Thorbecke, and D. Draganov, ''Relations between reflection and transmission responses of three-dimensional inhomogeneous media,'' Geophys. J. Int. **156**, 179–194 (2004).
[13] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[14] M. J. Buckingham, ''Spatial coherence of wind-generated noise in a shallow ocean channel,'' J. Acoust. Soc. Am. **70**, 1412–1420 (1981).
[15] T. C. Yang and K. Yoo, ''Modeling the environmental influence on the vertical directionality of ambient noise in shallow water,'' J. Acoust. Soc. Am. **101**, 2541–2554 (1997).
[16] G. M. Wenz, ''Acoustic ambient noise in the ocean: Spectra and sources,'' J. Acoust. Soc. Am. **34**, 1936–1956 (1962).
[17] R. K. Andrew, B. M. Howe, J. M. Mercer, and M. Dzieciu, ''Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast,'' ARLO **3**, 65 (2002). In this article, Wenzs results have been compared to more recent data.
[18] R. J. Urick, *Ambient Noise in the Sea* (Peninsula, Los Altos, CA, 1986).
[19] M. J. Buckingham, ''A theoretical model of ambient noise in a low-loss shallow water channel,'' J. Acoust. Soc. Am. **67**, 1186–1192 (1980).
[20] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994), pp. 33–34, 100–101, 183.
[21] C. H. Harrison, ''Formulas for ambient noise level and coherence,'' J. Acoust. Soc. Am. **99**, 2055–2066 (1996).
[22] M. R. Loewen and W. K. Melville, ''A model of the sound generated by breaking waves,'' J. Acoust. Soc. Am. **90**, 2075–2080 (1991).
[23] G. B. Deane, ''Sound generation and air entrainment by breaking waves in the surf zone,'' J. Acoust. Soc. Am. **102**, 2671–2689 (1997).
[24] G. B. Deane and M. D. Stokes, ''Scale dependence of bubble creation mechanisms in breaking waves,'' Nature. Am. **418**, 839–844 (2002).
[25] M. J. Buckingham, ''On surface-generated ambient noise in an upward refracting ocean,'' Philos. Trans. R. Soc. London **346**, 321–352 (1994).
[26] M. J. Buckingham and N. M. Carbone, ''Source depth and the spatial coherence of ambient noise in the ocean,'' J. Acoust. Soc. Am. **102**, 2637–2644 (1997).
[27] G. B. Deane and M. J. Buckingham, ''Vertical coherence of ambient noise in shallow water overlying a fluid seabed,'' J. Acoust. Soc. Am. **102**, 3413–3424 (1997).
[28] L. M. Brekhovskikh, *Waves in Layered Media, 2nd ed.* (Academic, New York), pp. 299–305.
[29] M. J. Buckingham, *Noise in Electronic Devices and Systems* (Ellis Horwood, Chichester, 1983), p. 30.
[30] L. Ding and D. M. Farmer, ''Observations of breaking surface wave statistics,'' J. Phys. Oceanogr. **24**(6), 1368–1387 (1994).
[31] I. Tolstoy and C. S. Clay, *Ocean Acoustics: Theory and Experiment in Underwater Sound* (American Institute of Physics, Woodbury, NY), pp. 77, 280–281.

Fine-tuning molecular acoustic models: sensitivity of the predicted attenuation to the Lennard–Jones parameters

Andi G. Petculescu and Richard M. Lueptow[a)]
Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Rd, Evanston, Illinois 60208 USA

(Received 7 July 2004; revised 30 September 2004; accepted 8 October 2004)

In a previous paper [Y. Dain and R. M. Lueptow, J. Acoust. Soc. Am. **109**, 1955 (2001)], a model of acoustic attenuation due to vibration-translation and vibration-vibration relaxation in multiple polyatomic gas mixtures was developed. In this paper, the model is improved by treating binary molecular collisions via fully pairwise vibrational transition probabilities. The sensitivity of the model to small variations in the Lennard–Jones parameters—collision diameter (σ) and potential depth (ϵ)—is investigated for nitrogen-water-methane mixtures. For a $N_2(98.97\%)$-$H_2O(338\,\text{ppm})$-$CH_4(1\%)$ test mixture, the transition probabilities and acoustic absorption curves are much more sensitive to σ than they are to ϵ. Additionally, when the 1% methane is replaced by nitrogen, the resulting mixture [$N_2(99.97\%)$-$H_2O(338\,\text{ppm})$] becomes considerably more sensitive to changes of σ_{water}. The current model minimizes the underprediction of the acoustic absorption peak magnitudes reported by S. G. Ejakov *et al.* [J. Acoust. Soc. Am. **113**, 1871 (2003)]. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1828547]

PACS numbers: 43.35.Ae, 43.35.Fj, 43.20.Hq [RR] Pages: 175–184

I. INTRODUCTION

In a gaseous medium, the molecules exchange energy via collisions. Excitation—relaxation processes transfer energy between internal (e.g., vibrational, rotational) and external (translational) degrees of freedom, and/or among internal degrees of freedom. Upon the passage of a sound wave, excited molecules do not exchange vibrational and/or rotational energy infinitely fast with the translational degrees of freedom associated with the temperature fluctuations. As a result, the total specific heat of the relaxing gas becomes complex valued and frequency dependent. Sound propagation is characterized by a frequency-dependent and complex-valued effective wave number, the real and imaginary parts of which yield the speed of sound and attenuation, respectively. In typical molecular acoustics experiments, the absorption and sound speed are measured over a frequency range within which the expected relaxation processes occur. From these curves, the relaxation times can be extracted and gas-kinetic properties such as collision rates and transition probabilities can be calculated. The present work, on the other hand, is motivated by the need for fast acoustic monitors for gaseous processes based on algorithms that track changes in the relaxation characteristics of gas mixtures.[1,2] As a result, it is critical to be able to predict the acoustic propagation characteristics for a wide range of molecular species.

The complex process in which the molecules change state upon collision is described analytically by quantum-mechanical inelastic scattering theory. The interaction is governed by an intermolecular potential typically characterized by a long-range attractive part and a short-range repulsive ''core.'' At present, the complex nature of the forces acting between molecules is not completely understood. The attraction forces (van der Waals or London dispersion forces) are thought to arise from the electric fields induced in each molecule by their rapidly fluctuating electric dipoles,[3] while other theories advocate the attraction forces between nuclei and their distorted electron clouds.[4] At very small separations, it is conjectured that the electronic clouds of the ''collision'' pair overlap, thus leading to strong, short-ranged repulsion.[5] The Lennard–Jones (or 12-6) intermolecular potential function has been used quite successful in modeling many nonpolar or weakly polar gases. Depending on the structure of the gas molecules and the ambient conditions, the Lennard–Jones potential can be supplemented by other interaction terms describing multipole electrostatic fields, induction fields between permanent dipoles and/or quadrupoles, chemical forces, etc.[5] Depending on the complexity of the interactions involved, other intermolecular potentials may be used.[5]

The relaxation times depend on the probabilities of transition between different quantum states after a collision. The transition probabilities, in turn, are contingent upon the form of the intermolecular potential. Ultimately, it is the interaction potential that will dictate the response of the gas to acoustic forcing. In a pioneering semiclassical approach to connect molecular transfer rates and the acoustic dispersion equation, Landau and Teller[6] used an exponential repulsive potential, arguing that the long-range attractive forces are unimportant for transitions. Based on the equidistant harmonic oscillator levels, they assumed that all relaxation processes occur effectively with a single relaxation time. Simultaneously, they advanced the concept of a *transition-favorable incident velocity*. For each temperature, out of the Maxwell distribution of incident velocities, a certain value can be picked out that maximizes the probability of quantum jumps. The salient feature of the Landau–Teller model is that only the molecules entering the interaction region with ve-

a)Corresponding author; electronic mail: r-lueptow@northwestern.edu

locities lying in the vicinity of the transition-favorable value are likely to induce transitions. Jackson and Mott[7] had developed the distorted-wave method for the quantum-mechanical treatment of inelastic scattering from an exponential potential. Through its purely repulsive (positive-valued) potential, the Landau–Teller model was able to predict the behavior of certain gases with varying degrees of success.

Shields[8] showed that the Landau–Teller approach using an exponential repulsive potential adequately predicts the temperature dependence of transition probabilities in CO_2. However, by not accounting for the attractive forces between molecules, the model is not widely applicable. On the other hand, using the Lennard–Jones function instead of an exponential can lead to mathematically intractable integrals since the transition probabilities involve calculation of the matrix elements of the scattering potential. That is why a substitute for the Lennard–Jones potential was sought in the form of a shifted exponential potential function that has both positive (repulsion) and negative (attraction) values and whose slope (force) and/or magnitude are adjusted to match those of the Lennard–Jones potential at specific points.[9] In this picture, the incident molecule is accelerated under the influence of a quasiuniform attractive potential until it enters the range of the strong repulsive ''core,'' shedding its kinetic energy at the classical turning point (or closest approach point).

Schwartz, Slawsky, and Herzfeld (SSH)[10] used the distorted-wave method to calculate vibration-translation and vibration-vibration transition probabilities by replacing the one-dimensional Lennard–Jones potential by a *shifted exponential function* tangent to the former at the classical turning point for the transition-favorable incident velocity. The SSH model addresses the problem of multiple relaxation times through the concept of complex interactions, in which each collision partner makes a single-quantum jump to a different vibrational state. Schwartz and Herzfeld[11] later extended the problem to three dimensions for diatomic molecules.

The choice of the classical turning point is not arbitrary; in general terms, it represents the molecular separation where the actual ''collision'' occurs. Herzfeld and Litovitz[9] pointed out that fitting the potential function at the classical turning point and at the zero-potential point, or collision diameter (for the hard-sphere approximation, this is the effective diameter of the colliding ''spheres''), results in better agreement with data than fitting the potential energy function and its slope (force) at the classical turning point. Tanczos[12] extended the SSH model further to include two-quantum exchanges for polar polyatomic molecules, specifically chloromethanes. He applied the SSH fitting method to the Krieger potential[13] for polar molecules. However, inconsistencies between the theory and measurements resulted due to the fact that the Krieger function does not accurately model collisions between polar molecules, as pointed out by Monchick and Mason.[14] Shields[15] described qualitatively the path of the relaxation processes—series (relaxation occurs in steps, via intermediary energy levels) or parallel (each excited level relaxes separately)—by resorting to electrical circuit analogs. In addition, he discussed the sensitivity of the absorption/dispersion curves to the relaxation path. Bauer *et al.*[16] developed a generalized model for acoustic dispersion and attenuation in binary polyatomic gas mixtures that allows for multi-quanta vibrational relaxation.

While the volume of theoretical and experimental work that has been done in the field of nonclassical attenuation and dispersion of sound in gases is impressive, a common thread that runs through many models is that they require one or more free parameters in order to match the experimental data. In the case of multicomponent mixtures, the large number of factors that affect the transition probabilities acts to hinder the development of universal predictive models. That is why systematic studies aimed at identifying the aforementioned factors are warranted. Moreover, the operation of a potential real-time acoustic gas sensor[1,2] would have to be sensitive to changes in the absorption and phase velocity as the composition varies (e.g., from leaks or foreign gases). Recently, Dain and Lueptow (DL)[17,18] extended the SSH theory to model vibrational relaxation in ternary mixtures of polyatomic gases at room temperature. While their results compared favorably with experimental data, there is clearly room for improvement in the model.

The goal of this paper is to investigate the sensitivity of the nonempirical DL model, based on the SSH theory, to small variations in the Lennard–Jones parameters. Since the SSH theory is based on vibration-translation and vibration-vibration relaxation and our focus is on processes at room temperature, we do not consider vibration-rotation and rotation-translation relaxation, which typically are important at high temperatures.[19–22] In addition, a minor oversight in the DL model is corrected and the DL model is refined by using full pairwise parametrization of two-molecule collisions. The gases chosen for the study are nitrogen, methane, and water vapor because they were studied in the original DL model and because they have practical implications for an acoustic gas sensor. Of the three species, perhaps the most ''troublesome'' is water vapor. A large number of publications have been devoted to understanding the forces arising between water molecules. Given the breadth of the work on water properties, which has yielded a wide range of results for the intermolecular forces,[23] it is only natural to determine the sensitivity of a particular relaxation model to the choice of the water force constants. Nitrogen is also of interest, for two reasons: it is the heaviest of the three species, and it has the lowest vibrational amplitude. In this paper, the Lennard–Jones (LJ) force constants are varied over ranges that agree with the bounds set by published data. The central part of the analysis consists of determining (i) the amount of change induced in and (ii) the functional dependence of the transition probabilities and relaxation frequencies as the force constants of water and nitrogen are changed incrementally. Then the sensitivity of the predicted attenuation to the force constants is evaluated.

II. SHIFTED-EXPONENTIAL APPROXIMATION OF THE LENNARD-JONES POTENTIAL

The core of a relaxational theory is the calculation of transition probabilities. These are based on the quantum-mechanical treatment of the inelastic scattering of molecules of species *a*, initially in vibrational state i_a from a ''target'' of species *b*, initially in state i_b. After the ''collision,'' mol-

ecules a are in the final state f_a and molecules b are in the final state f_b. The interaction potential $V_{ab}(\mathrm{r})$ is approximated by the Lennard–Jones (LJ) function

$$V_{LJ}(r)=4\epsilon_{ab}[(\sigma_{ab}/r)^{12}-(\sigma_{ab}/r)^{6}], \tag{1}$$

where r is the molecular separation, and ϵ_{ab} and σ_{ab} are the pairwise potential depth and the zero-potential point (or collision diameter), respectively. The latter are defined in terms of the ''individual'' parameters (describing collisions between like molecules)[24]

$$\epsilon_{ab}\equiv(\epsilon_{aa}\epsilon_{bb})^{1/2} \quad \text{and} \quad \sigma_{ab}\equiv(\sigma_{aa}+\sigma_{bb})/2. \tag{2}$$

Schwartz and Herzfeld[11] fit the magnitude and slope of a shifted exponential function to the Lennard–Jones potential at the point of closest approach or classical turning point r_c, for the transition-favorable incident velocity v_0^*. The shifted-exponential fit is given by

$$V_{\exp}(r)=(\epsilon_{ab}+E^*)e^{\alpha^*_{ab}(r_c-r)}-\epsilon_{ab}. \tag{3}$$

$E^*\equiv\mu v_0^{*2}/2$, where $\mu\equiv m_a m_b/(m_a+m_b)$ is the reduced mass of the collision pair, m_a and m_b are the masses of the interacting molecules, and α^*_{ab} is the decay parameter of the exponential function (we will refer to α^*_{ab} subsequently as the repulsion parameter). The key fitting point for the Lennard–Jones and exponential functions is the classical turning point r_c, where all the kinetic energy of the incident molecules E^* is converted into potential energy. The second fitting point is the hard-sphere collision diameter σ_{ab} [for which $V_{\mathrm{LJ}}(\sigma_{ab})=V_{\exp}(\sigma_{ab})=0$]. The two potential functions are matched iteratively. Starting with arbitrary initial values for the ''input'' variables r_c and α^*_{ab}, both variables are varied until $V_{\exp}(\sigma_{ab})=V_{LJ}(\sigma_{ab})=0$ and $V_{\exp}(r_c)=V_{\mathrm{LJ}}(r_c)=E^*$. Matching the two potential functions [Eqs. (1) and (3)] as described above establishes an intimate interconnectedness between the Lennard–Jones parameters (ϵ_{ab} and σ_{ab}), the shifted exponential parameters (r_c and α^*_{ab}), and, as will become apparent later, the transition probabilities. It is this very interconnectedness that the present paper aims to bring to light.

As an example, Fig. 1 shows the shifted exponential function [Eq. (3)] matched to the Lennard–Jones potential [Eq. (1)] for N_2-H_2O collisions using $\sigma_{12}=4.086$ Å and $r_c=3.604$ Å, where subscripts 1 and 2 refer, respectively, to N_2 and H_2O (see Table II). For clarity, the intermolecular potential and separation are normalized to the LJ depth $\epsilon_{12}=2.415\times10^{-21}$ J and the collision diameter σ_{12}, respectively. For the N_2-H_2O pair, $v_0^*=1593.2$ m/s, $E^*=2.310\times10^{-20}$ J, and $\alpha^*_{12}=4.887\ \text{Å}^{-1}$. It is quite clear that the shifted exponential function is a good approximation of the repulsive part of the Lennard–Jones potential ($r<\sigma_{12}$). The molecules lose their kinetic energy until they reach the classical turning point r_c which, for this particular case, is $0.88\sigma_{12}$. For the attractive portion of the potential energy ($r>\sigma_{12}$), the exponential function matches the LJ potential reasonably well in the vicinity of the Lennard–Jones well depth. Under the action of the shifted exponential potential, molecules coming in from infinity experience almost uniform attraction until the separation distance equals the zero-potential value σ_{12}. The discrepancy between the Lennard–Jones potential and the shifted exponential function at large r is inconsequential, since the collisions are assumed to be effective only at small separations, in the vicinity of the classical turning point.[6,10]

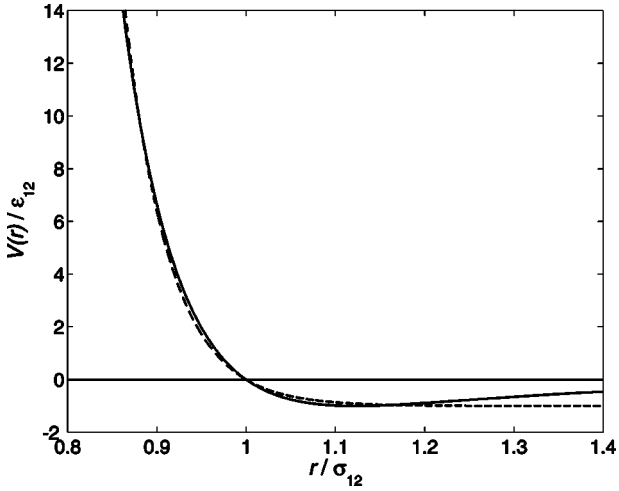


FIG. 1. The shifted exponential (dashed) and Lennard–Jones (solid) functions matched at $r=\sigma_{12}$ and $r=r_c\approx0.88\sigma_{12}$, for N_2-H_2O collisions at $T=300$ K. The repulsion parameter is $\alpha^*_{12}=4.887\ \text{Å}^{-1}$.

The parameters for the exponential function are incorporated into the expression for the transition probability, which is, after being corrected by Tanczos,[12]

$$P^{i_a\to f_a}_{i_b\to f_b}(a,b)=P_0(a)P_0(b)\frac{1.364}{1+C/T}\left(\frac{r_c}{\sigma_{ab}}\right)^2\overline{V}^2_{i_a,f_a}\overline{V}^2_{i_b,f_b}\times8\sqrt{\frac{\pi}{3}}\left(\frac{2\pi\mu\Delta E}{\alpha^{*2}_{ab}\hbar^2}\right)^2\zeta_{ab}^{1/2}\times\exp\left[-3\zeta_{ab}+\frac{\Delta E}{2k_BT}+\frac{\epsilon_{ab}}{k_BT}\right], \tag{4a}$$

where

$$\zeta_{ab}\equiv\frac{\mu v_0^{*2}}{2k_BT}=\left(\frac{\Delta E^2\mu\pi^2}{2\alpha^{*2}_{ab}\hbar^2k_BT}\right)^{1/3}. \tag{4b}$$

$\Delta E=\hbar\omega_a(i_a-f_a)+\hbar\omega_b(i_b-f_b)$ is the energy exchanged with translational degrees of freedom during a collision process, ω_a and ω_b are the vibrational angular frequencies, $P_0(a)$ and $P_0(b)$ are nonsphericity (or steric) factors, C is the Sutherland constant, $\hbar=1.0546\times10^{-34}$ Js is the reduced Planck's constant, $k_B=1.3807\times10^{-23}$ J/K is the Boltzmann constant, and T is the temperature in degrees kelvin. The vibrational factors $\overline{V}^2_{i,f}$ factors represent the squares of the transition matrix elements between the i and f harmonic oscillator states for each molecular species. For zero- and one-quantum jumps, the vibrational factors for species a are[10]

$$\overline{V}^2_{i_a,f_a}=\begin{cases}1, & \text{for } f_a=i_a\\ (i_a+1/2\pm1/2)(\alpha^*_{aa})^2\overline{A}^2\hbar/2\omega_a, & \text{for } f_a=i_a\pm1'\end{cases} \tag{5}$$

TABLE I. Normal modes of vibration, mode degeneracies, and vibrational amplitude coefficients for N_2, H_2O, and CH_4 from Ref. 25. The modes in bold type are those assumed to be active around room temperature.

Species	Normal modes (cm^{-1})		g	$\overline{V^2}$ (amu^{-1})
N_2	ν	**2331**	**1**	**0.0354**
H_2O	ν_1	3657	1	0.9539
	ν_2	**1596**	**1**	**0.9527**
	ν_3	3756	1	0.9241
CH_4	ν_1	2915	1	0.9921
	ν_2	**1534**	2	**0.9921**
	ν_3	3019	1	0.9923
	ν_4	**1306**	3	**0.8368**

where $\overline{A^2}$ is the vibrational amplitude coefficient representing the averaged displacement of surface atoms for unit change of the normal coordinate for a given vibration.[25] Inherent in the probabilities as described by Eq. (4a) is the existence of two ways in which the exchange of translational and vibrational energy occurs. First, there are processes whereby only one collision partner changes vibrational state. For example, the probability of molecule *a* losing one quantum ($i_a=1\rightarrow f_a=0$) while the vibrational state of molecule *b* is unaltered ($i_b=0\rightarrow f_b=0$) would be described by $P_{00}^{10}(a,b)$, or simply $P^{10}(a,b)$, where the arrows in the superscript and subscript have been omitted. This process, where a single quantum is exchanged with translation, is called a ''vibration-translation,'' or V-T, interaction. Second, complex collisions may exist in which *both* partners gain/lose one quantum each, the difference $\hbar\omega_a-\hbar\omega_b$ being changed into kinetic energy.[9] Such a process, where a total of two quanta are exchanged with translation and which is described by the probability $P_{01}^{10}(a,b)$, is called a ''vibration-vibration,'' or V-V, interaction.

Equations (3), (4a), and (4b) show explicitly the sensitivity of a relaxational model to the values of the parameters ϵ_{ab}, σ_{ab}, α_{ab}^*, and r_c. Because of the matching procedure, the LJ collision diameter (σ_{ab}) and potential well depth (ϵ_{ab}) are interconnected with the repulsion parameter α_{ab}^* and the classical turning point r_c of the shifted exponential function. As a result, a change in σ_{ab} or ϵ_{ab} leads immediately to changes in α_{ab}^* and r_c, which in turn affect the transition probabilities. This dependence of the transition probabilities on the force constants and their impact on acoustic propagation in gas mixtures are the focus of this study.

As in Ref. 17, a mixture of water vapor, methane, and nitrogen is considered. Table I provides the vibrational modes, degeneracies *g*, and vibrational coefficients of nitrogen, water, and methane. At temperatures around 300 K, it is assumed that only the lowest modes are significant. The contribution of higher modes to molecular energy transfer at these temperatures is small. Consequently, it is assumed that energy exchange will occur with a high probability only between the following vibrational modes: $\nu=2331\ cm^{-1}$ of N_2, $\nu_2=1596\ cm^{-1}$ of H_2O, and $\nu_2=1534\ cm^{-1}$ and $\nu_4=1306\ cm^{-1}$ of CH_4. The second vibrational mode $\nu_2=1534\ cm^{-1}$ of CH_4 is included in the relaxation process due to its near resonance with the mode $\nu_2=1596\ cm^{-1}$ of H_2O. In the analysis, the following indices are used: 1 for N_2, 2 for $H_2O(\nu_2)$, 3 for $CH_4(\nu_4)$, and 4 for $CH_4(\nu_2)$, as in Ref. 17.

TABLE II. Reference parameters for the Lennard–Jones and exponential potentials (the values for σ_{aa} and ϵ_{aa} ($a=1...4$) are similar to those used in Refs. 17 and 18, based on Ref. 5). The quantities that will change in the analysis are shown in boldface.

Pair (ab)	σ_{ab} (Å)	ϵ_{ab}/k_B (K)	α_{ab}^* ($Å^{-1}$)	r_c (Å)
11	***3.704***	***80.01***	**5.093**	**2.941**
22	***4.468***	***382.43***	**4.530**	**4.002**
33	*4.075*	*143.91*	4.796	3.490
44	*4.075*	*143.91*	4.775	3.465
21; **12**	**4.086**	**174.92**	**4.887**	**3.604**
23; **32**	**4.271**	**234.60**	**4.924**	**3.965**
24; **42**	**4.271**	**234.60**	**5.207**	**4.118**
31; **13**	**3.890**	**107.3**	**4.989**	**3.292**
34; 43	4.075	143.91	5.081	3.733
41; **14**	**3.890**	**107.31**	**5.023**	**3.328**

In the present sensitivity analysis, the Lennard–Jones potential parameters for nitrogen ($\sigma_{11},\epsilon_{11}$) and water ($\sigma_{22},\epsilon_{22}$) are varied from the original values used in Ref. 17, which are based on viscosity.[5] As a result, the parameters σ_{ab} and ϵ_{ab} for all the interactions except those between the two CH_4 modes will change; for every interaction pair, the repulsion parameter α_{ab}^* is determined by matching the associated exponential and LJ potentials as described following Eq. (3). Table II provides the reference values of the Lennard–Jones force constants and the corresponding exponential fit parameters for the four vibrational modes. The original model described in Refs. 17 and 18 used only the italicized quantities in Table II. For the model described here, all of the values in Table II are used in an effort to improve accuracy. Boldface indicates the values that will change in the analysis. The values for α_{aa}^* used in the original DL model were slightly different than those shown in Table II (within 2% for nitrogen and methane, and 6% for water). Also, in the original DL model, the classical turning point was not used explicitly. Instead, the ratio $(r_c/\sigma_{aa})^2$ was approximated by imposing $V_{LJ}(r_c)\approx k_BT$. Only the values in the first two lines of Table II will be varied systematically; the other boldface values will change as a result of varying the ''11'' and ''22'' values.

In the first step of the analysis, the LJ force constants of water are kept constant ($\sigma_{22}=4.468$ Å, $\epsilon_{22}/k_B=382.43$ K) while the values of σ_{11} and ϵ_{11}/k_B of nitrogen are allowed to vary, in turn, from 3.334 to 3.704 Å, and from 80.01 to 96.01 K, respectively. In the second step, the force constants for N_2 are held constant ($\sigma_{11}=3.704$ Å, $\epsilon_{11}/k_B=80.01$ K) and those of H_2O are allowed to change: σ_{22} from 3.217 to 4.468 Å, and, separately, ϵ_{22}/k_B from 302.12 to 474.21 K. The directions of variation of the water and nitrogen force constants are such that, starting from initial reference values used in Table II, they stay within the ranges found in the literature.[5,14,23] Thus σ_{11} varies between 3.334 and 3.704 Å, ϵ_{11}/k_B between 80.01 and 96.01 K, σ_{22} between 3.217 and 4.468 Å, and ϵ_{22}/k_B between 302.12 and 474.21 K for water. Values of $\sigma_{22}<3$ Å and $\epsilon_{22}/k_B>700$ K appear in the literature but they often involve nonphysical assumptions such as

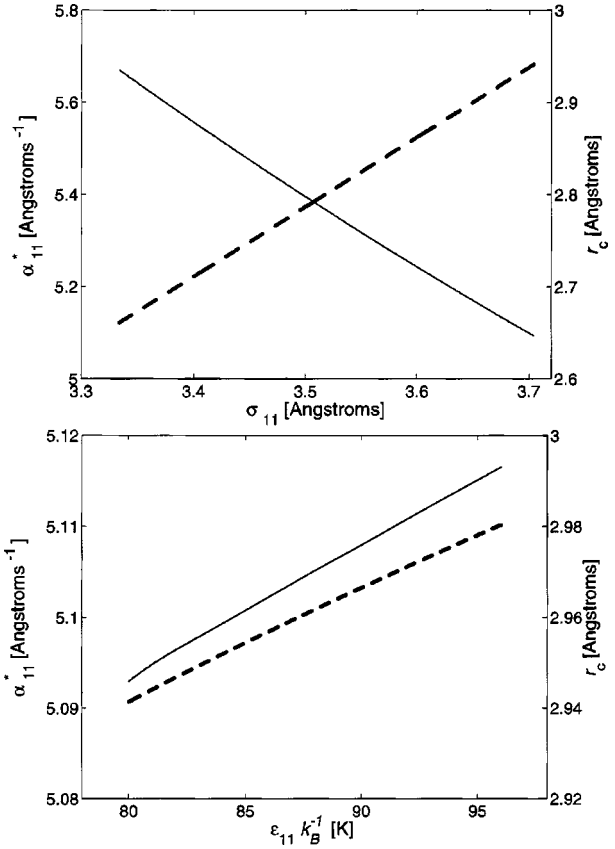


FIG. 2. Variation of the nitrogen repulsion parameter α_{11}^{*} (solid) and closest approach point r_c (dashed) with the nitrogen collision diameter, σ_{11} (top), and with the nitrogen LJ potential depth, ϵ_{11} (bottom).

embedded point dipoles[14] and/or single-point interactions,[23] so they are not considered here.

The transition probabilities do not have an explicit functional dependence on the LJ potential parameters, σ and ϵ. Changes in σ and ϵ alter the shifted-exponential repulsion parameter α^* and the classical turning point r_c, which also enter the expression for the transition probabilities. Therefore, an analysis of the sensitivity to values of σ and ϵ should start with showing how these parameters affect α^* and r_c. Figure 2 shows the dependence of the N_2-N_2 repulsion parameter (α_{11}^{*}) and turning point (r_c) on σ_{11} and ϵ_{11}. It is clear that α_{11}^{*} and r_c vary monotonically with σ_{11} and ϵ_{11} over the parameter range found in the literature.

Of course, the interspecies transition probabilities involving nitrogen will also be affected by the variation in these parameters, according to Eqs. 4(a), and 4(b). It is these transition probabilities that affect the acoustic attenuation and dispersion in gas mixtures, which are the focus of this study. Figure 3 shows selected transition probabilities as a function of the corresponding pairwise LJ parameters [for the sake of brevity, only N_2-N_2 ("1,1") and N_2-H_2O ("1,2") cases are shown; transition probabilities between nitrogen and the two modes of methane exhibit a behavior similar to that of Figs. 3(b) and 3(d)]. The dependence of the transition probabilities on pairwise collision diameter σ_{11} is notably nonlinear for the V-T collisions among nitrogen molecules [$P^{10}(1,1)$]. The dependence on σ_{12} is only slightly nonlinear for the V-V nitrogen-water interaction $P_{01}^{10}(1,2)$. The results

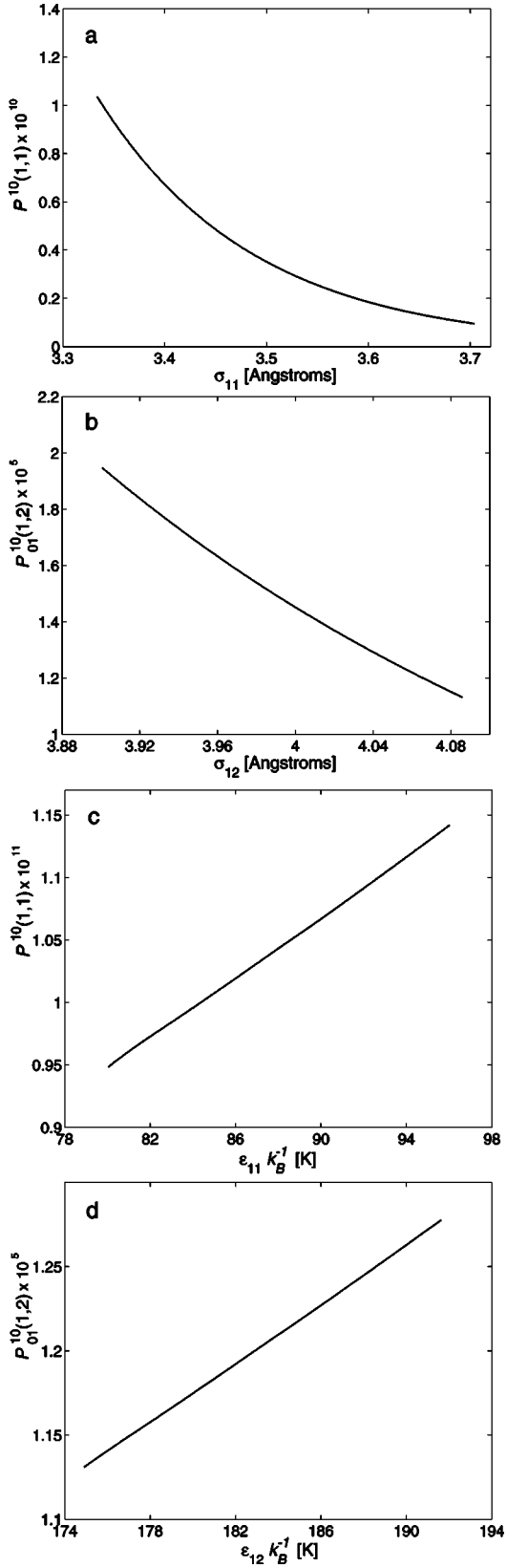


FIG. 3. The effects of varying the N_2 force constants ϵ_{11} and σ_{11} on transition probabilities for N_2-N_2 ("1,1") and N_2-H_2O ("1,2") collisions. (a) dependence V-T probability dependence on σ_{11}, for $\sigma_{22}=4.468$ Å, (b) V-V probability dependence on σ_{12}, for $\sigma_{22}=4.468$ Å, (c) V-T probability dependence on ϵ_{11}/k_B, for $\epsilon_{22}/k_B=382.43$ K, and (d) V-V probability dependence on ϵ_{12}/k_B, for $\epsilon_{22}/k_B=382.43$ K. Note the considerably larger sensitivity to the values of the collision diameters.

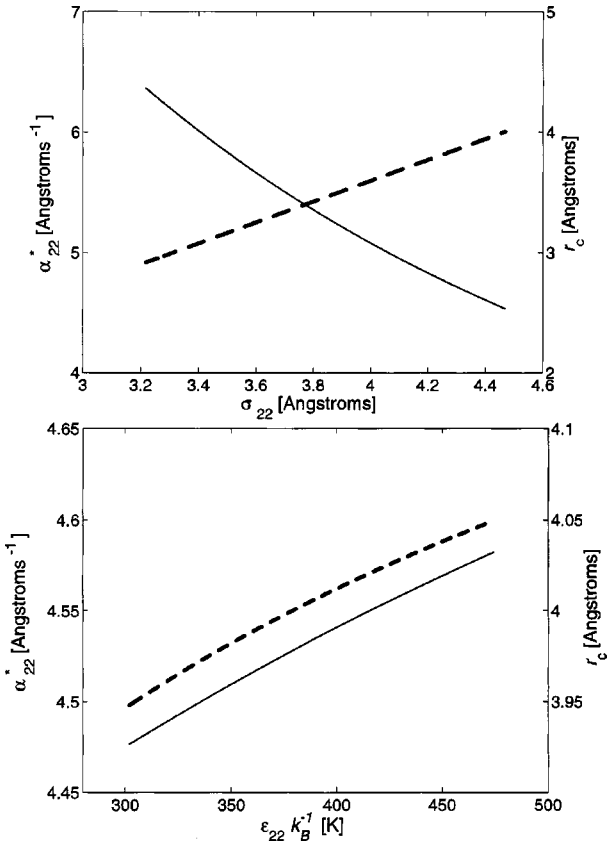


FIG. 4. Variation of the water repulsion parameter α_{22}^{*} (solid) and closest approach point r_c (dashed) with the water collision diameter, σ_{22} (top), and with the water LJ potential depth, ϵ_{22} (bottom).

for $P_{01}^{10}(2,1)$, which are not shown, are similar to those for $P_{01}^{10}(1,2)$. The transition probabilities increase nearly linearly over the range of the potential depths (ϵ_{11} and ϵ_{12}). The more important observation, however, is that the transition probabilities are quite sensitive to variations in the collision diameters, varying by a factor of almost 11 for $P^{10}(1,1)$ and a factor of almost 2 for $P_{01}^{10}(1,2)$ and $P_{01}^{10}(2,1)$. The transition probabilities are much less sensitive to variations in the potential depths, varying by less than 20%.

The dependence of the repulsion parameters and turning points, and of selected transition probabilities on the water force constants are shown in Figs. 4 and 5, respectively. Figure 4 shows the dependence of the H_2O-H_2O (“22”) repulsion parameter (α_{22}^{*}) and turning point (r_c) on σ_{22} and ϵ_{22}. Again, α_{22}^{*} and r_c vary monotonically with σ_{22} and ϵ_{22} in the range that is considered. The transition probabilities have a nonlinear dependence on the collision diameters [notably so for $P^{10}(2,2)$ for V-T transitions among water molecules]. In this case, the dependence of the V-T probability on the potential depth ϵ_{22}/k_B is also somewhat nonlinear, although the dependence of the V-V probability is nearly linear. Again, it is evident from Fig. 5 that the transition probabilities are quite sensitive to changes in the collision diameters [a factor of almost 127 for $P^{10}(2,2)$ and a factor of nearly 6 for $P_{01}^{10}(1,2)$]. However, the V-T transition probability is also quite sensitive to the potential depth of water ϵ_{22}/k_B, unlike the dependence on the potential depth of nitrogen ϵ_{11}/k_B.

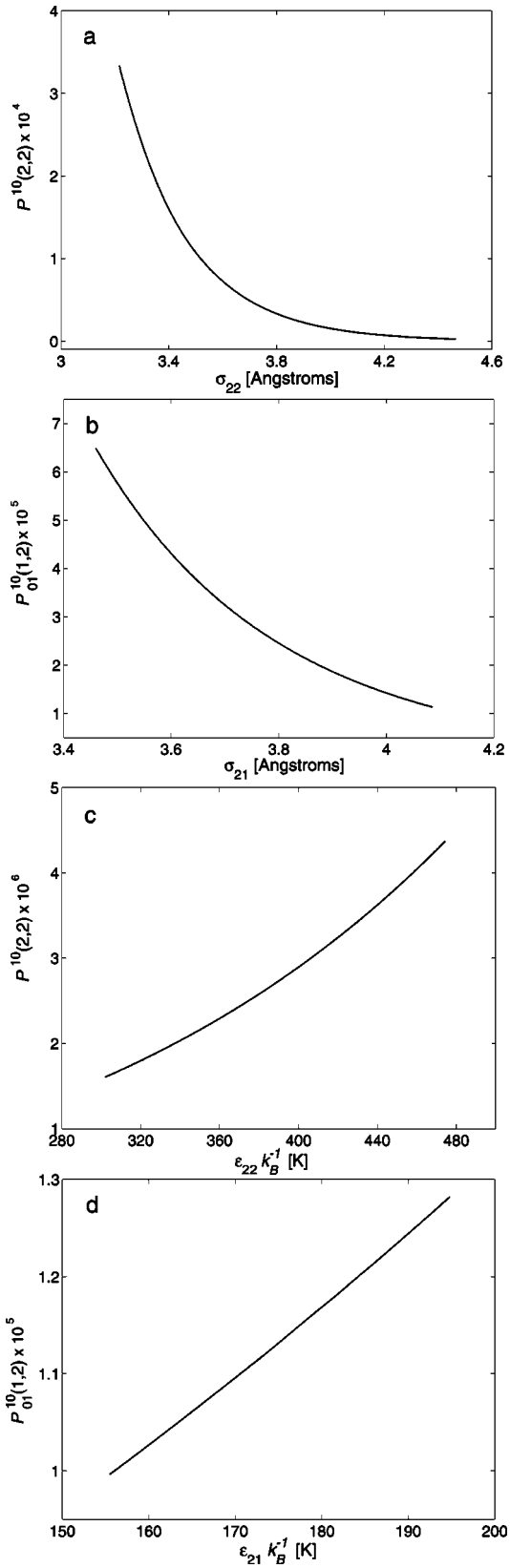


FIG. 5. The effects of varying the H_2O force constants ϵ_{22} and σ_{22} on transition probabilities for H_2O-H_2O (“2,2”) and N_2–H_2O (“1,2”) collisions. (a) V-T probability dependence on σ_{22}, for $\sigma_{11}=3.704$ Å, (b) V-V probability dependence on $\sigma_{21}(=\sigma_{12})$, for $\sigma_{11}=3.704$ Å, (c) V-T probability dependence on ϵ_{22}/k_B, for $\epsilon_{11}=80.01$ K and (d) V-V probability vs $\epsilon_{21}/k_B(=\epsilon_{12}/k_B)$, for $\epsilon_{11}=80.01$ K. Note the considerably larger sensitivity to values of the collision diameters.

III. SENSITIVITY OF ACOUSTIC ATTENUATION TO LENNARD–JONES PARAMETERS

The DL model[17] is centered on the set of relaxation equations

$$\frac{d\Delta T^{\mathrm{vib}}}{dt} = -\mathbf{A}\Delta T^{\mathrm{vib}} + \mathbf{q}\Delta T, \tag{6}$$

where ΔT^{vib} and ΔT are the fluctuations of the vibrational and translational temperatures, respectively, with respect to the equilibrium temperature T_0. $\mathbf{q}$ is a vector and $\mathbf{A}$ is the 4×4 relaxation matrix incorporating the inverses of the relaxation times (effective relaxation frequencies). Their components are given by the following relations:[17]

$$q_a = \frac{1}{\tau_a^{VT}} + \sum_{\substack{b=1\\ b\neq a}}^{4} \frac{1}{\tau_{a,b}^{VV}} \frac{1-\exp(-\hbar\omega_a/k_BT_0)}{1-\exp(-\hbar\omega_b/k_BT_0)}\left[1-\frac{\omega_b}{\omega_a}\right], \tag{7}$$

$$A_{aa} = \frac{1}{\tau_a^{VT}} + \sum_{\substack{b=1\\ b\neq a}}^{4} \frac{1}{\tau_{a,b}^{VV}} \frac{1-\exp(-\hbar\omega_a/k_BT_0)}{1-\exp(-\hbar\omega_b/k_BT_0)},$$

$$a = 1,...,4,$$

$$A_{ab} = -\frac{1}{\tau_{a,b}^{VV}} \frac{1-\exp(-\hbar\omega_a/k_BT_0)}{1-\exp(-\hbar\omega_b/k_BT_0)} \frac{\omega_b}{\omega_a}, \tag{8}$$

$$a,b = 1,...,4, \quad a\neq b.$$

The translational (V-T) relaxation times that appear in Eq. (5) are

$$\frac{1}{\tau_a^{VT}} = \sum_{b=1}^{3} \frac{\alpha_b}{\tau_{a,b}^{VT}}, \quad a = 1,...,3,$$

$$\frac{1}{\tau_4^{VT}} = \sum_{b=1}^{2} \frac{\alpha_b}{\tau_{4,b}^{VT}} + \frac{\alpha_3}{\tau_{4,4}^{VT}} \tag{9}$$

where α_i are the molar fractions of the three species ($\alpha_4 = \alpha_3$, for the two methane modes considered). The paired V-T and V-V inverse relaxation times for excitation processes with one and two vibrational modes involved are, respectively,[17]

$$(\tau_{a,b}^{VT})^{-1} = Z(a,b)P^{10}(a,b)[1-\exp(-\hbar\omega_a/k_BT_0)]$$

$$(\tau_{a,b}^{VV})^{-1} = \alpha_b g_b Z(a,b)P_{01}^{10}(a,b), \quad a,b = 1,...,4, \tag{10}$$

$$a\neq b, \quad \alpha_3 = \alpha_4.$$

Here $Z(a,b)$ are the rates of collision of molecules of species a with molecules of species b. The collision rate per molecule is based on the kinetic theory for a gas of rigid spheres[25]

$$Z(a,b) = 2N_a\sigma_{ab}^2\sqrt{2\pi k_BT_0\frac{m_a+m_b}{m_am_b}}, \tag{11}$$

where N_a is the number density of the "a" molecules, and m_a, m_b are the molecular masses of the two species. The calculation of the effective acoustic wave number is described in detail in Ref. 17. It is set up from linear acoustic equations for ideal gases. Substitution of a harmonic plane-wave solution of the wave equation into Eq. (6) results in the following algebraic set of equations:[17]

$$(i\omega\mathbf{I}+\mathbf{A})\Delta T^{\mathrm{vib}} = \mathbf{q}\Delta T, \tag{12}$$

where $\mathbf{I}$ is the identity matrix. Equation (12) connects the temperature fluctuations of the internal (vibrational) and external (translational) degrees of freedom. The complex-valued, frequency dependent effective wave number is given by the following expression:

$$\tilde{k}^2 = k_0^2\frac{C_V^0+\Sigma_{i=1}^4\alpha_iC_i^{\mathrm{vib}}(\Gamma_i-1)}{C_P^0+\Sigma_{i=1}^4\alpha_iC_i^{\mathrm{vib}}(\Gamma_i-1)}, \tag{13}$$

where k_0, C_V^0, and C_P^0 are the static ($\omega\to0$) values of the wave number and the translational isochoric and isobaric specific heats of the mixture, respectively, C_i^{vib} is the vibrational specific heat of species i, and $\Gamma_i \equiv \Delta T_i^{\mathrm{vib}}/\Delta T_i$ is the temperature fluctuation ratio. The imaginary and real parts of the effective wave number $\tilde{k}$ determine the attenuation and phase velocity of the acoustic wave. The original model developed in Ref. 17 inadvertently used $C_V^0+\Sigma_{i=1}^4\alpha_iC_i^{\mathrm{vib}}\Gamma_i$ and $C_P^0+\Sigma_{i=1}^4\alpha_iC_i^{\mathrm{vib}}\Gamma_i$ instead of the corresponding expressions in the numerator and the denominator, respectively, of the right-hand side of Eq. (13). The correction moderately increases the magnitude of the theoretical attenuation peaks presented in Ref. 2. For example, for pure methane, the correction raises the height of the normalized attenuation peak by about 15%, to within 2% of the experimental data. Also for methane, using the pairwise model described in Sec. I reduces the disagreement between the predicted and measured attenuation peaks from 19% to 4%.

The model described by Eqs. (6)–(13) is applied to calculate the effective wave number, the imaginary and real parts of which yield the frequency-dependent acoustic absorption $\alpha(\omega)$ and speed of sound $c(\omega)$, in a mixture of $N_2(98.97\%)$-$H_2O(338\,\mathrm{ppm})$-$CH_4(1\%)$. This particular mixture was considered because experimental data is available for the acoustic attenuation over a range of frequencies,[27] it includes the effects of water vapor, and it was previously considered in Ref. 17. The sensitivity of the attenuation to changes in the water and nitrogen force constants is shown in Fig. 6. The LJ force constants are varied from their reference values (Table II), which happen to be near one end of the range of values in the literature, to values at the other end of the range. Specifically, σ_{11} and σ_{22} are varied from 3.704 to 3.334 Å (10%) and from 4.468 to 3.217 Å (28%), respectively, while ϵ_{11}/k_B and ϵ_{22}/k_B are varied from 80.01 Å to 96.01 K (20%) and from 382.43 K to 474.21 K (24%), respectively. The frequency/pressure ratios at which the absorption peaks occur in Fig. 6 are indicative of the relaxation processes occurring in the mixture; their relative magnitudes depend on the concentrations of the component molecular species. As a result, the magnitudes of the peaks are unaltered as the force constants are varied, but the relaxation frequencies do depend on the choice of the force constants, σ and ϵ. It is apparent that the attenuation of the mixture is considerably more sensitive to changes in σ_{11}, the collision diameter of nitrogen, than to σ_{22}, ϵ_{11} or ϵ_{22}. The

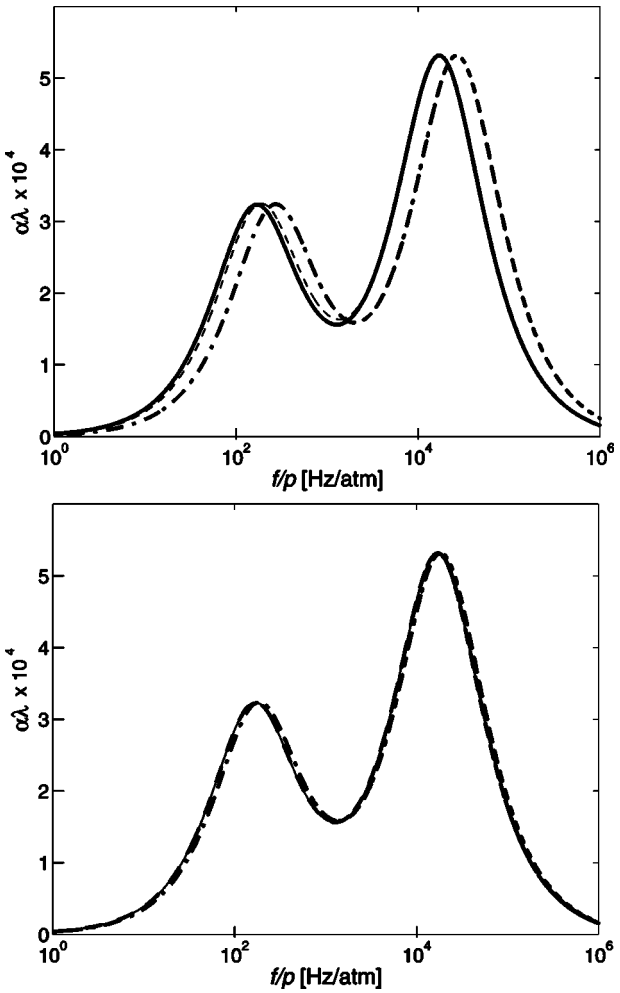


FIG. 6. Effects of the variation of the collision diameter and potential depth for the N_2(98.97%)-H_2O(338 ppm)-CH_4(1%) mixture on the attenuation nondimensionalized by the acoustic wavelength. Top: σ_{11} and σ_{22} are varied by 10% and 28%, respectively, from their reference values (Table II). Bottom: ϵ_{11}/k_B and ϵ_{22}/k_B are varied by 20% and 24%, respectively, from their reference values (Table II). Solid: using reference values ($\sigma_{11}=3.704$ Å, $\sigma_{22}=4.468$ Å, $\epsilon_{11}/k_B=80.01$ K, $\epsilon_{22}/k_B=382.43$ K). Dashed: effect of changing H_2O parameters, so that $\sigma_{22}=3.217$ Å and $\epsilon_{22}/k_B=474.21$ K, with σ_{11} and ϵ_{11}/k_B at their reference values. Note that the effect of changing ϵ_{11} is so small that the dashed is hidden by the solid curve in the lower figure. Dot-dashed: Effect of changing N_2 parameters, so that $\sigma_{11}=3.334$ Å and $\epsilon_{11}/k_B=96.01$ K, with σ_{22} and ϵ_{22}/k_B at their reference values.

effective relaxation frequencies (inverse relaxation times) are the eigenvalues of the matrix **A** in the relaxation equation, Eq. (12), the eigenvectors of which specify the normalized contribution of each species (see Ref. 17 for details and examples). The strong shift of the absorption peak due to the overall change in σ_{11} is not surprising given the large nitrogen concentration and the wide variation in $P^{10}(1,1)$ with σ_{11} [Fig. 3(a)]. On the other hand, the larger change in $P^{10}(2,2)$ with σ_{22} [Fig. 5(a)] as well as the stronger dependence of $P_{01}^{10}(1,2)$ on σ_{12} [Fig. 5(b)] are not enough to cause an absorption peak shift comparable to that due to the variation of the nitrogen collision diameter. We conjecture that this is the case due, first, to the very small water vapor concentration (338 ppm), and, second, to a possible ''buffering'' effect of methane with its 1534 cm^{-1} vibrational level which is quasiresonant with the 1596 cm^{-1} mode of water. In addition, the V-V transition probabilities between the two modes of methane considered and water are several orders of magnitude higher than the V-V probabilities between methane and nitrogen, which may also affect the attenaution.

The effect of the Lennard–Jones force constants on the acoustic attenuation may be more easily understood in terms of the behavior of the relaxation frequencies. The explicit dependence of the two main relaxation frequencies (normalized absorption peaks in Fig. 6) on the values of the nitrogen LJ parameters is shown in Fig. 7 ($f_{\mathrm{relax},1}$ and $f_{\mathrm{relax},2}$ refer to the frequencies of the first and second attenuation peaks, respectively). The large effect of the nitrogen collision diameter σ_{11} is obvious. Varying σ_{11} results in large changes in both relaxation frequencies by factors of 2.7 and 2.4 for $f_{\mathrm{relax},1}$ and $f_{\mathrm{relax},2}$, respectively. The change in the potential depth ϵ_{11} brings about a considerably smaller variation in the relaxation frequencies (about 10% or less). When the water LJ parameters are varied (not shown), the change in the collision diameter of water molecules σ_{22} brings about relatively small changes in the two main relaxation frequencies (about 5%); similarly, the effects of changing ϵ_{22} amount to less than 1%. The dependence of the two relaxation peaks on σ_{11} is slightly nonlinear, apparently related to the sensitivity of the transition probabilities to the potential parameters (Fig. 3).

If the 1% molar fraction of methane is replaced by nitrogen so that the new mixture is 338 ppm water and the rest nitrogen, and the Lennard–Jones parameters of the two constituents are allowed to vary as before, the effect on the relaxation curves changes drastically: Water now dominates the sensitivity of the mixture. In this case, changes in both the water and nitrogen collision diameters σ_{22} and σ_{11} induce considerable shifts of the main relaxation frequency, as shown in Fig. 8. As in the case of the ternary mixture, the variations in the N_2-H_2O mixture attenuation due to changes in ϵ_{11} and ϵ_{22} are insignificant, and therefore are not shown. The behavior of the main relaxation frequency as a function of the LJ parameters of water and nitrogen (not shown) is similar to that shown in Fig. 7.

The difference in the degree of the impact of changing σ_{22} between the attenuation curves in Figs. 6 and 8 could be explained in terms of the vibrational factors, which are proportional to the ratio of vibration amplitudes and frequencies and play a significant role in the evaluation of transition probabilities.[26] Table I indicates that the vibrational factor of the nitrogen molecules is more than one order of magnitude smaller than those of the other species present in the mixture. The lightest of the three gases, methane, has the highest vibrational factor. The 1534 cm^{-1} mode of methane very likely couples near resonantly with the 1596 cm^{-1} mode of water. As a result, changing the force constants of water does not amount to a sizable effect on the relaxation curve in Fig. 6. However, when the methane is entirely replaced by nitrogen, water molecules lose a ''favorite'' interaction partner, with a commensurately high vibration factor. The water-nitrogen mixture is a composition of extremes: Very small amounts of water are able to ''push'' the relaxation frequency from that of pure nitrogen (<0.1 Hz/atm) (Ref. 27) to around 6–7

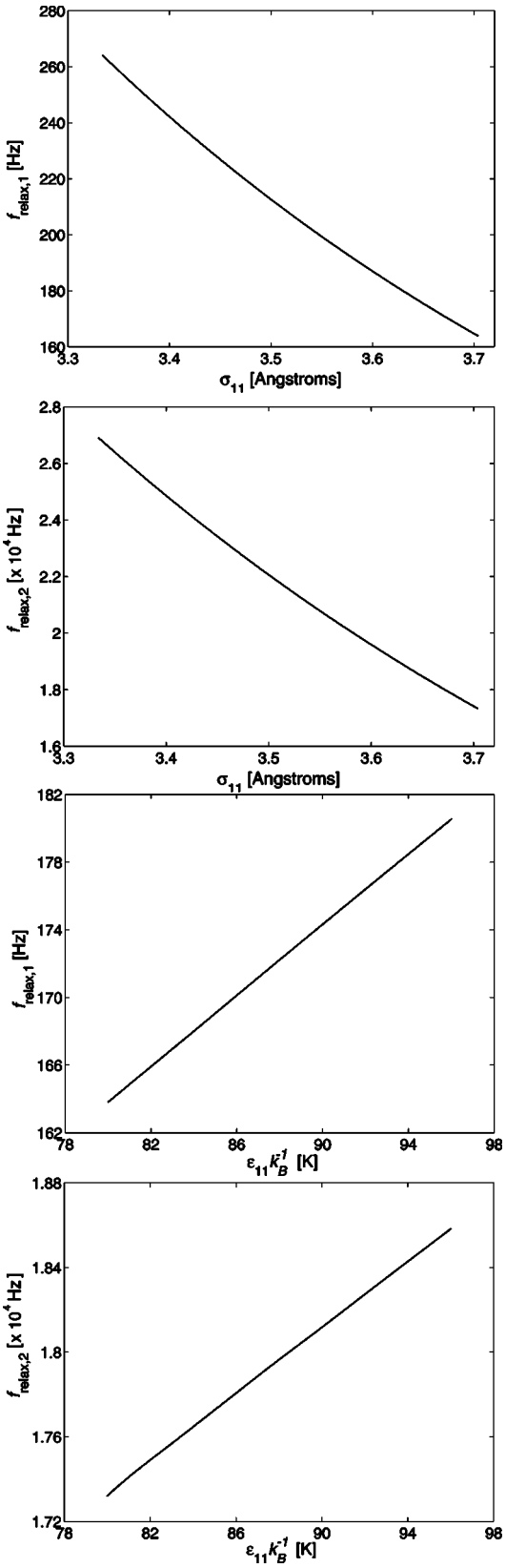


FIG. 7. Dependence of the two main relaxation frequencies (normalized absorption peaks in Fig. 6), at $P_0=1$ atm, on the values of the LJ parameters of nitrogen. Note the higher sensitivity of the relaxation frequencies to σ_{11}.

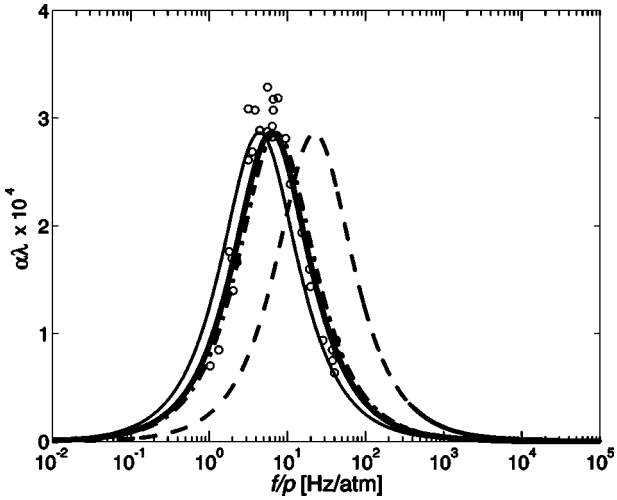


FIG. 8. Effects of the variation of the collision diameter for a mixture of 338 ppm H_2O and the rest N_2, on the normalized attenuation. Thin solid: using reference values for $\sigma_{11}=3.704$ Å and $\sigma_{22}=4.468$ Å (Table II). Dashed: effect of 28% change of σ_{22} from its reference value of 4.468 Å to 3.217 Å. Dot-dashed: effect of 10% change of σ_{11} from its reference value of 3.704 to 3.334 Å. Circles: experimental data of Zuckerwar and Griffin (Ref. 28). Bold solid: using $\sigma_{11}=3.704$ Å and $\sigma_{22}=4.200$ Å (6% change from reference value).

Hz/atm. Beside the large changes induced by the large quantity of heavy nitrogen molecules, varying the potential parameters of the water molecules, which are characterized by a considerably higher vibration factor, shifts the relaxation frequencies to a large extent.

Experimental data obtained by Zuckerwar and Griffin[28] for the same nitrogen-water vapor mixture is also plotted. By slightly adjusting the hard-sphere collision diameter σ of water or nitrogen, the absorption curve can be made to agree very well with the data. As an example, by making the water collision $\sigma_{22}=4.200$ Å (a 6% change from its reference value), the predicted curve can be made to agree fairly well with the data.

IV. CONCLUSIONS

The paper addresses acoustic propagation in gaseous media, associated with vibration-translation and vibration-vibration relaxation, as influenced by the choice of the Lennard–Jones parameters. Although certain molecular interaction specifics are described in some detail, the drive behind the current work is acoustics, specifically the ongoing quest for precise acoustic propagation models to be used in gas sensing applications.

The salient point of the SSH-Tanczos treatment of molecular relaxation in gases is the approximation of the Lennard–Jones interaction potential by a shifted exponential function in the vicinity of the classical turning point. The probability of quantum transitions between vibrational levels depends nontrivially on the Lennard–Jones potential depth ϵ and hard-sphere collision diameter σ. However, the values for both ϵ and σ can vary substantially in the literature, since they are quite difficult to measure. What has not been clear is how these variations in the fundamental collision parameters play out with regard to the values for the shifted exponential repulsion parameter α^* and the classical turning point r_c, which, in turn, affect the transition probabilities and subsequently the acoustic attenuation. Furthermore, in multicom-

ponent gas mixtures, these two parameters interact such that a small variation in one of the parameters of the intermolecular energy (here, approximated by the Lennard–Jones potential) could result in a large change in the acoustic attenuation in the mixture.

To address the sensitivity of the acoustic attenuation in gas mixtures to the Lennard–Jones parameters, the exchange of energy with translation is treated via a full pairwise parametrization of the transition probabilities, assuming binary collisions. The shifted-exponential parameters α^* and r_c depend on the choice of the Lennard–Jones parameters, ϵ and σ. The specific choice of these parameters affects the transition probabilities and, subsequently, the acoustic attenuation (and phase velocity) in model mixtures of nitrogen, methane, and water vapor. The behavior of the transition probabilities as a function of changing the values of the LJ parameters (decreasing with increasing collision diameter, and increasing with potential depth—see Figs. 3 and 5) can be explained, in a conjectural context, starting with the behavior of the repulsion parameter and classical turning point. Figures 2 and 4 show that r_c increases with both σ and ϵ, while α^* decreases with increasing σ and increases with ϵ. An increasing r_c means that the point where two ''colliding'' molecules transfer their relative kinetic energy to potential energy occurs at smaller separations. The parameter α^* is the ''slope'' of the shifted-exponential potential function. It determines the ''strength'' of the interaction potential: Large values of α^* imply ''strong'' interaction, while small values of α^* imply ''weak'' interaction. Thus, as the value of the collision diameter σ increases, it can be conjectured that there is a progressively weaker interaction occurring at progressively shorter intermolecular separations; hence the decrease in the transition probabilities. On the other hand, an increasing potential depth ϵ leads to a progressively stronger interaction at shorter separations and thus to an increase in the transition probabilities. The transition probabilities are extremely sensitive to the value of collision diameter σ rather than to the depth of the Lennard–Jones potential well ϵ. The sensitivity of the acoustic relaxation model to the choice of an LJ parameter (the collision diameter for water, for the mixture considered) is itself ''sensitive'' to the concentration of the molecular species making up the mixture.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of NASA (Grant No. NAG9-1544) as well as fruitful discussions with Douglas Shields.

[1] S. Phillips *et al.*, ''Theory for a gas composition sensor based on acoustic properties,'' Meas. Sci. Technol. **14**, 70–75 (2003).
[2] S. G. Ejakov *et al.*, ''Acoustic attenuation in gas mixtures with nitrogen: experimental data and calculations,'' J. Acoust. Soc. Am. **113**, 1871–1879 (2003).
[3] F. London, ''The general theory of molecular forces,'' Trans. Faraday Soc. **33**, 8–26 (1937).
[4] R. P. Feynman, ''Forces in molecules,'' Phys. Rev. **56**, 340–343 (1939).
[5] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, Chap. 13, Wiley, New York, 1964.
[6] L. Landau and E. Teller, ''Zur Theorie der Schalldispersion,'' Phys. Z. Sowjetunion **10**, 34–43 (1936).
[7] J. M. Jackson and N. F. Mott, ''Energy exchange between inert gas atoms and a solid surface,'' Proc. R. Soc. London **137**, 703–717 (1932).
[8] F. D. Shields, ''Thermal relaxation in carbon dioxide as a function of temperature,'' J. Acoust. Soc. Am. **29**, 450–454 (1957).
[9] K. F. Herzfeld and T. H. Litovitz, *Absorption and Dispersion of Ultrasonic Waves*, Chap. 8, Academic Press, New York, 1959.
[10] R. N. Schwartz, Z. I. Slawsky, and K. F. Herzfeld, ''Calculation of vibrational relaxation times in gases,'' J. Chem. Phys. **20**, 1591–1599 (1952).
[11] R. N. Schwartz and K. F. Herzfeld, ''Vibrational relaxation times in gases (three-dimensional treatment),'' J. Chem. Phys. **22**, 767–773 (1954).
[12] F. I. Tanczos, ''Calculation of vibrational relaxation times of the chloromethanes,'' J. Chem. Phys. **25**, 439–447 (1956).
[13] F. J. Krieger, ''The viscosity of polar gases,'' Project Rand Report RM-646, July, 1, 1951.
[14] L. Monchick and E. A. Mason, ''Transport properties of polar gases,'' J. Chem. Phys. **35**, 1676–1697 (1961).
[15] F. D. Shields, ''On obtaining transition rates from sound absorption and dispersion curves,'' J. Acoust. Soc. Am. **47**, 1262–1268 (1970).
[16] H.-J. Bauer, F. Douglas Shields, and H. E. Bass, ''Multimode vibrational relaxation in polyatomic molecules,'' J. Chem. Phys. **57**, 4624–4628 (1972).
[17] Y. Dain and R. M. Lueptow, ''Acoustic attenuation in three-component gas mixtures—theory,'' J. Acoust. Soc. Am. **109**, 1955–1964 (2001).
[18] Y. Dain and R. M. Lueptow, ''Acoustic attenuation in a three- gas mixture—results,'' J. Acoust. Soc. Am. **110**, 2974–2979 (2001).
[19] C. B. Moore, ''Vibration-rotation energy transfer,'' J. Chem. Phys. **43**, 2979–2986 (1965).
[20] H. E. Bass, J. R. Olson, and R. C. Amme, ''Vibrational relaxation in H_2O vapor in the temperature range 373–946 K,'' J. Acoust. Soc. Am. **56**, 1455–1460 (1974).
[21] H. K. Shin, ''Vibration-to-rotation energy transfer in water, heavy water, and ammonia,'' J. Phys. Chem. **77**, 346–351 (1973).
[22] H. K. Shin, ''Self-relaxation of vibrationally excited H_2O molecules,'' J. Chem. Phys. **98**, 1964–1978 (1993).
[23] A comprehensive database of past and current studies of water structure is available at www.lsbu.ac.uk/water
[24] J. M. Prausnitz, *Molecular Thermodynamics of Fluid-Phase Equilibria*, Prentice-Hall, New Jersey, 1969, p. 72.
[25] J. D. Lambert, *Vibrational and Rotational Relaxation in Gases*, Clarendon Press, Oxford, 1977.
[26] Ref. 25, p. 46.
[27] A. J. Zuckerwar and W. A. Griffin, ''Resonant tube for measurement of sound absorption in gases at low frequency/pressure ratios,'' J. Acoust. Soc. Am. **68**, 218–226 (1980).
[28] A. J. Zuckerwar and W. A. Griffin, ''Effect of water vapor on sound absorption in nitrogen at low frequency/pressure ratios,'' J. Acoust. Soc. Am. **69**, 150–154 (1981).

A method for estimating the parameters of electrodynamic drivers in thermoacoustic coolers

Insu Paek,[a] Luc Mongeau, and James E. Braun
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 South Intramural Drive, West Lafayette, Indiana 47907-2031

(Received 9 June 2004; revised 29 September 2004; accepted 8 October 2004)

The electroacoustic efficiency of high-power actuators used in thermoacoustic coolers may be estimated using a linear model involving a combination of six parameters. A method to identify these equivalent driver parameters from measured total electrical impedance and velocity-voltage transfer function data was developed. A commercially available, moving-magnet driver coupled to a functional thermoacoustic cooler was used to demonstrate the procedure experimentally. The method, based on linear electrical circuit theory, allowed for the possible frequency and amplitude dependence of the driver parameters to be estimated. The results demonstrated that driver parameters measured *in vacuo* using this method can be used to predict the driver efficiency and performance for operating conditions which may be encountered under load. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1828500]

PACS numbers: 43.35.Ud [RR] Pages: 185–193

LIST OF SYMBOLS

Bl	force constant
f	frequency
f_d	damped natural frequency
f_o	undamped natural frequency
F	force
$H_{u/V}$	velocity-voltage transfer function
I	current
kl	Helmholtz number
K_{gas}	gas spring stiffness
K_m	mechanical stiffness
L_e	coil inductance
M_m	mechanical moving mass
P	pressure
P_o	mean pressure
R_a	real part of S^2Z_a
R_e	electrical resistance
R_m	mechanical resistance
S	piston cross-sectional area
S_p	cross-sectional area of cavity
u	velocity
U	volume velocity
V	voltage
$\overline{V}$	volume
x_o	initial piston displacement
x_p	piston peak displacement
X	imaginary part of Z_{ma}
Z_a	acoustic impedance
Z_b	blow-by impedance
$Z_{e,a}$	equivalent impedance due to acoustic loading in the electrical domain
$Z_{e,m}$	equivalent mechanical impedance with no current in the electrical domain
Z_i	total electrical impedance
Z_e	mechanical-motion blocked electrical impedance
Z_{ma}	$R_m+j(\omega M_m-K_m/\omega)+S^2Z_a$
η	electroacoustic efficiency
ω	angular frequency
γ	specific heat ratio

[a]Electronic mail: paek@ecn.purdue.edu

I. INTRODUCTION

Electrodynamic drivers are used in a class of electrically driven thermoacoustic refrigeration systems.[1] The mechanical and electrical characteristics of the driver, in conjunction with the acoustic load impedance at the driver piston, determine the electroacoustic efficiency of the actuator.[2,3] The electroacoustic efficiency is, of course, a key factor in the overall efficiency of the cooling system. For this reason, it is useful to develop models that allow the efficiency of any such driver to be predicted for varying operating conditions and loads.

A detailed description of linear models of loudspeakers using equivalent electrical circuits is readily available (for example, Ref. 4 or 5). Several methods based on such linear models have been proposed in order to determine the model parameters experimentally. Measurement methods may be categorized as static or dynamic.[6–8] The electrical resistance can be measured, say, using a four-wire technique, a dc ammeter, and a dc power source. The mechanical stiffness can be measured statically using a force gauge, a dc current source, and a position sensor. Dynamic methods use free decay curves of the driver response with known weights added to the piston. Alternately, one may measure the moving mass and characterize the force constant statically using similar procedures as for the stiffness determination. The mechanical damping parameter is normally inferred from free decay curves of the driver response. These measurement methods are not always convenient because they require dedicated experimental setups and instrumentation, with often the need to remove the driver from the cooling system.

Methods for parameter identification without direct measurements have been developed. Most of them have been applied to loudspeakers, which require a uniform frequency response and modest electroacoustic efficiencies over a large frequency range.[9–11] The possible influence of piston displacement amplitude on the parameters has not typically been considered in these studies. Jacobsen[9] obtained driver parameters using the measured voltage response for a step current excitation. The decay curves were found to be only approximately exponential, with unevenly spaced zero crossings. This was attributed to nonlinear effects, which result in amplitude and time-dependent equivalent driver parameters. The parameters identified using this method were accurate within 10% compared with parameter values measured directly on a test bench. Wright[10] suggested an empirical impedance model for the measured electrical impedance. He modeled the electrical resistance and the inductance as exponential functions of frequency, and found the model parameters using a curve fitting technique. A significant increase of the electrical resistance with frequency was observed, reportedly due to eddy current effects. Knudsen *et al.*[11] studied the low-frequency behavior of a loudspeaker using a time-domain system identification method featuring a low-pass filtered square-wave input. This method yielded estimates within 4% compared with directly measured values for all driver parameters, except the coil inductance. It was suggested that the coil inductance could be neglected for drivers having an inductance much smaller than the ratio of the electrical resistance and the angular frequency, due to its relatively small contribution to the overall input impedance.[11] This simplification is not appropriate for power actuators in thermoacoustic systems because the coil inductance is often comparable to the ratio of electrical resistance and angular frequency.

Other efforts to model nonlinear effects have been made to overcome the limitations of linear driver models. Olson[12] used a third-order polynomial to model the nonlinear suspension stiffness with respect to the cone displacement, and solved the associated differential equations of motion. Birt[13] obtained the nonlinear force constant profile with respect to the voice-coil displacement, and the voice-coil current for an assembled loudspeaker driver. He proposed a method to eliminate nonlinearity using a programmable read-only memory in conjunction with a compensating coil and center pole extension. Kaizer[14] used a truncated Volterra series expansion to solve a governing nonlinear differential equation for enclosed loudspeakers having nonlinearities in the electric inductance, the suspension stiffness and the force constant. Reasonable agreement was obtained between the calculated and directly measured responses. Klippel[15] further extended Kaiser's work by considering bass reflex loudspeakers using the same Volterra series expansion. A fair agreement between measured and calculated responses was obtained. More recently, Scott *et al.*[16] proposed a way to determine nonlinear driver parameters using an impedance sweep for a fixed coil displacement, assuming the linear equivalent circuits are accurate in this condition. The measured data was used to obtain seventh-order regression curves for the amplitude-dependent parameters. This method is simpler than Kaiser's and Klippel's methods, yet requires additional measurements of the cone's displacement for a DC current excitation and the force constant at zero coil displacement.

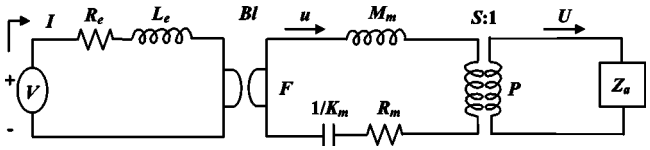


FIG. 1. Linear electrical equivalent circuit of an electroacoustic driver.

In contrast with hi-fi loudspeakers, electrodynamically driven thermoacoustic cooling systems require drivers that can provide much greater acoustic power at one single target frequency, with good efficiency. This problem is well described in a recent investigation of the optimization of electrodynamic drivers for high electroacoustic efficiency using linear models based on equivalent electrical circuits.[17] Drivers in thermoacoustic coolers need to operate only in the vicinity of the target frequency. Therefore it is not necessary to characterize the driver over a wide frequency range.

Methods specific for thermoacoustic coolers have been proposed before. Ballister and McKelvey[18] found the force constant using a three-parameter regression of the impedance measured using an impedance analyzer. Either the moving mass or the mechanical stiffness was required. The frequency and amplitude dependence of the linear parameters was not considered. Smith[7] assumed that the electrical resistance is a linear function of frequency at low frequencies, based on measured total electrical impedance data. This model was found to supply accurate estimates of the measured total electrical impedance of a 2-kW model C-2A CFIC moving magnet driver.

The objective of the present study was to develop a simple method to identify the frequency- and amplitude-dependent driver parameters without time-consuming static or dynamic measurements or elaborate test facilities. The goal was to improve upon Scott *et al.*'s method using a simple model derived from linear electrical equivalent circuit theory. Only electrical impedance and velocity-voltage transfer function data are required for two distinct but similar frequencies, and a fixed piston amplitude. A direct analytical formulation derived from the linear electrical equivalent circuit theory was used to identify all the driver parameters. The parameters obtained *in situ* using this method were compared with the values measured on a dedicated test bench, and values provided by the manufacturer.

II. THEORETICAL BACKGROUND

A schematic of the linear electrical equivalent circuit model for an electroacoustic driver is shown in Fig. 1. The model parameters are R_e, the electrical resistance; L_e, the coil inductance; M_m, the mechanical moving mass; K_m, the mechanical stiffness; R_m, the mechanical resistance; Bl, the force constant; S, the piston cross-sectional area; and Z_a, the complex acoustic impedance at the piston. All parameters except Z_a are assumed to be real. The electrical circuit is connected to the mechanical circuit through a gyrator and the

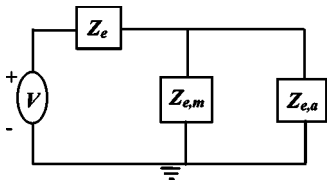


FIG. 2. Electrical equivalent circuit with transformed mechanical and acoustical elements.

mechanical circuit is connected to the acoustical circuit through a transformer. The electrical, mechanical, and acoustical circuits can be brought into the electrical domain and represented as shown in Fig. 2. In Fig. 2, Z_e is the electrical impedance with the mechanical motion blocked, given by

$$Z_e = R_e + j\omega L_e, \tag{1}$$

where ω is the angular frequency.

The equivalent mechanical impedance, $Z_{e,M}$, measured without current in the electrical domain, is given by

$$Z_{e,M} = \frac{(Bl)^2}{R_m + j(\omega M_m - K_m/\omega)}, \tag{2}$$

and $Z_{e,a}$, the equivalent impedance due to the acoustic loading in the electrical domain, is given by

$$Z_{e,a} = \frac{(Bl)^2}{S^2 Z_a}. \tag{3}$$

The total electrical impedance at the driver, Z_i, is then expressed by

$$\begin{aligned} Z_i &= Z_e + \frac{1}{1/Z_{e,M} + 1/Z_{e,a}} \\ &= Z_e + \frac{(Bl)^2}{R_m + j(\omega M_m - K_m/\omega) + S^2 Z_a} \\ &= Z_e + \frac{(Bl)^2}{Z_{ma}}, \end{aligned} \tag{4}$$

where

$$Z_{ma} = R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) + S^2 Z_a. \tag{5}$$

The force provided by the driver is

$$F = Bl \cdot I = Z_{ma} \cdot u, \tag{6}$$

and therefore

$$\frac{u}{I} = \frac{Bl}{Z_{ma}}. \tag{7}$$

Finally, the transfer function between the piston velocity and the voltage input to the driver may be calculated using Eqs. (4) and (7), and is given by

$$H_{u/V} = \frac{u}{I} \cdot \frac{1}{Z_i} = \frac{Bl/Z_{ma}}{Z_e + (Bl)^2/Z_{ma}}. \tag{8}$$

The following expression results from Eq. (4):

$$\frac{Bl}{Z_{ma}} = \frac{Z_i - Z_e}{Bl}. \tag{9}$$

Substitution of Eq. (9) into Eq. (8) yields

$$\begin{aligned} H_{u/V} &= \frac{Bl/Z_{ma}}{Z_e + (Bl)^2/Z_{ma}} = \frac{(Z_i - Z_e)/Bl}{Z_e + [(Z_i - Z_e)/Bl]Bl} \\ &= \frac{Z_i - Z_e}{BlZ_i}, \end{aligned} \tag{10}$$

which also supplies the relation,

$$H_{u/V} Bl Z_i = Z_i - Z_e = Z_i - R_e - j2\pi f L_e. \tag{11}$$

Considering a pair of frequencies f_1 and f_2, with f_1 very close to the frequency of interest f_2, the frequency-dependent electrical resistance, R_e, and inductance, L_e, can be assumed to be constant over the range between f_1 and f_2. Thus the following equation can be obtained from Eq. (11):

$$\begin{aligned} &(H_{u/V}|_{f_1} Z_i|_{f_1} - H_{u/V}|_{f_2} Z_i|_{f_2}) Bl \\ &\quad = Z_i|_{f_1} - Z_i|_{f_2} - j2\pi(f_1 - f_2) L_e. \end{aligned} \tag{12}$$

Finally, the transduction coefficient, Bl, is given by the expression

$$Bl = \text{real}\left(\frac{Z_i|_{f_1} - Z_i|_{f_2} - j2\pi(f_1 - f_2)L_e}{H_{u/V}|_{f_1} Z_i|_{f_1} - H_{u/V}|_{f_2} Z_i|_{f_2}} \right). \tag{13}$$

Although Bl should be a complex number, the lagging phase angle is very close to zero near the driver resonance frequency. In thermoacoustic coolers, the electroacoustic drivers are generally tuned to have resonance around the operating frequency. Therefore Bl was assumed to be a real number in Eq. (13).

Equations for R_e and L_e can then be obtained directly from Eq. (11) and the expression for Bl. These are

$$R_e = \text{real}(Z_i|_{f_2} - H_{u/V}|_{f_2} Bl Z_i|_{f_2}) \tag{14}$$

and

$$L_e = \frac{\text{imag}(Z_i|_{f_2} - H_{u/V}|_{f_2} Bl Z_i|_{f_2})}{2\pi f_2}. \tag{15}$$

The transduction factor, Bl, is initially obtained from Eq. (13) by assuming that the L_e term is negligible compared with other terms. The parameters R_e and L_e are then evaluated using the approximate value for Bl. The calculation is iterated until Bl, R_e, and L_e converge to satisfy Eqs. (13)–(15).

The input impedance of the mechanical circuit with acoustical load is given by

$$\frac{F}{u} = R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) + S^2 Z_a. \tag{16}$$

From the well known relation,

$$F = Bl \cdot I, \tag{17}$$

in conjunction with Eqs. (1) and (2), F may be eliminated from Eq. (16) as follows:

$$R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) = Bl\frac{I}{u} - S^2 Z_a = \frac{Bl}{Z_i H_{u/V}} - S^2 Z_a . \tag{18}$$

Now, R_m can be obtained from the measured transfer functions at f_2 and Bl as follows:

$$R_m = \text{real}\left(\frac{Bl}{Z_i|_{f_2} H_{u/V}|_{f_2}} - S^2 Z_a|_{f_2}\right). \tag{19}$$

The equivalent mass parameter, M_m, and the stiffness, K_m are calculated from Eq. (18) with the help of the measured transfer functions at f_1:

$$M_m = \frac{\text{imag}(Bl/Z_i|_{f_1} H_{u/V}|_{f_1} - S^2 Z_a|_{f_1}) f_1 - \text{imag}(Bl/Z_i|_{f_2} H_{u/V}|_{f_2} - S^2 Z_a|_{f_2}) f_2}{2\pi(f_1^2 - f_2^2)}. \tag{20}$$

Finally, the equivalent stiffness parameter, K_m, is obtained using

$$K_m = 4\pi^2 f_2^2 M_m - \text{imag}\left(\frac{Bl}{Z_i|_{f_2} H_{u/V}|_{f_2}} - S^2 Z_a|_{f_2}\right). \tag{21}$$

All the driver parameters are expressed in closed form in terms of the driving frequencies, f_1 and f_2, the total electrical impedance, Z_i, the velocity-voltage transfer function, $H_{u/V}$, and the acoustic impedance, Z_a with the exception of Bl; the latter is calculated in conjunction with R_e and L_e as previously described. The relations (13)–(15) and (19)–(21), although simple to derive, are not readily available in the literature. The indirect parameter identification method based on these equations will be referred to as ''method A'' in subsequent sections of this paper, for convenience.

According to Wakeland,[17] the electroacoustic efficiency, η, of the driver is related to the driver parameters and the acoustic impedance through

$$\frac{1}{\eta} = \frac{R_e R_m}{(Bl)^2}\frac{R_m}{R_a}\left(1 + \frac{R_a}{R_m}\right)^2 + \left(1 + \frac{R_m}{R_a}\right) + \frac{R_e R_m}{(Bl)^2}\frac{X^2}{R_m R_a}, \tag{22}$$

where $R_a = \text{Re}[S^2 Z_a]$, and $X = \text{Im}[Z_{ma}]$. Equation (22) can be further simplified to

$$\eta = \frac{R_a (Bl)^2}{R_e\{X^2 + (R_m + R_a)^2\} + (Bl)^2 (R_a + R_m)}. \tag{23}$$

Equation (23) was used to calculate electroacoustic efficiency using parameter values inferred from Eqs. (13)–(15) and (19)–(21), i.e., using method A.

III. GUIDANCE IN SELECTING FREQUENCY DIFFERENTIALS

The direct formulation, method A, for the identification of R_e, L_e, and Bl is valid when the two driving frequencies are nearly the same, and the piston displacement is fixed. This is because R_e and L_e are assumed constants, and the term including L_e in Eq. (13) is initially neglected to find Bl. Therefore, small frequency intervals that ensure the imaginary part of the term including L_e in Eq. (13) is much smaller than the imaginary part of its preceding term are essential for estimating R_e, L_e, and Bl.

However, the use of two very similar frequencies could result in large errors in the identification of M_m and K_m if there are small errors in the source frequencies due to the term $f_1^2 - f_2^2$ in Eq. (20). The accuracy of the source frequency from the DSP board used for the experiments is about 0.02 Hz. A simulation was done assuming 0.02-Hz frequency errors to estimate uncertainties in the identified parameters. Relative errors for M_m and K_m of up to 45% can result when a 0.1-Hz frequency differential is used. The uncertainty in K_m was very close to M_m because M_m is included in the equation for K_m. Errors for the other parameters associated with the added 0.02-Hz frequency errors were less than 6%. The relative errors of M_m and K_m were reduced to 2.3% when a 2-Hz frequency differential was used. Based on these results, a 2-Hz frequency increment was selected for M_m and K_m, and a 0.1-Hz frequency increment was used for all the other parameters.

An approximate estimate of the uncertainty in M_m can be developed from Eq. (20). The uncertainty in K_m is approximately the same as that for M_m. By assuming the uncertainty of the numerator is negligible compared with that of the denominator, Eq. (20) can be rewritten as

$$M_m \approx \frac{C}{2\pi(f_1^2 - f_2^2)}, \tag{24}$$

where C is a constant. The uncertainty in M_m is then given by[19]

$$\delta M_m = \sqrt{\left(\frac{\partial M_m}{\partial f_1}\delta f_1\right)^2 + \left(\frac{\partial M_m}{\partial f_2}\delta f_2\right)^2} = \frac{C\sqrt{f_1^2 \delta f_1^2 + f_2^2 \delta f_2^2}}{\pi(f_1^2 - f_2^2)^2}, \tag{25}$$

where δf_1 and δf_2 are uncertainties in f_1 and f_2, respectively. Finally, the relative uncertainty of M_m is

$$\frac{\delta M_m}{M_m} = \frac{2\sqrt{f_1^2 \delta f_1^2 + f_2^2 \delta f_2^2}}{(f_1^2 - f_2^2)}. \tag{26}$$

By assuming f_1 is close to f_2, Eq. (26) can be simplified to

$$\frac{\delta M_m}{M_m} = R\sqrt{2}\frac{f}{\Delta f}, \tag{27}$$

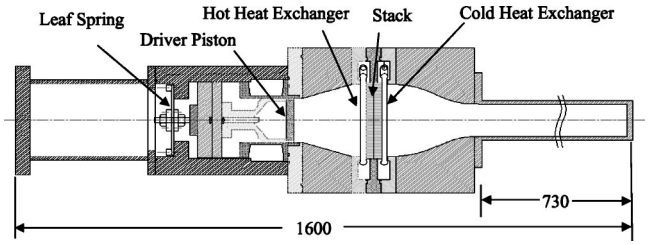


FIG. 3. Schematic of the standing wave thermoacoustic cooler (dimensions in millimeters).

where R is the relative uncertainty of the frequency, f is the frequency of the source, and Δf is the frequency differential. Equation (27) yields 28.9% uncertainty for a 0.1-Hz frequency differential and 1.4% uncertainty for a 2-Hz frequency interval when 170 Hz with the relative uncertainty of 0.012% is used. Although Eq. (27) is approximate, it can be used to select a proper frequency differential for M_m and K_m in the absence of knowledge of other parameters and transfer functions.

IV. EXPERIMENTAL APPARATUS

A thermoacoustic cooler prototype[20] was used to investigate the driver parameter characterization measurement method. A schematic of the thermoacoustic cooler is shown in Fig. 3. The driver was a moving magnet linear actuator (CFIC, model B-300) mounted on an added metal leaf spring to provide additional suspension stiffness. The driver's natural frequency without the added spring was near 33 Hz. The driver was designed to deliver more than 200 W of acoustic power at 33 Hz, with an electro-acoustic transduction efficiency of 70% and a maximum displacement amplitude of 6 mm. The driver's natural frequency was increased to about 163 Hz, and its maximum displacement amplitude was reduced to less than 1 mm due to the added leaf spring. More information on the detailed driver characteristics is available in Ref. 21. The nominal driver parameters without the added spring provided from the manufacturer are given in Table I. The nominal parameters were directly measured by the manufacturer using various methods.[22] The electrical resistance was measured using a four-wire technique, and a dc current from a dc power source. The stiffness and the force constant were measured statically using a force gauge and a dc current source. The mechanical resistance was measured using the free decay response method.

In the present study, an accelerometer (PCB, 353B15) was mounted on the driver piston to measure the axial acceleration. A dynamic pressure sensor (PCB, 102A07) was installed in a port near the piston to measure the driver pressure. An Agilent VXI mainframe (E8403A) and a DSP board (E1432A) were used for data acquisition. The input voltage, the resulting current, the piston velocity, and the driver pressure were measured simultaneously. One source channel was used to operate the moving magnet driver. A data acquisition and display program was developed using Agilent Vee, and was used for the measurement.

TABLE I. Driver parameters provided by manufacturer.

R_e (Ω)	0.11
L_e (mH)	0.86
Bl (N/A)	9
M_m (kg)	1.6
K_m (MN/m)	0.74
R_m (N-s/m)	25

The piston displacement was monitored and the voltage input to the driver was adjusted to keep the piston displacement constant during the measurements. Seven frequencies in the range from 168 to 174 Hz, which encompassed the maximum electroacoustic efficiency of the driver, were chosen. The transfer functions were measured at frequencies both 0.1 and 2 Hz below and above the chosen frequencies, as well as at the chosen frequency. Two different sets of parameters were identified for each chosen frequency using forward and backward frequency steps. The two values were averaged.

V. BENCH MEASUREMENTS OF DRIVER PARAMETERS

To verify the parameters calculated using method A, the driver parameters were also measured directly with the exception that manufacturers' electrical inductance was used.

The undamped natural frequency, f_o, is related to the moving mass and the mechanical stiffness through

$$f_o^2 = \frac{1}{4\pi^2}\frac{K_m}{M_m}. \tag{28}$$

The relation between the undamped and the damped natural frequency is

$$f_o = f_d\sqrt{1+\left(\frac{R_m}{4\pi f_d M_m}\right)^2}, \tag{29}$$

where f_d is the damped natural frequency.[23] The free decay response of the driver was measured after adding a known mass to the piston, assuming that damping was small enough not to affect the natural frequency. The second term in the square root in Eq. (29) is much smaller than one, and f_o is approximately equal to f_d. This allowed the moving mass and the mechanical stiffness to be estimated.

The mechanical resistance was obtained from measured decay curves of the piston displacement, x, which can be described by

$$x = x_o e^{-R_m/2M_m}\cos(2\pi f_d t), \tag{30}$$

where x_o is the initial piston displacement. The force constant, Bl, was obtained from the measured free decay curves of the acceleration and the voltage. The voltage leads the acceleration by 90°, and the ratio of the voltage and the integrated acceleration yields the force constant when the circuit is electrically open. This can be verified from Eq. (8) when Z_e is zero.

The electrical resistance, R_e, was obtained from the electrical impedance measured at frequencies far different from the natural frequency of the driver. For operating frequencies either much greater or smaller than the natural frequency, the real part of the electrical impedance is about the same as the electrical resistance. This can be verified from

TABLE II. Driver parameters obtained from method B.

	0.2 mm	0.3 mm	0.4 mm
$R_e\,(\Omega)$	0.2	0.2	0.2
Bl	8.5	8.6	8.6
M_m (kg)	1.71	1.71	1.71
K_m	1.78	1.78	1.78
R_m	47.2	37.6	35.5

Eq. (4). The measured electrical impedance values above and below the natural frequency were linearly interpolated to find the frequency-dependent electrical resistance.

The direct method described above will subsequently be referred to as ''method B,'' to distinguish it from ''method A'' described in Sec. II. The driver parameters measured with method B for three different piston displacements are given in Table II. The large mechanical stiffness value in Table II compared to that in Table I is due to the addition of a leaf spring to the driver. The electrical resistances were arithmetically averaged over the frequency range between 168 and 172 Hz. The mechanical resistance decreased with increasing piston displacement, as discussed later.

VI. RESULTS

A. Driver parameter estimation

The total electrical impedance and the velocity-voltage transfer functions were measured *in vacuo* over the frequency range from 168 to 174 Hz for different fixed piston peak displacements, x_p, of 0.2, 0.3, and 0.4 mm, respectively. The six driver parameters were then calculated using method A.

The results are shown in Figs. 4–6. The identified electrical resistance and the inductance increase with frequency. The trend of an exponential increase has been reported and attributed to eddy current effects.[24] Wright also observed this trend for frequencies up to 20 kHz.[10] At low frequencies, the increase can be approximated as linear.[7] The curves in Fig. 4 indicate a nearly linear trend. The resistance values, R_e, obtained for the frequencies of interest are much larger than the dc resistance measured by the manufacturer (Table I). Therefore the maximum electroacoustic efficiency of the driver calculated with the nominal dc electrical resistance using the usual method[17] may not be achieved for high frequencies.

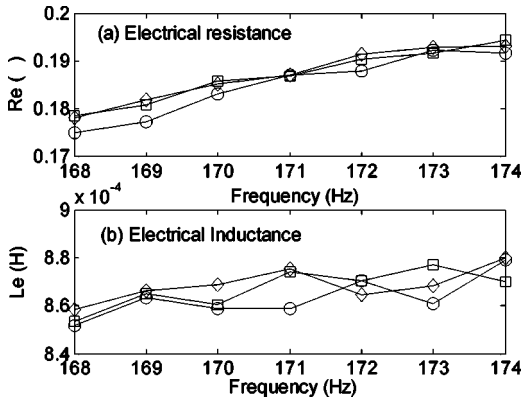


FIG. 4. Driver equivalent parameters identified using method A versus frequency: (a) electrical resistance and (b) electrical inductance. ○: $x_p = 0.2$ mm, □: $x_p = 0.3$ mm, and ◇: $x_p = 0.4$ mm.

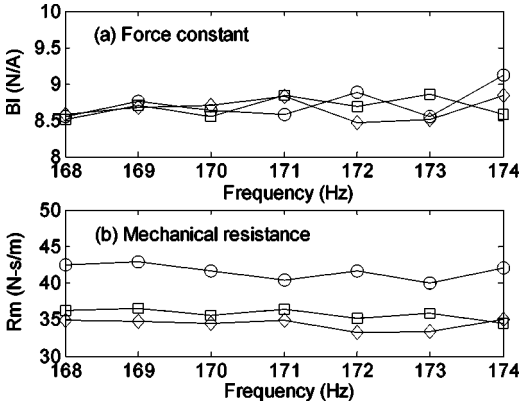


FIG. 5. Driver equivalent parameters identified using method A versus frequency: (a) force constant and (b) mechanical resistance. ○: $x_p = 0.2$ mm, □: $x_p = 0.3$ mm, and ◇: $x_p = 0.4$ mm.

The force constant, Bl, and the mechanical resistance, R_m, are shown in Fig. 5. The force constant increases slightly with the piston amplitude, and asymptotically approaches a constant for large piston displacements. Although not shown in Fig. 5, Bl is expected to sharply decrease with piston displacements for much larger piston displacements.[7,13] This occurs when the moving magnet oscillation displacement is much larger than the stationary electromagnet structure, and thus portions of the moving magnet and the stator are less effective in creating a magnetic flux within the gap.[13]

The resistance, R_m, was obtained using Eq. (30) applied for added mass tests. The mechanical resistance R_m decreased as the piston displacement increased. The mechanical resistance lumps all dissipative phenomena such as internal damping in the spring element, or losses in the voice-coil centering mechanism (moving magnet in this case).[4] Based on Fig. 5, both Bl and R_m are clearly functions of piston displacement, but they are almost independent of operating frequency. This is in apparent contradiction with Smith's experimental observations, which report a decrease of R_m with increasing frequency.[7]

The identified spring stiffness was amplitude and fre-

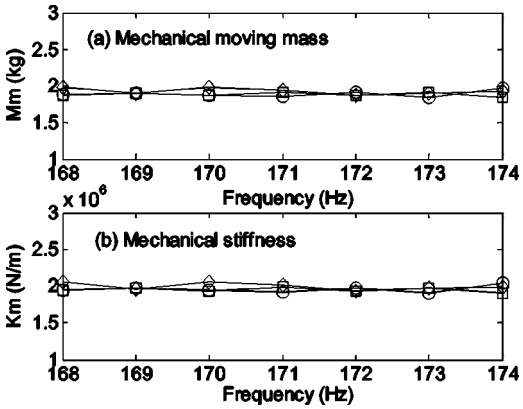


FIG. 6. Driver equivalent parameters identified using method A versus frequency: (a) mechanical moving mass and (b) mechanical stiffness. ○: $x_p = 0.2$ mm, □: $x_p = 0.3$ mm, and ◇: $x_p = 0.4$ mm.

TABLE III. Frequency-averaged driver parameters obtained from method A.

	0.2 mm	0.3 mm	0.4 mm
R_e (Ω)	0.19	0.19	0.19
L_e (mH)	0.86	0.87	0.87
Bl	8.7	8.7	8.7
M_m (kg)	1.89	1.88	1.92
K_m	1.96	1.95	2.00
R_m	42	36	34

quency independent, as shown in Fig. 6. This is in contrast with earlier observations.[11,12] The values of the mechanical stiffness obtained using method B (Table II) did not increase with amplitude, either. The amplitude-dependent characteristics of the spring may be truly linear because the piston displacement is small due to the added leaf spring. Both the mechanical stiffness and the moving mass obtained with method A were within 15% of the values obtained with method B. The effective moving mass was also amplitude and frequency independent as shown in Fig. 6.

B. Electroacoustic efficiency predictions

The equivalent parameters obtained using method A were arithmetically averaged over the frequency span, neglecting the effects of frequency, for electroacoustic efficiency predictions. A model for the effects of a back cavity is needed in addition to the parameters measured in vacuum to predict the driver efficiency in a pressurized gas mixture. When the product kl is much less than unity, the fluid in the back cavity can be considered as a lumped element.[25] Therefore the gas spring stiffness can be calculated using

$$K_{gas}=\frac{\gamma\cdot P_o}{V}S_p^2, \tag{31}$$

where γ is the specific heat ratio, P_o is the mean pressure of the gas, V is the volume of the cavity, and S_p is the cross-sectional area of the cavity. The calculated gas-spring stiffness was about 10 000 N/m. This value was added to the mechanical stiffness in the model. The frequency-averaged parameters combined with the measured acoustic impedance yield the electroacoustic efficiency of the driver. The frequency-averaged parameters are shown in Table III.

The electroacoustic efficiency was calculated using Eq. (23), with identified parameter values from method A, and the acoustic load impedance measured in the pressurized system. The calculated efficiency was compared with directly measured efficiencies for operating conditions near the resonance frequency. The electroacoustic efficiency was also calculated from parameters obtained using method B, i.e., the parameter values of Table II, along with the nominal electrical inductance of Table I. Figures 7–9 show the predicted and the measured efficiency with respect to frequency for 0.2, 0.3, and 0.4 mm piston peak displacements, respectively. For all cases, the efficiency predictions using parameters obtained from method A yielded less than 17% relative errors from the measured efficiency. The efficiency calculated using the parameters obtained from method B also yielded efficiency predictions within 17%. For a 0.4-mm piston amplitude, the efficiency calculated from the frequency-averaged equivalent parameters using method A yielded a slightly better estimate of the measured efficiency than that calculated from method B.

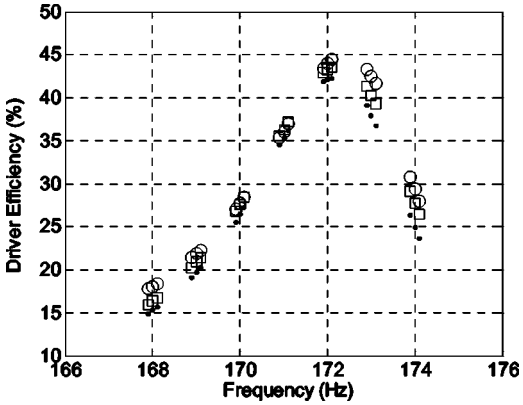


FIG. 8. Electroacoustic driver efficiency versus frequency when $x_p = 0.3$ mm. ●: calculated efficiency using the averaged identified parameters obtained from method A in vacuum, ○: measured efficiency in pressure, and □: calculated efficiency using the parameters obtained from method B.

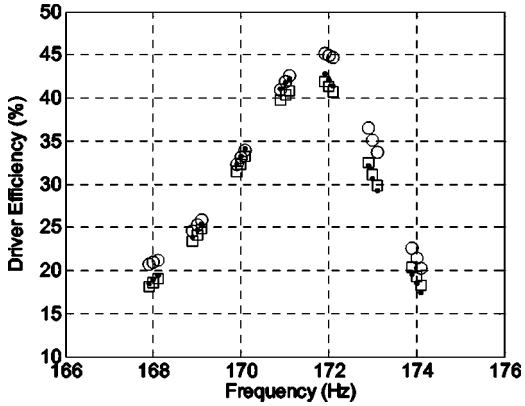


FIG. 7. Electroacoustic driver efficiency versus frequency when $x_p = 0.2$ mm. ●: calculated efficiency using the averaged identified parameters obtained from method A in vacuum, ○: measured efficiency in pressure, and □: calculated efficiency using the parameters obtained from method B.

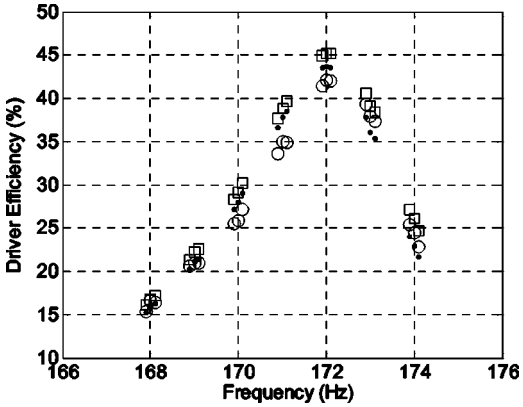


FIG. 9. Electroacoustic driver efficiency versus frequency when $x_p = 0.4$ mm. ●: calculated efficiency using the averaged identified parameters obtained from method A in vacuum, ○: measured efficiency in pressure, and □: calculated efficiency using the parameters obtained from method B.

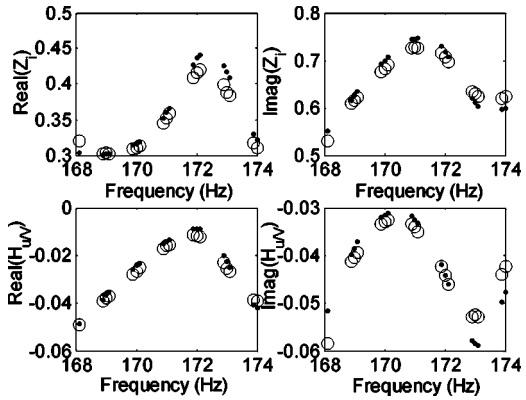


FIG. 10. Measured and calculated transfer functions when $x_p = 0.4$ mm. (a) Real part of total electrical impedance; (b) imaginary part of total electrical impedance; (c) real part of velocity-voltage transfer function; and (d) imaginary part of velocity-voltage transfer function. ●: measured transfer functions in pressure, ○: calculated transfer functions using the averaged identified parameters obtained from method A in vacuum and the measured acoustic impedance in pressure.

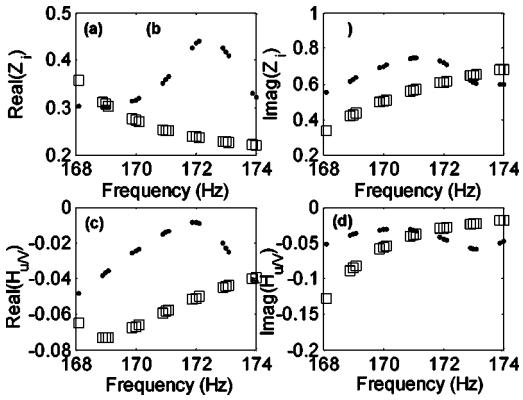


FIG. 12. Measured and calculated transfer functions when $x_p = 0.4$ mm and the blow-by impedance is zero. (a) Real part of total electrical impedance; (b) imaginary part of total electrical impedance; (c) real part of velocity-voltage transfer function; and (d) imaginary part of velocity-voltage transfer function. ●: measured transfer functions in pressure and □: calculated transfer functions using the averaged identified parameters obtained from method A in vacuum and the measured acoustic impedance in pressure.

VII. DISCUSSION

A. Effects of piston leakage

The transfer functions measured in the pressurized system for a 0.4-mm piston displacement are compared with the same quantities calculated from *in vacuo* parameters identified using method A in Fig. 10. The calculated values were slightly underpredicted compared to the measured values around the frequency of peak electroacoustic efficiency. One possibility that could explain the slight discrepancy is the blow-by flow that might occur through a small gap between the piston and the piston housing. To investigate this phenomenon, an unknown blow-by impedance, Z_b, was added in parallel to the acoustic impedance, as shown in Fig. 11.

If the acoustic impedance, Z_a, is zero (like in vacuum), all the flow occurs through u_1, and there is no blow-by flow, i.e., $u_2 = 0$. When Z_b is much larger than $S^2 Z_a$, all the flow occurs through u_1, and this again corresponds to no blow-by flow. For an arbitrary, nonzero acoustic load, Z_a, flow through, u_2, and an additional black box impedance, Z_b, contribute to the measured impedance.

The total electrical impedance and the velocity-voltage transfer function including the unknown impedance are given by

$$Z_i = Z_e + \frac{(Bl)^2}{Z_{mb}}, \tag{32}$$

$$H_{u/V} = \frac{Bl}{Z_e Z_{mb} + (Bl)^2}, \tag{33}$$

where

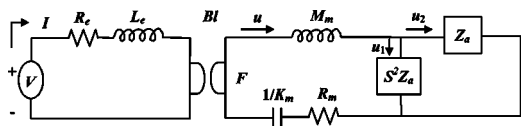


FIG. 11. Linear electrical equivalent circuit of an electroacoustic driver including blow-by flow.

$$Z_{mb} = R_m + j\left(\omega M_m - \frac{K_m}{\omega}\right) + \frac{S^2 Z_a \times Z_b}{S^2 Z_a + Z_b}. \tag{34}$$

The identified parameters in vacuum for a 0.4-mm piston displacement were substituted into Eqs. (32) and (33) and compared with the measured transfer functions in the pressurized system. The blow-by impedance, Z_b, was artificially adjusted to see its effect on the transfer functions. Figure 12 shows the results for Z_b equal to zero. This means that the blow-by impedance is much smaller than the acoustic impedance, and all the flow occurs through the blow-by impedance. The fluid in front of the piston is not compressed by the oscillating piston but flows back and forth through the gap between the piston and the piston housing. As expected, there is no peak in the total electrical impedance due to the acoustic resonance. The same results can be obtained when Z_a is zero, as occurs in a vacuum. When the acoustic impedance is not zero, and the blow-by impedance, Z_b, is twice $S^2 Z_a$, the calculated transfer functions approaches the shape of the measured transfer functions as shown in Fig. 13. When Z_b reaches about ten times $S^2 Z_a$, the plot of the measured and the calculated transfer functions are the same as that without blow-by flow as shown in Fig. 10. This implies that, if blow-by occurs, the calculated transfer function values would have been greater than the measured values, and not smaller as observed in Fig. 10. Based on this model, it was concluded that piston leakage effects were insignificant, and that other unknown reasons were responsible for the discrepancies in the model predictions.

VIII. CONCLUSIONS

A method to identify the driver parameters was developed based on the assumption that the driver parameters are constant for two closely spaced, distinct frequencies and for a fixed piston amplitude. The total electrical impedance and the velocity-voltage transfer function were simultaneously measured to enable the evaluation of the six driver parameters in terms of known quantities. The parameter identifica-

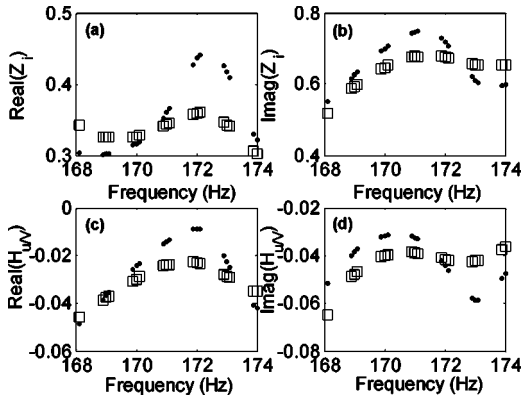


FIG. 13. Measured and calculated transfer functions when $x_p = 0.4$ mm, and the blow-by impedance is twice S^2Za. (a) Real part of total electrical impedance; (b) imaginary part of total electrical impedance; (c) real part of velocity-voltage transfer function; and (d) imaginary part of velocity-voltage transfer function. ●: measured transfer functions in pressure and □: calculated transfer functions using the averaged identified parameters obtained from method A in vacuum and the measured acoustic impedance in pressure. V/I: total electrical impedance.

tion method was applied to the case of a moving magnet electrodynamic driver used in a thermoacoustic cooler for three different piston peak displacements over the frequency range for maximum driver efficiency. The calculated equivalent parameters clearly showed the dependency of some parameters on the piston displacement and the operating frequency. R_e and L_e increased with frequency almost linearly, and R_m decreased with piston amplitude. Bl and K_m are typically amplitude-dependent, but they did not exhibit an amplitude dependence in the present study. This is because the piston amplitudes were very small due to the added leaf spring. The averaged identified driver parameters were used to calculate the electroacoustic efficiency in the pressurized vessel. These supplied a good estimate of the measured electroacoustic efficiency, with relative errors within 17%.

The proposed driver parameter identification method is useful in estimating driver parameters and efficiency at an amplitude and frequency of interest. A knowledge of these parameters is useful in achieving optimal electroacoustic efficiency.

[1] G. W. Swift, *Thermoacoustics: A unifying perspective for some engines and refrigerators* (Acoustical Society of America, New York, 2002).
[2] Y. Li, B. L. Minner, G. Chiu, L. Mongeau, and J. E. Braun, ''Adaptive tuning of an electrodynamically driven thermoacoustic cooler,'' J. Acoust. Soc. Am. **111**, 1251–1258 (2002).
[3] Y. Li, I. Paek, G. Chiu, L. Mongeau, and J. E. Braun, ''Reply to comment on 'Adaptive tuning of an electrodynamically driven thermoacoustic cooler,'" J. Acoust. Soc. Am. **115**, 976–979 (2004).
[4] F. V. Hunt, *Electroacoustics: The Analysis of Transduction, and its Historical Background* (Acoustical Society of America, New York, 1982). (Original work published by Wiley, New York, 1954).
[5] L. L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1982). (Original work published by McGraw–Hill, New York, 1954).
[6] D. A. Harris and R. E. Volkert, ''Design and calibration of an electrodynamic driver for the space thermoacoustic refrigerator,'' M.S. thesis, Physics Department, Naval Postgraduate School, 1989.
[7] R. W. Smith, ''High efficiency two kilowatt acoustic source for a thermoacoustic refrigerator,'' M.S. thesis, Engineering Science and Mechanics Department, The Pennsylvania State University, 2000.
[8] J. F. Heake, ''Characterization of a 10-kilowatt linear motor/alternator for use in thermoacoustic refrigeration,'' M.S. thesis, Acoustics Department, The Pennsylvania State University, 2001.
[9] O. Jacobsen, ''Measuring loudspeaker constants by a transient method,'' J. Audio Eng. Soc. **30**, 112–116 (1982).
[10] J. R. Wright, ''An empirical model for loudspeaker motor impedance,'' J. Audio Eng. Soc. **38**, 749–754 (1990).
[11] M. H. Knudsen, J. G. Jensen, V. Julskjaer, and P. Rubak, ''Determination of loudspeaker driver parameters using a system identification technique,'' J. Audio Eng. Soc. **37**, 700–708 (1989).
[12] H. F. Olson, ''Analysis of the effects of nonlinear elements upon the performance of a back-enclosed, direct radiator loudspeaker mechanism,'' J. Audio Eng. Soc. **10**, 156–162 (1962).
[13] D. R. Birt, ''Nonlinearities in moving-coil loudspeakers with overhung voice coils,'' J. Audio Eng. Soc. **39**, 219–231.
[14] A. J. M. Kaizer, ''Modeling of the nonlinear response of an electrodynamic loudspeaker by a Volterra series expansion,'' J. Audio Eng. Soc. **35**, 421–433 (1987).
[15] W. Klippel, ''Dynamic measurement and interpretation of the nonlinear parameters of electrodynamic loudspeakers,'' J. Audio Eng. Soc. **38**, 944–955 (1990).
[16] J. Scott, J. Kelly, and G. Leembruggen, ''New method of characterizing driver linearity,'' J. Audio Eng. Soc. **44**, 258–265 (1996).
[17] R. W. Wakeland, ''Use of electrodynamic drivers in thermoacoustic refrigerators,'' J. Acoust. Soc. Am. **107**, 827–832 (2000).
[18] S. C. Ballister and D. J. McKelvey, ''Shipboard electronics thermoacoustic cooler,'' M.S. thesis, Physics Department, Naval Postgraduate School, 1995.
[19] J. R. Taylor, *An Introduction to Error Analysis* (University Science Books, Sausalito, CA, 1997), Eq. (3.47).
[20] L. Mongeau, A. Alexander, B. L. Minner, I. Paek, and J. E. Braun, ''Experimental investigations of an electro-dynamically driven thermoacoustic cooler,'' in *Proceedings of the 2001 International Mechanical Engineering Congress and Exposition* (American Society of Mechanical Engineers, New York, 2001), pp. 1–12.
[21] G. A. Yarr and J. A. Corey, ''Linear Electrodynamic Machine,'' U.S. Patent No. 5389844. February 14, 1995.
[22] Private communication.
[23] L. Meirovitch, *Elements of Vibration Analsys* (McGraw–Hill, New York, 1986), Sec. 1.7.
[24] J. R. Bowler, N. Harfield, and N. P. Merricks, ''A theoretical analysis of Eddy-current effects in loudspeaker motors,'' J. Audio Eng. Soc. **48**, 668–678 (2000).
[25] D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000), Sec. 3C.

Numerical description of the edge mode at the beveled extremity of a plate

Nicolas Wilkie-Chancellier, Hugues Duflo, Alain Tinel, and Jean Duclos
Laboratoire d'Acoustique Ultrasonore et d'Electronique, L.A.U.E.—CNRS (UMR 6068), Université du Havre, IUT, Place Robert Schuman, 76610 Le Havre, France

(Received 15 March 2004; revised 13 October 2004; accepted 14 October 2004)

This paper deals with the reflection of a two-dimensional harmonic Lamb wave at the beveled end of a plate. The existence of a resonant edge mode is described by a numerical model. It is proved that the edge mode is the resonance of different complex modes. The behavior of this mode as a function of the bevel angle is studied. Its amplitude decreases and its resonance frequency shifts as the bevel angle decreases from 90 to 85 deg. An unexpected strong variation of the repartition of the reflected energy is linked to this phenomenon. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1828675]

PACS numbers: 43.35.Pt [ANN] Pages: 194–199

I. INTRODUCTION

Lamb waves are often used in the nondestructive testing of structures because they theoretically propagate without attenuation. Much work has been devoted to the reflection of a Lamb wave normally incident at the free edge of a plate and to its conversion in reflected modes. A particularity appears when the incident wave is the symmetric S_0 mode: the generation of the edge mode. This phenomenon has already been observed[1–5] on various materials. For example, Shaw observed the edge resonance of a thick barium titanate disk. An explanation of this phenomenon, in harmonic conditions, has been given by several authors who describe the displacements inside a semi-infinite plate as a sum of the modes of the infinite plate verifying the Lamb wave equation. Gazis and Mindlin[4] obtained an approximate solution, assuming only three modes at the end of the plate in addition to the incident wave. Torvik[1] increased the number of modes to 21 and obtained a more precise description of a phenomenon called edge mode. A study was carried out by Auld and Tsao[2] using a variational method. In these previous works, reducing the computation time was a priority so analytic calculations were used as well as the property of orthogonality of Lamb modes. The mode coefficients are the solution to a simple homogeneous linear system.

These studies show that the edge mode is a large resonance at the end of the plate if several conditions are verified: The incident wave must be the first symmetric Lamb mode S_0 and must have a given frequency. The edge mode is equivalent to the thickness resonance of complex modes which have a large amplitude at the end of the plate.

In this paper we use a collocation method initiated by Shen *et al.*[6] The starting point is also the modal theory used to describe the displacements in the plate. This method involves a linear system of numerous equations and unknowns. As a consequence, the computation time increases and the convergence of results must be examined carefully. The characteristics of the reflected waves are obtained with a good accuracy checked by the calculation of the balances of energy.

The edge mode is predicted at the straight end of a stainless steel plate because the complex modes have a high amplitude if the frequency thickness product is close to 2.3 MHz mm. Last, the influence of the bevel angle on the generation of the edge mode is shown. We prove that the edge mode decreases gradually and show an unexpected reflection of the incident wave.

II. THEORETICAL COMPUTATION OF REFLECTED WAVES

The study of a Lamb wave conversion has been initiated by Torvik[1] who assumed that a harmonic Lamb wave which propagates in a semi-infinite free plate was converted into Lamb waves at the straight cut. In this paper, the modal theory used by Shen *et al.*[6] and Zhang *et al.*[7] is extended to the case of a beveled plate (Fig. 1). The waves propagating in the x_1 direction, incident or reflected, verify the dispersion equation of Lamb waves and have different symmetries (symmetric or antisymmetric). Their analytic expression is as follows:

$$\underline{u}_i = \sum_{m=1}^{4} \underline{U}_{\mathrm{im}} \exp[j(-\underline{K}_1 x_1 + \underline{K}_{2m} x_2 + \omega t)]$$

$$= \varphi_i(x_2) \exp[j(-\underline{K}_1 x_1 + \omega t)], \quad (i = 1,2).$$

In an infinite plate of thickness E, real, complex, and purely imaginary harmonic modes can exist, depending on whether their wave vector component $\underline{K}_1 = K_1' + jK_1''$ is real, complex, or purely imaginary at the frequency F. An incident wave has $K_1' > 0$ and a reflected one has $K_1' < 0$. A finite number of real solutions to the equation of dispersion has been found for a given frequency-thickness product FE. These solutions are the Lamb waves. An infinite number of complex modes also exist. The dispersion curves of all these modes (Lamb, imaginary, and complex modes) are shown in Fig. 2 for FE products ranging from 1 to 4 MHz mm. Only the complex modes, whose real part K_1' is positive, are represented, and the label "C_{1s}" (for example) is associated to a pair of modes (K_1' and $-K_1'$).

In order to describe each mode at a given frequency-thickness product, a program successively computes the

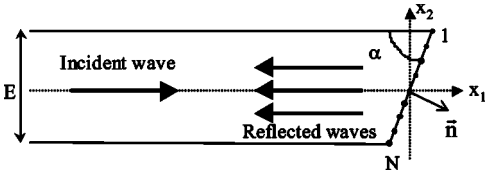


FIG. 1. Geometry of the beveled plate.

wave number $\underline{K}_1$, the amplitudes $\underline{U}_i(x_1,x_2,t)$, the stresses $\underline{T}_{ij}(x_1,x_2,t)$ (i and $j=1, 2$), and the Poynting vector flow Φ through a straight section of the plate (set arbitrarily at one meter wide in the x_3-direction).

Then a boundary condition must be imposed at the end of the plate. This condition is the nullity of the components $\underline{F}_1$ and $\underline{F}_2$ of the force $\underline{F}$ due to the modes on the inclined section.[1] These components depend on the stresses $\underline{T}_{11}$, $\underline{T}_{12}$, $\underline{T}_{21}$, and $\underline{T}_{22}$ and the components n_1, n_2 of the normal to the beveled surface (Fig. 1)

$$\begin{cases} \underline{F}_1 = \underline{T}_{11} n_1 + \underline{T}_{12} n_2 = \underline{T}_{11} \sin(\alpha) - \underline{T}_{12} \cos(\alpha) = 0 \\ \underline{F}_2 = \underline{T}_{21} n_1 + \underline{T}_{22} n_2 = \underline{T}_{21} \sin(\alpha) - \underline{T}_{22} \cos(\alpha) = 0 \end{cases}.$$

The boundary conditions at the beveled edge of the plate are expressed with a sufficiently large number of reflected modes. If the real modes only are taken into account, energy conservation cannot be verified. The nonreal modes must therefore be added in the expression of the stresses. Their wave vector component $\underline{K}_1$ has a real part K_1' positive or negative and an imaginary part K_1'' positive and weak (in order to impose a long decay in the negative x_1 direction). The determination of the wave amplitudes is carried out using a collocation method. Weighting coefficients are assigned to the wave displacements,[8] respectively, 1 to the incident wave and r_m to the mth-additional mode. The stresses of the incident and reflected waves are numerically computed at N points uniformly distributed on the inclined extremity (Fig. 1). So, M modes are added to the incident mode and the stresses of these modes are calculated at the N points. The zero-stress condition at the beveled extremity is N times written as follows:

$$\begin{cases} \underline{F}_1^{\text{inc}}(i) + \sum_{m=1}^{M} \underline{r}_m \underline{F}_1^m(i) = 0 \\ \underline{F}_2^{\text{inc}}(i) + \sum_{m=1}^{M} \underline{r}_m \underline{F}_2^m(i) = 0 \end{cases}, \quad 1 \leqslant i \leqslant N.$$

These $2N$ equations can be written as a linear system

$$F + Ar = 0,$$

where F is the column vector $(2N*1)$ associated to the incident wave, A is the $(2N*M)$ stress matrix, and r is the $(M*1)$ column vector of the r_m coefficients. The solution is written in the least-squares meaning

$$r = -(A^T A)^{-1} A^T F.$$

In this expression, the T and -1 exponents denote the transpose and inverse of a matrix.

In order to perform a quantitative study, the energies of the reflected waves are computed and summed. Only reflected real modes have to be taken into account in this sum because the complex modes do not transport energy. Indeed there are no other waves that Lamb waves at a large distance from the end of the plate. If P_m and P_{inc} are, respectively, the Poynting vector flow of the mth reflected and the incident modes ($P_m = |\underline{r}_m|^2 \cdot \Phi_m$ and $P_{\text{inc}} = \Phi_{\text{inc}}$), the reflection coefficients can be written: $R_m = P_m / P_{\text{inc}} = |\underline{r}_m|^2 \cdot \Phi_m / \Phi_{\text{inc}}$. So the energy balance to check is written $P_{\text{inc}} = \Sigma_{m=1}^{M_L} R_m = 1$, if M_L is the number of reflected Lamb waves. To ensure that this equality is accurate enough, the number M ($M < 2N$) of additional modes must be increased until each R_m reaches a stable value. The number M_L of Lamb modes is equal to 3 at the FE products studied.

Stainless steel plates (velocities $c_L = 5850\,\text{m s}^{-1}$ and $c_T = 3150\,\text{m s}^{-1}$) are used for the computations. Only one result is presented here: The S_0 Lamb mode is incident on a straight end ($\alpha = 90$ deg). There are no reflected antisymmetric Lamb modes (A_0 and A_1) because a change of mode symmetry cannot exist. For the computations, the number N of points must be quite large in order to obtain a good sampling of all the wave displacements. The energy balances obtained with $N = 800$ points and $M = 45$ modes are presented in Fig. 3 when the frequency-thickness products are

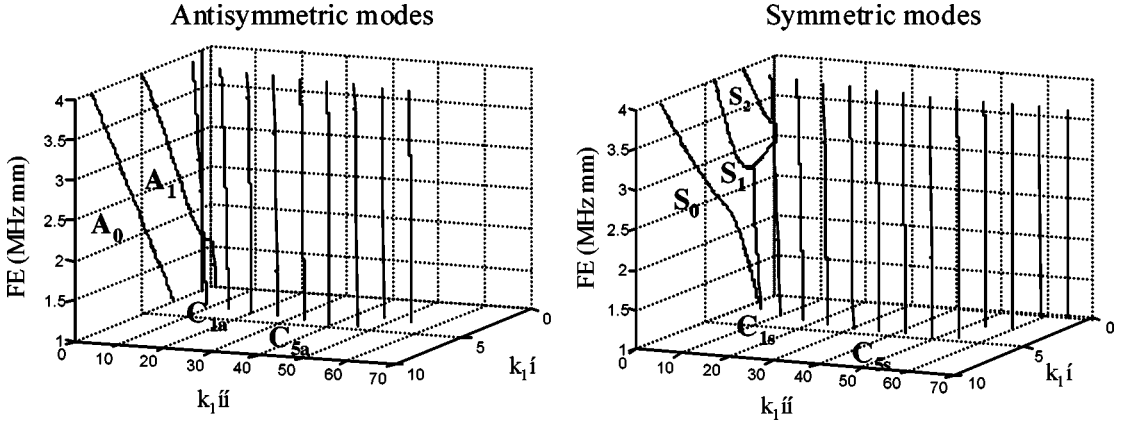


FIG. 2. Dispersion curves in the complex plane versus the FE product.

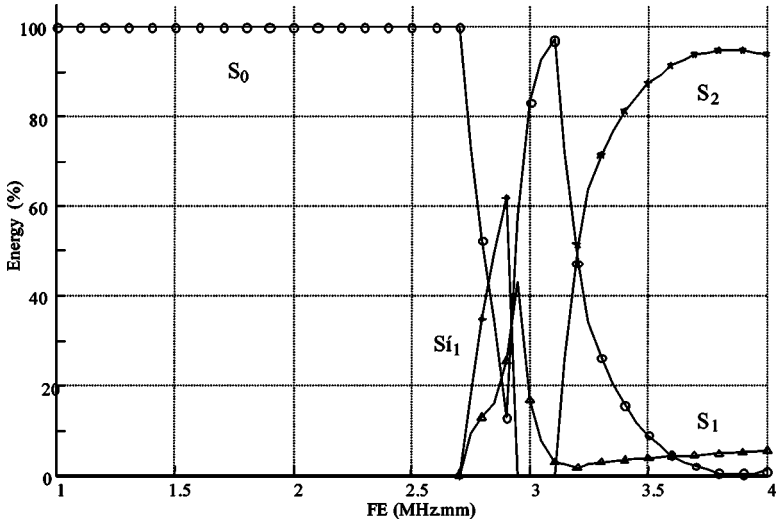


FIG. 3. Energy balance versus FE (S_0 incident mode and α=90 deg).

included between 1 and 4 MHz mm in 0.05 MHz mm step. So the accuracy of the results is better than 1% Under these conditions, the R_{S_0} coefficient, associated to the reflected S_0 mode, is equal to 1 if the S_1 mode is absent ($FE < 2.72$ MHz mm).

The complete study of the wave reflection has been presented in our previous studies[9] as a function of the frequency-thickness product FE (FE is less than 4 MHz mm) and the α angle of the bevel for several incident Lamb waves (from 60° to 90°). As predicted by the modal theory, the results must have the same accuracy whatever α but we need a larger number of modes to achieve a correct result if the α angle is less than 90°.

III. EDGE MODE AT THE STRAIGHT END OF A PLATE

A. Search of the edge mode

No phenomenon related to the existence of the edge mode appears in these studies (see Fig. 3) but computation errors are detected when the incident wave is the S_0 mode, when the FE product is close to 2.31 MHz mm and if the number of modes taken into account is too small ($M=45$ modes for example). The obtained energy balances are then false. The accuracy of the computation improves if M is at least twice larger, i.e., $M=100$ and the energy balances become correct. The edge mode exists precisely for the FE values and conditions mentioned above.

Since the investigations performed by Torvik, it has been known that the edge mode was a vibration imposed by the S_0 harmonic incident wave at a particular frequency. Under these conditions, the amplitudes of the C_{1s} and C_{2s} complex modes clearly exceed those of the incident wave.[4,5] We have computed the amplitude of the different waves present at the extremity. In Fig. 4, the normalized amplitudes of the first complex modes at the extremity are represented as a function of FE. A normalized amplitude is the ratio of the largest amplitude of a complex mode to the largest amplitude of the S_0 reflected wave anywhere in the plate thickness. The C_{1s} modes have an amplitude reaching 35 times those of the S_0 mode at $FE=2.3162$ MHz mm. This resonance is rather sharp (its half middle width is equal to 4.4 kHz mm).

B. Waves at the end of the plate

Figure 5 shows the internal displacements of the various waves which contribute to the edge mode at the resonance frequency ($FE=2.3162$ MHz mm) in a two millimeter thick plate. They are the C_{1s}, C_{2s}, and C_{3s} complex modes and the S_0 reflected wave. The displacements of the following complex waves ($C_{4s}, C_{5s}, \ldots,$) oscillate in the thickness of the plate but with lesser amplitudes. The larger the mode index, the higher the oscillation number.

The edge mode is especially due to the C_{1s} mode oscillations. A physical description can be given for the two C_{1S} modes (K_1' and $-K_1'$). They give rise to standing waves: In the x_1 direction (with attenuation as x_1 decreases) and in the x_2 direction ($x_2=0$ is the symmetry axis). All the complex modes give similar standing waves but with lesser amplitudes. The incident and reflected S_0 Lamb waves are associated in exact standing waves in the x_1 direction (they have same amplitudes and are not attenuated).

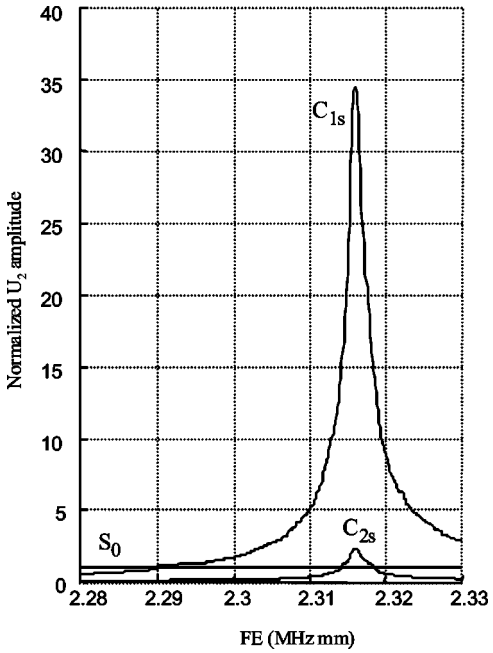


FIG. 4. Maxima of the reflected modes at α=90° (U_2 displacement).

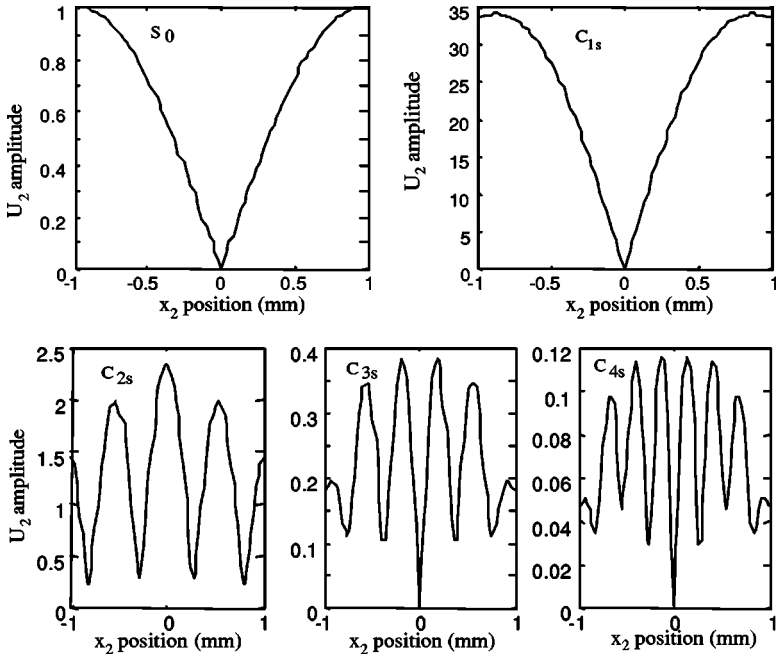


FIG. 5. Normalized U_2 displacement of the S_0, C_{1s}, C_{2s}, C_{3s}, and C_{4s} modes ($FE=2.316$ MHz mm, $\alpha=90°$).

C. Vibrations on the surface at the resonance

The global displacement can be computed (all the waves being taken into account) at the surface of the plate starting from its end. The amplitude of the U_2 displacement component is calculated over the first fifteen millimeters from the bevel (Fig. 6). This makes it possible to visualize the stationary waves produced by the two S_0 waves (incident and reflected). Indeed, beyond a distance larger than 6 mm, the signal becomes periodic. The periodicity is 1.84 mm in the considered stainless steel, which is the S_0 half wavelength $\Lambda_{S_0}/2$ at $FE=2.3162$ MHz mm. Close to the end of the plate, the complex modes have a larger contribution. As expected, the amplitude reaches 35 times the amplitude of the incident Lamb wave. Complex modes have an obvious influence up to three times the thickness of the plate from the end of the plate, which confirms an observation made in another case.[10]

IV. INFLUENCE OF THE α ANGLE ON THE EDGE MODE

The variations of angle result in strong modifications of the Lamb mode reflection. The edge mode also depends on the bevel angle and the resonance is expected to be destroyed at the edge, as the symmetry of the plate is broken when the α angle is different from 90 deg.

A. Shift of the resonance if the α angle varies

A simple way of studying the resonance consists in a computation of the maximum amplitude of the C_{1s} complex modes versus the frequency-thickness product when α varies. If the angle value decreases from 90 to 85 deg (Fig. 7), the C_{1s} amplitude (normalized with respect to S_0) drops very quickly: It is divided by 17. Simultaneously, the resonance frequency decreases slightly ($\Delta FE=-5$ kHz mm). The

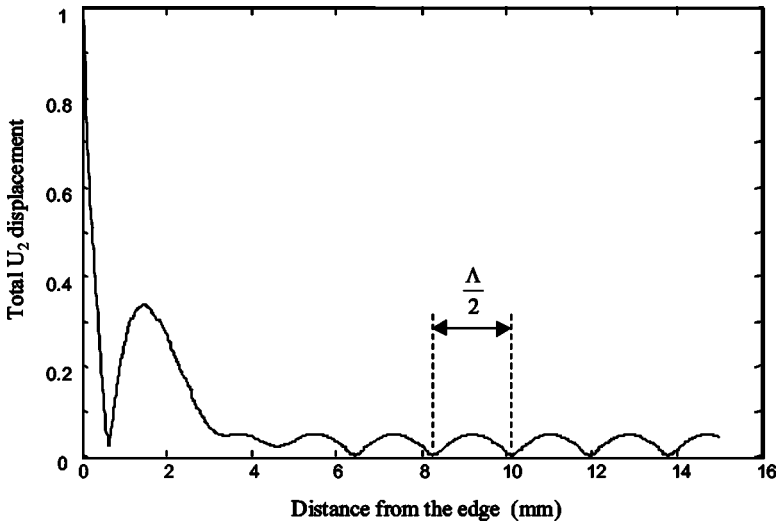


FIG. 6. Evolution of the U_2 total amplitude at the surface versus the distance of propagation (S_0 incident mode, $FE=2.3162$ MHz mm, $\alpha=90°$).

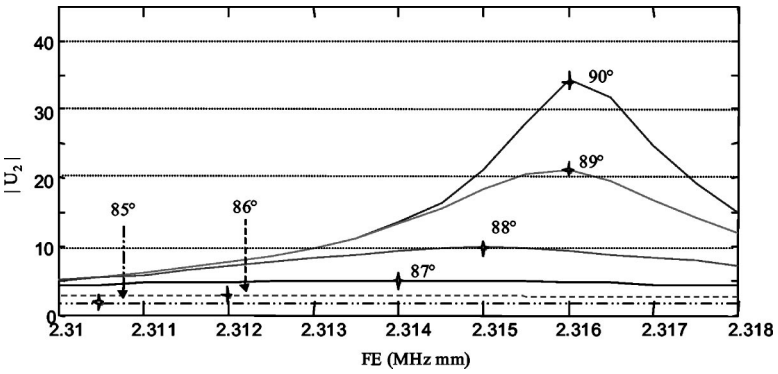
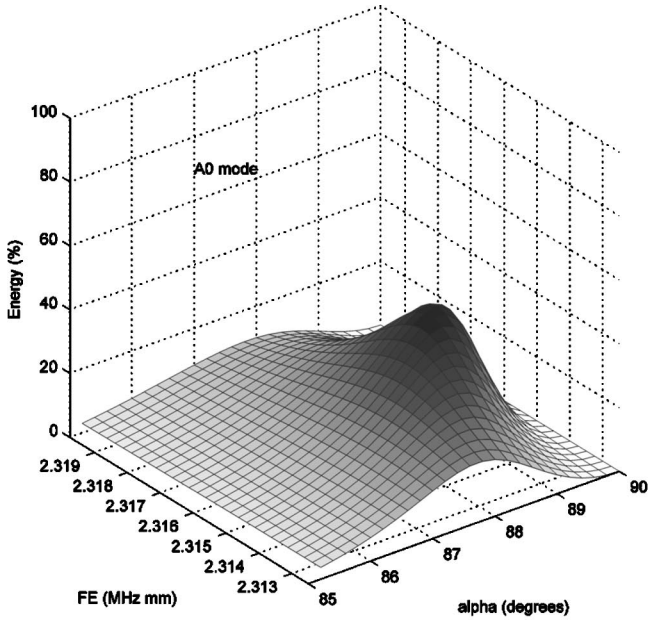
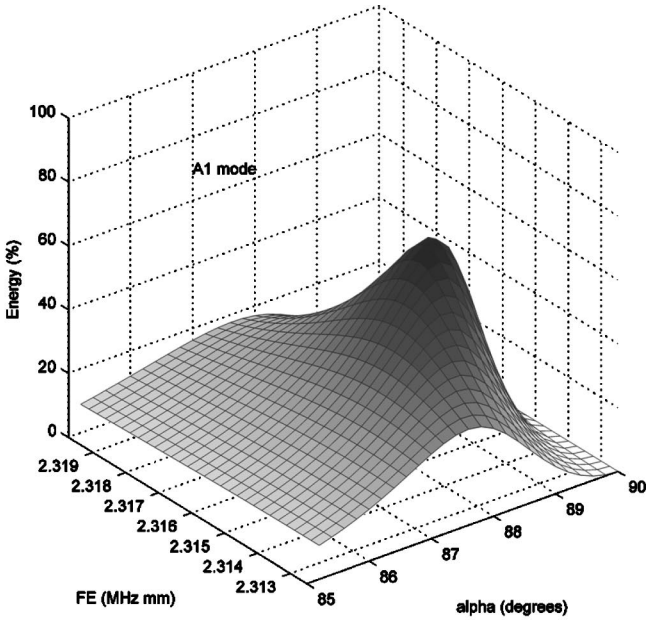


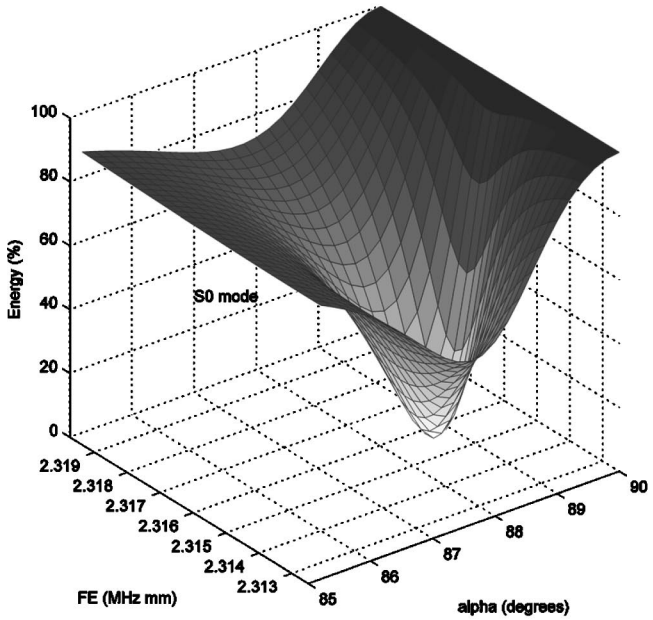
FIG. 7. Maxima of the first complex mode C_{1s} versus FE and α angle (U_2 component).



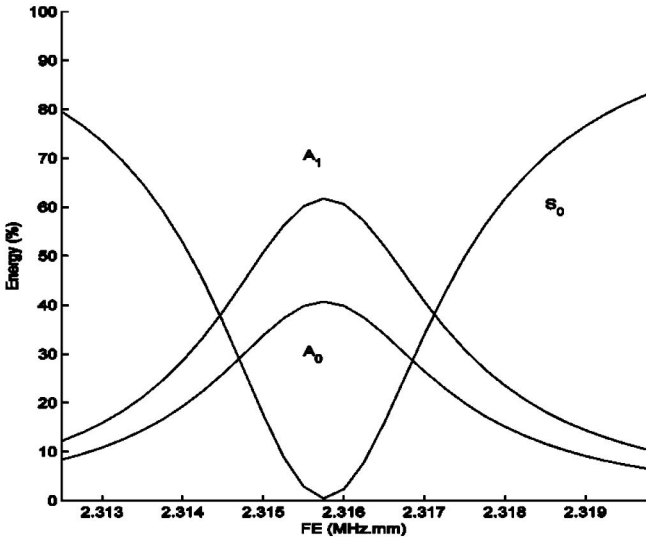
(a)



(c)



(b)



(d)

FIG. 8. (a) Energy rate of the A_0 mode versus FE and α. (b) Energy rate of the S_0 mode versus FE and α. (c) Energy rate of the A_1 mode versus FE and α. (d) Energy balance of the S_0 mode versus FE at α=88.8 deg.

same weakening is observed for the other complex modes. The standing waves, calculated at $\alpha=90°$, cannot settle in the plate thickness.

B. Influence on the energy balances

As in the previous study,[9] we submit the behavior of the reflected wave amplitudes (S_0, A_0, and A_1 modes). In order to make the comparisons easier, the energy rates (obtained when the α angle decreases from 90 deg) are shown in Fig. 8, which represent the reflected energies versus α and FE (respectively, for the A_0, S_0, and A_1 modes in Figs. 8(a)–(c)). When $\alpha=90$ deg, the energy is entirely reflected into a S_0 wave; then a very fast reduction occurs if α decreases. This energy is reduced to zero when $\alpha=88.8$ deg and $FE=2.3158$ MHz mm. At the same time, the reflected energy is transferred into the A_0 and A_1 modes. If the α angle is equal to 88.8 deg and if the FE product varies, the very small value of the S_0 energy and the maximum reached by A_0 and A_1 (respectively, 40% and 60%) clearly appear [Fig. 8(d)]. It is obviously difficult to explain this resonant phenomenon ($\Delta FE=3.3$ kHz mm and $\Delta FE/FE=1.3\times10^{-3}$) and to relate it to the edge mode. When the energy of the reflected S_0 mode is null, the edge mode still exists its amplitude being halved. The extinction of an incident mode is not usual if the α angle is different from 90°. We observe this phenomenon in this case only. For $\alpha=90°$ and $FE=1.85$ MHz mm, we note[9] similar cases for the A_0 and A_1 incident waves: A_0 reflected in A_1 and reciprocally. Nevertheless, the resonance band frequency is quite larger than 1 MHz mm.

V. CONCLUSION

The computation method used in this paper describes the reflection of a harmonic Lamb wave at the beveled end of an elastic plate. This method based on the modal theory is used in order to prove the existence of the edge mode in a stainless steel plate beveled at 90 deg. It is confirmed that this mode is a resonance of complex modes (mainly C_{1s}), when the S_0 Lamb wave is incident at $FE=2.3162$ MHz mm. The computed resonance is sharp and does not modify the energy balances. Even when the edge mode involves an accumulation of energy at the end of the plate, an exact energy balance is computed. The incident energy is entirely converted into the reflected waves (only one wave in this case). Under these conditions, the displacements at the end of the plate are close to 35 times the incident S_0 mode one. This ratio appears to be weaker than the results given by other researchers but depends on the Poisson coefficient of the material and also depends on the definition of the ratio.

The resonant phenomenon is deteriorated by a modification of the bevel. The edge mode disappears quickly when the α angle of the bevel changes from 90 to 85 deg. More precisely, this reduction is accompanied by an unexpected phenomenon. At $\alpha=88.8$ deg and $FE=2.3158$ MHz mm, the incident S_0 mode is completely converted into A_0 and A_1 waves. This kind of phenomenon has not occurred in previous studies. A systematic study can be undertaken in order to correlate the resonance frequency and some properties of the plate material (velocities of sound, Poisson coefficient).

[1] P. J. Torvik, "Reflection of wave trains in semi infinite plates," J. Acoust. Soc. Am. **41**(2), 346–353 (1967).

[2] B. A. Auld and E. M. Tsao, "A variational analysis of edge resonance in a semi-infinite plate," IEEE Trans. Sonics Ultrason. **SU24**(5), 317–326 (1977).

[3] E. A. G. Shaw, "On the resonant vibrations of thick barium titanate disks," J. Acoust. Soc. Am. **28**(1), 38–50 (1956).

[4] D. C. Gazis and R. D. Mindlin, "Extensional vibrations and waves in a plate circular disk and a semi-infinite plate," J. Appl. Mech. **27**, 541–547 (1960).

[5] E. Le Clezio, M. V. Predoï, M. Castaings, B. Hosten, and M. Rousseau, "Numerical predictions and experiments on the free-plate edge mode," Ultrasonics **41**, 25–40 (2003).

[6] J. Z. Shen, S. Y. Zhang, and C. F. Ying, "The reflection of the Lamb wave in a semi-infinite plate," Chin. J. Acoustics **9**(1), 27–35 (1990).

[7] S. Y. Zhang, J. Z. Shen, and C. F. Ying, "The reflection of the Lamb wave by a free plate edge: visualization and theory," Mater. Eval. **46**, 638–641 (1988).

[8] B. Morvan, N. Wilkie-Chancellier, H. Duflo, A. Tinel, and J. Duclos, "Lamb wave reflection at the free edge of a plate," J. Acoust. Soc. Am. **113**(3), 1417–1425 (2003).

[9] N. Wilkie-Chancellier, "Réflexion et conversion d'une onde de Lamb à l'extrémité biseautée d'une plaque," («Reflection and conversion of a Lamb wave at a beveled extremity of a plate») Thesis, University of Le Havre, France, 2003.

[10] M. J. S. Lowe and O. Diligent, "Reflection of the fundamental Lamb modes from the ends of plates," *Review of Progress in Quantitative Non Destructive Evaluation*, Vol. 20, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, New York, 2001), pp. 89–96.

Response of bare 1–3 piezocomposite array to localized electrical excitation

Jérôme Guyonvarch,[a] Dominique Certon, Leong Ratsimandresey,[b] Frédéric Patat, and Marc Lethiecq
GIP ULTRASONS—LUSSI, EIVL rue de la chocolaterie BP 3410, 41034 BLOIS Cédex, France

(Received 28 January 2003; revised 24 September 2003; accepted 4 October 2004)

The theoretical response of a 1–3 piezocomposite plate submitted to localized electrical excitation was studied with the theory of guided waves. The theoretical modeling was based on the global matrix method, and the piezocomposite material was considered as a homogeneous medium. To validate the theoretical results, experimental displacement measurements were performed with an interferometric probe on two piezocomposite plates, one with a single element and one with an array of electrodes. The measured response on the single-element plate was mainly supported by the S_0 and S_3 modes of the plate. Homogenization limits of the composite in terms of frequency and wave number are defined on the basis of data from this sample. Within these limits, the piezocomposite material operates as a homogeneous medium, and comparison between theoretical and experimental results allows the equivalent electroacoustic parameters to be evaluated. A second sample was measured to study the effects of the electrode array on the electroacoustic response of the plate. Two kinds of electrical excitation were studied.
[DOI: 10.1121/1.1829259]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.40.At [AJZ] Pages: 200–209

I. INTRODUCTION

Many piezoelectric transducer arrays for medical imaging or nondestructive evaluation (NDE) currently use 1–3 piezocomposite as an active material.[1] These materials possess high electromechanical coupling factors and are easier to match to water than pure piezoceramics. The simulation process (with analytical or numerical models) is an essential stage in the design of such devices, since it allows prediction of all the characteristics of the array such as sensitivity, bandwidth, directivity pattern of one element, and cross-coupling levels. Many authors have shown the importance of the last two criteria on the final quality of the image. The cross-coupling element increases the aperture of the element, thereby reducing its directivity[2,3] and disturbance of the radiation pattern.[4]

One of the main limitations of the simulation process is the determination of input parameters of models which, for the piezocomposite material, are the equivalent dielectric, elastic, and piezoelectric (i.e., electroacoustic) constants. Generally, equivalent parameters of the thickness and radial modes are easily determined from samples of circular shape. The electrical impedance from an analytical vibration model with adjustable electromechanical material parameters is fit to the measured impedance data. Unknown parameters[5] are thereby derived. However, the number of electroacoustic constants determined this way is not sufficient for the simulation. In the case of homogeneous piezoelectric materials, the missing constants are obtained by manufacturing rods and plates with various poling directions on which electrical impedance can be measured.[5] Here again, using a fitting technique for each sample geometry, all electroacoustic coefficients can be determined. Lamberti[6] has proposed a method in which one rectangular sample with a fixed thickness to width ratio is sufficient to determine all electroacoustic parameters of the ceramic. However, these methods are not suitable for 1–3 piezocomposites.[7] Their mechanical structure does not allow manufacture of samples with rod geometry. Equivalent electromechanical properties of a single piezocomposite within an array must be directly determined from a plate with a grounded plane on the lower surface and an electrode array on the upper surface. We report here an experimental and theoretical study of the electroacoustic response of such a piezocomposite plate to the electrical excitation of one element in the array. The aim was to provide an analytical model to predict the behavior of a bare piezocomposite plate and to determine the equivalent electroacoustic tensor of the material.

The response of a plate to a localized source on one face has been extensively studied. Several methods have been proposed to solve this problem, the first being the normal mode expansion technique,[8,9] in which the fields in the layer are expressed as a sum of Lamb modes. The main difficulty is to determine the complete dispersion curve for each mode, including propagating, evanescent, and nonpropagating modes. The second consists of calculating the elastodynamic Green function of the system using the Cagniard de Hoop method.[10,11] This method expresses the response of the plate in terms of the generalized ray theory. The last method is based on a propagation model in a multilayered structure. Thomson[12] and Haskell[13] have introduced a method to compute a transfer matrix that connects stresses and displacements between two adjacent interfaces. The global matrix method, an alternative method presented by Knopoff,[14] consists of assembling a single matrix which represents the com-

[a]Electronic mail: guyonvarch@univ-tours.fr
[b]VERMON SA, 180 avenue du général Renault, 37038 TOURS Cédex, France

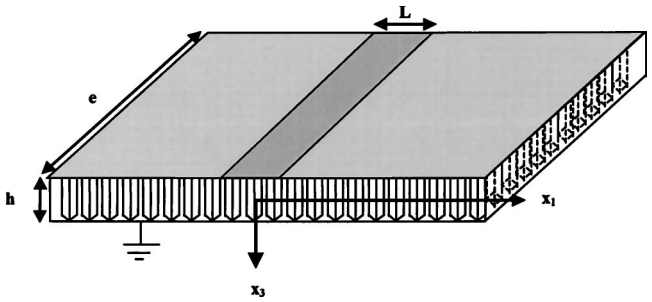


FIG. 1. Geometry of the problem (plate in vacuum).

plete system and contains all the boundary conditions. It has already been applied to piezoelectric media submitted to mechanical or electrical excitation.[15]

The first part of this paper describes an extension of the global matrix method applied to piezocomposite material considered as a homogeneous medium. The excitation source is introduced with electrical boundary conditions. The calculation of electrical impedance of one element in the array is described, as well as displacement fields on the surface of the plate. The second part shows the experimental results in order to compare them with theoretical predictions. Displacement field measurements obtained with laser interferometer and experimental electrical impedance are presented. Two plates were tested with the same piezocomposite structure and the same geometry, but with different electrical boundary conditions. The first possessed a single element, with a single electrode on the excited face. The excited face of the other plate was covered with an array of 32 electrodes, which are the elementary transducers of the device.

II. THEORETICAL ANALYSIS

In this part, a matrix method to simulate the response of a 1–3 piezocomposite plate submitted to a single-element excitation is presented. The geometry of the plate and the definition of axis are given in Fig. 1. The upper face was covered with a single electrode of width L and the lower face was totally metallized. The periodic structure of the material was not considered in this theoretical development. The piezocomposite plate is modeled as a homogeneous medium with a transversely isotropic symmetry class (6-mm symmetry), the same as the piezoelectric phase.

The response of the plate depends on the electrical boundary conditions applied to the excited surface. In our problem, the electrical boundary conditions were mixed, the potential being equal to the excitation voltage on the electrode, and the charge density being zero outside. This problem can be solved with a two-step calculation. First, the electrical boundary condition are supposed to be uniform and expressed in terms of a voltage or a charge excitation. Next, the technique described by Milsom[16] is applied to take into account the mixing of the boundary conditions and to calculate the charge density induced by the voltage applied to the electrode.

A. Basic equations and boundary conditions

We considered a two-dimensional system in which all field components are constant in the x_2 direction, and the electrode is considered as an infinite extent source in the x_2 direction. Waves propagate in the (Ox_1, Ox_3) plane and the direction x_1 corresponds to the propagation direction of Lamb modes in the plate. Two kinds of solution exist for material having 6-mm symmetry. The first is a nonpiezoelectric horizontal shear wave polarized along the Ox_2 axis, and the second corresponds to piezoelectric Lamb modes polarized in the (Ox_1, Ox_3) plane. For problems concerning electrical excitation at the surface, horizontal shear waves are not excited and are not taken into account here.

Assuming that the quasistatic approximation is valid, the displacement and the electric potential in the (Ox_1, Ox_3) plane are governed by[17]

$$c_{ijkl}\frac{\partial^2 u_k}{\partial x_1 \partial x_i} + e_{kij}\frac{\partial^2 \phi}{\partial x_i \partial x_k} = \rho \frac{\partial^2 u_j}{\partial t^2}, \quad i,j,k = 1,3, \tag{1}$$

$$\epsilon_{ij}\frac{\partial^2 \phi}{\partial x_i \partial x_j} - e_{ijk}\frac{\partial^2 u_k}{\partial x_i \partial x_j} = 0, \quad i,j,k = 1,3, \tag{2}$$

where ρ is the density, c_{ijkl}, e_{ijk}, ϵ_{ij} are the equivalent elastic, dielectric, and piezoelectric constants; u_k is the displacement in the kth direction, and Φ is the potential.

Following the same procedure as described in Ref. 15, we assume solutions of the form

$$U_i(k_1, x_3, \omega) = A_i \exp j(\omega t - k_1 x_1 - k_3 x_3), \quad i = 1,3, \tag{3a}$$

$$\Phi(k_1, x_3, \omega) = A_4 \exp j(\omega t - k_1 x_1 - k_3 x_3), \tag{3b}$$

where (k_1, k_3) are the wave vector coordinates.

Equations (1) and (2) become a linear system (Christoffel's equation)

$$[\Gamma_{ij} - \rho\omega^2 \delta_{ij}][A_i] = 0, \tag{4}$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. $[\Gamma]$ is a symmetrical matrix whose elements are given by[17]

$$\Gamma_{ij} = c_{3i3j}k_3^2 + (c_{1i3j} + c_{3i1l})k_3 k_1 + c_{1i1j}k_1^2, \quad i = 1,3, \quad j = 1,3, \tag{5a}$$

$$\Gamma_{i4} = \Gamma_{4i} = e_{3i3}k_3^2 + (e_{1i3} + e_{3i1})k_3 k_1 + e_{1i1}k_1^2, \quad i = 1,3, \tag{5b}$$

$$\Gamma_{44} = -[\epsilon_{33}k_3^2 + (\epsilon_{13} + \epsilon_{31})k_3 k_1 + \epsilon_{11}k_1^2]. \tag{5c}$$

To obtain a nontrivial solution, it is required that the determinant of the matrix $[\Gamma_{ij} - \rho\omega^2\delta_{ij}]$ be equal to zero. For each value of (ω, k_1), there are six solutions for k_3 (k_3^p) and six three-component eigenvectors $A^p = (A_1^p, A_2^p, A_3^p)$, $p = 1,...,6$. The general solution can be written as a linear combination of all the partial waves

$$U_i(k_1, x_3, \omega) = \sum_{p=1}^{6} C_p A_i^p \exp(jk_3^p x_3) \times \exp j(k_1 x_1 - \omega t), \quad i = 1,3, \tag{6a}$$

$$\Phi(k_1, x_3, \omega) = \sum_{p=1}^{6} C_p A_3^p \exp(jk_3^p x_3) \exp j(k_1 x_1 - \omega t), \tag{6b}$$

where Cp are weighting coefficients which are determined from the electrical and mechanical boundary conditions expressed at each interface. Stresses and electrical displacement in the plate are written as

$$T_{3j}=c_{3jkl}\frac{\partial u_k}{\partial x_l}+e_{k3j}\frac{\partial \phi}{\partial x_k}, \tag{7}$$

$$D_3=e_{3kl}\frac{\partial u_k}{\partial x_l}-\epsilon_{3k}\frac{\partial \phi}{\partial x_k}. \tag{8}$$

The waves must satisfy the mechanical boundary conditions on both faces of the plate

$$T_{3i}\left(k_1,\pm\frac{h}{2},\omega\right)=0. \tag{9}$$

The electrical boundary conditions depend on the electrical state of the surface. For a grounded metallized surface, the electrical condition is reduced to

$$\Phi\left(k_1,\pm\frac{h}{2},\omega\right)=0. \tag{10}$$

For a free surface (without metallization), the electrical boundary condition is expressed by the electrical induction continuity at the vacuum/plate interface

$$D_3^{\text{plate}}\left(k_1,\pm\frac{h}{2},\omega\right)-D_3^{\text{vacuum}}\left(k_1,\pm\frac{h}{2},\omega\right)=0, \tag{11}$$

where $D_3^{\text{vacuum}}(k_1,\pm h/2,\omega)=\pm k_1\epsilon_0\Phi^{\text{vacuum}}(k_1,\pm h/2,\omega)$, and the potential continuity

$$\Phi^{\text{plate}}\left(k_1,\pm\frac{h}{2},\omega\right)=\Phi^{\text{vacuum}}\left(k_1,\pm\frac{h}{2},\omega\right). \tag{12}$$

B. Response of the plate submitted to uniform electrical excitation

The geometry of the problem to be analyzed is that shown in Fig. 1. On the lower surface, the grounded surface means that

$$\Phi(x_1,t)=0, \quad x_3=h/2. \tag{13}$$

The single element, located on the surface at $x_3=-h/2$, allows introduction of the excitation source in the electrical boundary conditions. For a uniform voltage excitation, the electrical boundary condition at $x_3=-h/2$ becomes

$$\Phi(x_1,t)=V_0\,\text{rect}(x_1)\,\delta(t), \quad x_3=-h/2, \tag{14}$$

where V_0 is the voltage excitation amplitude and $\text{rect}(x_1)$ a gate function whose width is equal to that of the electrode. The temporal dependence of the potential is equal to a Dirac function to calculate the impulse response of the plate.

For a uniform charge excitation, the boundary conditions can be written as

$$\Delta D_3(x_1,t)=\sigma_0\,\text{rect}(x_1)\,\delta(t), \quad x_3=-h/2, \tag{15}$$

where σ_0 is the excitation amplitude of the surface charge density.

Equations (6a) and (6b), coupled with the mechanical (9) and electrical boundary conditions (13)–(15) for a charge excitation, or (13)–(14) for a voltage excitation, lead to six linear equations with six unknown coefficients.

For a charge excitation, the system can be written in a matrix form

$$\begin{pmatrix} M_{11}\overline{E_1} & M_{12}\overline{E_2} & M_{13}\overline{E_3} & M_{14}\overline{E_4} & M_{15}\overline{E_5} & M_{16}\overline{E_6} \\ M_{21}\overline{E_1} & M_{22}\overline{E_2} & M_{23}\overline{E_3} & M_{24}\overline{E_4} & M_{25}\overline{E_5} & M_{26}\overline{E_6} \\ N_{31}\overline{E_1} & N_{32}\overline{E_2} & N_{33}\overline{E_3} & N_{34}\overline{E_4} & N_{35}\overline{E_5} & N_{36}\overline{E_6} \\ M_{11}E_1 & M_{12}E_1 & M_{13}E_1 & M_{14}E_1 & M_{15}E_1 & M_{16}E_1 \\ M_{21}E_1 & M_{22}E_2 & M_{23}E_3 & M_{24}E_4 & M_{25}E_5 & M_{26}E_6 \\ E_1 & E_2 & E_3 & E_4 & E_5 & E_6 \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ F_{\text{exc}}^{\sigma}(\omega,k_1) \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{16}$$

where

$$M_{1p}=-i(e_{133}k_1+c_{2323}k_3^pV_p+c_{2323}k_1W_p), \quad p=1,...,6,$$

$$M_{2p}=-i(e_{333}k_3^p+c_{1133}k_1V_p+c_{1133}k_3^pW_p), \quad p=1,...,6,$$

$$M_{3p}=i(\epsilon_{33}k_3^p-e_{311}k_1V_p-e_{333}k_3^pW_p), \quad p=1,...,6,$$

$$N_{3p}=(M_{3p}+k_1\epsilon_0), \quad p=1,...,6,$$

$$E_p=e^{-ik_3^ph/2}, \quad \overline{E_p}=e^{ik_3^p(h/2)}, \quad p=1,...,6.$$

$F_{\text{exc}}^{\sigma}(\omega,k_1)$ is the excitation force vector obtained by spatial Fourier transform of Eq. (15)

$$F_{\text{exc}}^{\sigma}(\omega,k_1)=\sigma_0 L\,\text{sinc}(\pi k_1 L). \tag{17}$$

For a voltage excitation, the system becomes

$$\begin{pmatrix} M_{11}\overline{E_1} & M_{12}\overline{E_2} & M_{13}\overline{E_3} & M_{14}\overline{E_4} & M_{15}\overline{E_5} & M_{16}\overline{E_6} \\ M_{21}\overline{E_1} & M_{22}\overline{E_2} & M_{23}\overline{E_3} & M_{24}\overline{E_4} & M_{25}\overline{E_5} & M_{26}\overline{E_6} \\ \overline{E_1} & \overline{E_2} & \overline{E_3} & \overline{E_4} & \overline{E_5} & \overline{E_6} \\ M_{11}E_1 & M_{12}E_1 & M_{13}E_1 & M_{14}E_1 & M_{15}E_1 & M_{16}E_1 \\ M_{21}E_1 & M_{22}E_2 & M_{23}E_3 & M_{24}E_4 & M_{25}E_5 & M_{26}E_6 \\ E_1 & E_2 & E_3 & E_4 & E_5 & E_6 \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ F_{\text{exc}}^V(k_1,\omega) \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{18}$$

where $F_{\text{exc}}^V(\omega,k_1)$ is the excitation force vector obtained by spatial Fourier transform of Eq. (14)

$$F_{\text{exc}}^V(\omega,k_1) = V_0 L \,\text{sinc}(\pi k_1 L) \tag{19}$$

From the C_p coefficient determination, Eqs. (6a) and (6b) are solved to calculate the displacement U_i and potential Φ as a function of the wave number k_1 and the angular frequency ω. The response of the plate as a function of the coordinates x_1 and time t are obtained by a double-inverse Fourier transform

$$u_i(x_1,x_3,t) = \frac{1}{4\pi^2}\int\int U_i(k_1,x_3,\omega)e^{j(k_1x_1-\omega t)}dk_1\,d\omega, \quad i=1,3, \tag{20a}$$

$$\phi(x_1,x_3,t) = \frac{1}{4\pi^2}\int\int \Phi(k_1,x_3,\omega)e^{j(k_1x_1-\omega t)}dk_1\,d\omega. \tag{20b}$$

Functions $U_i(k_1,x_3,\omega)$ and $\Phi(k_1,x_3,\omega)$ have real singularities which correspond to free modes of the plate. These real singularities $k_1(\omega)$ cancel the determinant of the global matrix and correspond to the dispersion curves of the Lamb waves. For piezocomposite material, mechanical and dielectric losses must be taken into account. They are introduced using complex stiffness and dielectric constants. Thus, poles of the integrand move away from the real axis, and the double integration of Eqs. (20a) and (20b) may be easily achieved using a double-fast Fourier transform algorithm.

Simulation of the harmonic response of the plate submitted to a voltage and a charge excitation was performed using Eq. (6a). The two theoretical (ω,k_1) diagrams of the displacement are shown in Figs. 2 and 3. The amplitudes are normalized by the maximum and shown in a logarithmic scale. Contour levels between 0 and -30 dB are presented by steps of -3 dB. Theoretical dispersion curves of S_0, S_1, and S_3 modes are superimposed in the two diagrams using dotted lines to identify Lamb waves. These curves were calculated for a plate with a shorted (Fig. 2) and free (Fig. 3) excited surface, using the method described by Ref. 18. Looking at Fig. 2, most of the displacement occurred for (ω,k_1) coordinates corresponding to the S_0, S_1, and S_3 Lamb mode dispersion curves calculated when the two surfaces of the plate were metallized. Maximum amplitude displacement was reached for $k_1=0$ rad/m and $f=3.3$ MHz. This was the cutoff frequency of the S_3 mode. The S_1 mode seemed more weakly coupled than the other two. The maximum amplitude for charge excitation (Fig. 3) was then located at 3.85 MHz for $k_1=0$ rad/m. This was the new cutoff frequency of the S_3 mode when one surface was metallized and the other open. The position of the S_0 dispersion curve was also modified. This mode seemed weakly coupled since its amplitude had diminished. Finally, the S_1 mode disappeared. These two theoretical results show that, as already

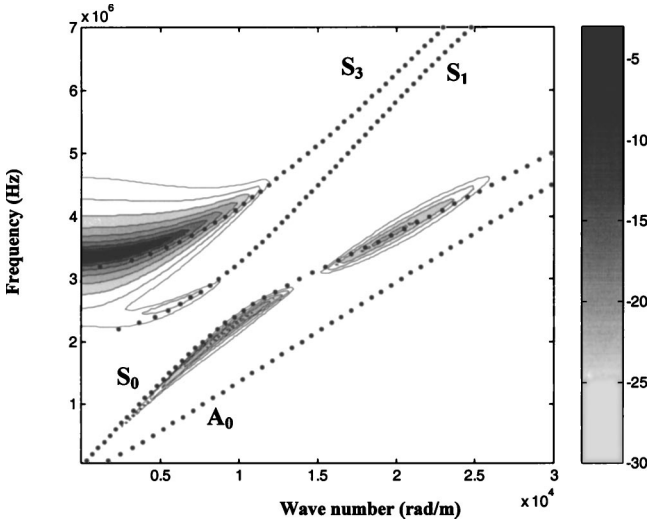


FIG. 2. Level contour representation of theoretical displacement diagram $U(\omega,k_1)$ for uniform voltage excitation (in dB). Theoretical dispersion curves are superimposed using dotted lines to identify Lamb modes.

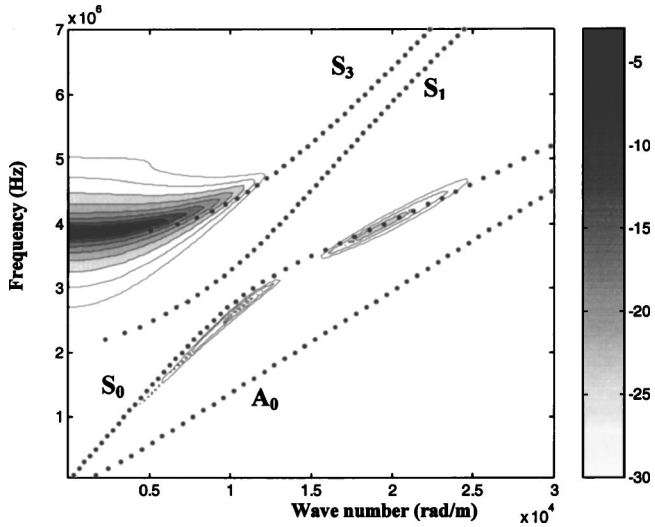


FIG. 3. Level contour representation of theoretical displacement diagram $U(\omega,k_1)$ for uniform charge excitation (in dB). Theoretical dispersion curves are superimposed using dotted lines to identify Lamb modes.

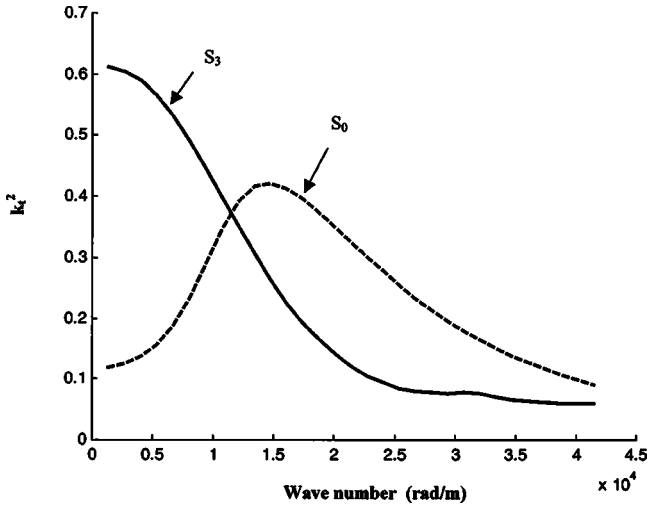


FIG. 4. Theoretical variations of electromechanical coupling coefficient of S_0 (dashed) and S_3 (continuous) modes.

observed,[19] the response of the plate was mainly governed by the S_0 and S_3 modes because they had the highest electromechanical coupling coefficient. Figure 4 presents the theoretical electromechanical coupling coefficient versus wave number for S_0 and S_3 modes. The behavior of these two modes may be analyzed from the shapes of these curves. It can be seen that their displacement amplitudes follow the same variations as the associated electromechanical coupling coefficients for S_0 and S_3 modes. However, we note that when the electromechanical coupling of the S_0 mode reaches maximum ($k_1 = 1500$ rad/m), the displacement amplitude reaches a local minimum. This minimum corresponds to the first minimum of the excitation function [Eqs. (17) and (19)], which is a sinus cardinal function.

C. Response of the plate submitted to nonuniform electrical excitation

In this part, the theoretical approach proposed by Milsom[16] was developed to calculate the electro-acoustic Green function of a piezoelectric half-space with mixed electrical boundary conditions.

The potential and the charge density at the surface of piezoelectric material are related by

$$\Phi(x_1,\omega) = G(x_1,\omega) * \sigma(x_1,\omega), \tag{21}$$

where $G(x_1,\omega)$ is the Green function. In our problem, $\Phi(x_1,\omega)$ was equal to the excitation voltage whenever x_1 was on the electrode, and $\sigma(x_1,\omega)$ was zero over all the unmetallized zone. The aim of this calculation was to determine the charge density on the electrode and the potential outside.

We first calculated $G(x_1,\omega)$ considering that the charge excitation is a Dirac function $\delta(x_1)$ [Eq. (15)] and by solving the equation set (16)–(17) and (20). Note that the force excitation $F^{\sigma}_{\text{exc}}(\omega,k_1)$ [Eq. (17)] is equal to 1. We defined a set of M points x_i (Fig. 5) existing only on the electrode, where they have a small spacing Δx. Equation (21) is written as

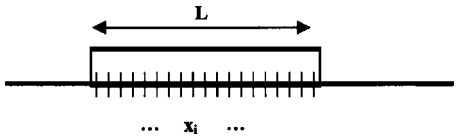


FIG. 5. Subdivision of the electrode for computation of the charge density profile.

$$\Phi_i(\omega) = \sum_{j=1}^{M} A_{ij}\sigma_j(\omega), \quad i = 1,...,M, \tag{22}$$

where the A_{ij} coefficients are equal to $\Delta x G(x_i - x_j,\omega)$.

$\sigma_i(\omega) = \sigma(x_i,\omega)$ may now be expressed in terms of $\Phi_i(\omega) = \Phi(x_i,\omega)$

$$\sigma_i(\omega) = \sum_{j=1}^{M} B_{ij}\Phi_j(\omega), \quad i = 1,...,M, \tag{23}$$

where B_{ij} is the reciprocal of the matrix A_{ij}.

A constant voltage excitation $\Phi_j(\omega) = V_0$ was introduced in Eq. (23) in order to calculate the charge density produced on the electrode. This new charge density was then used as an excitation source in system (16) to obtain the electromechanical response of the plate. The current $I(\omega)$ entering the electrode was provided by numerical integration of the charge density along the electrode. The ratio between the excitation voltage V_0 and $I(\omega)$ supplies the electrical impedance $Z_{\text{el}}(\omega)$ versus frequency.

III. EXPERIMENTAL RESULTS

Two arrays were manufactured of the same type of piezocomposite material, with 50% ceramic volume fraction, plate thickness of 470 μm, array pitch of 500 μm, and thickness mode frequency of 3.5 MHz. Both faces were first totally electroplated for each sample. Then, a single electrode and a 32-electrode array were patterned on the reverse of the samples (Fig. 6).

The surface displacements over piezocomposite array were measured using a commercially available interferom-

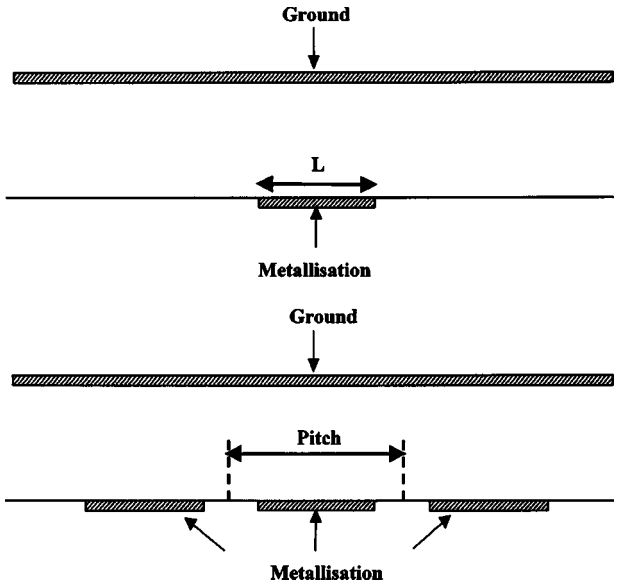


FIG. 6. Structure of the two samples.

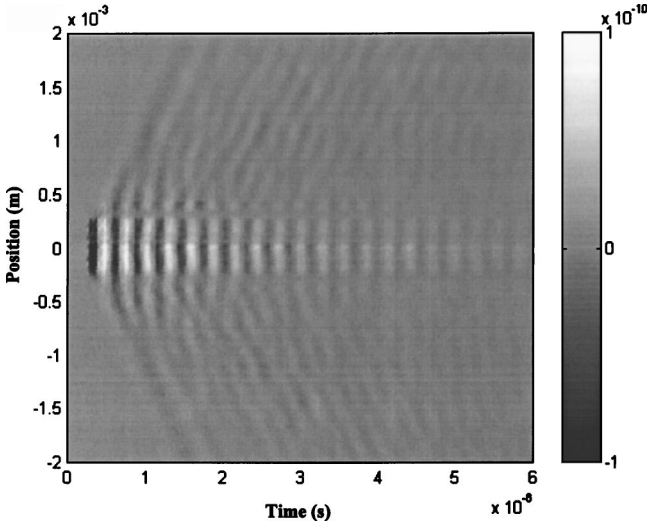


FIG. 7. Experimental displacement diagram $u(x,t)$ (in meters).

eter probe (SH 140; Thales Lasers, France). According to the experimental procedure defined in a previous article,[20] the central element of the probe was excited with a pulse generator. The amplitude delivered by the generator was 50 V; the transient response was recorded by scanning the surface and electrical impedance was measured using a network analyzer (HP4195A).

A. Homogenization limits

The first experimental results are for the single element. The experimental response $u(x_1,t)$ of the displacement field as a function of position x_1 and time t is shown in Fig. 7, where the horizontal axis corresponding to the time axis is graduated in seconds, and the vertical axis shows the position in meters. Near $t=0$, it can be seen that the excited element vibrates in its thickness mode immediately after the excitation and that, as expected, the maximum displacement was located on it. Propagation of Lamb waves inside the plate clearly appeared afterwards, outside the excited element, through the two oblique traces on the diagram.

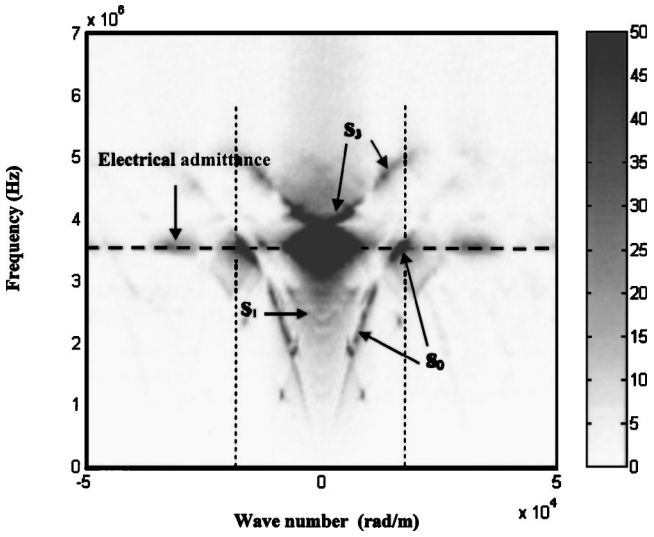


FIG. 8. Decomposition of plane waves $U(\omega,k_1)$ of the experimental displacement (arbitrary units).

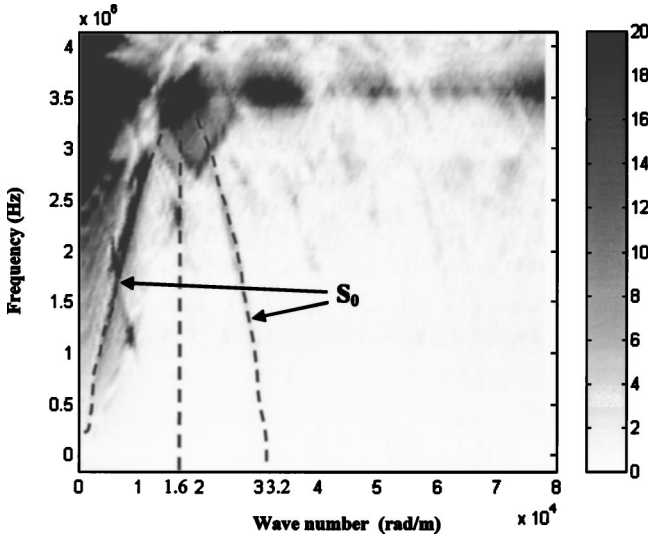


FIG. 9. Decomposition of plane waves $U(\omega,k_1)$ of the experimental displacement (arbitrary units).

The plane wave decomposition of the field $u(x_1,t)$ was computed using a 2D Fourier transform[21] in order to identify these Lamb waves. The units used were wave number (k_1) and frequency (ω). The experimental $U(\omega,k_1)$ displacement field is given in Fig. 8. A horizontal line around 3.5 MHz can be observed on the diagram, which shows that most of the energy was localized around this frequency. Looking at the electrical admittance curves in Fig. 10, it can be seen that they correspond to the maximum of the real part of the admittance, i.e., for a maximum of the current entering the electrode. This result was expected. Since the output impedance of the generator was much smaller than that of the element, the generator acts here as a current generator. Just above this line at 3.5 MHz, the first mechanical resonance of the thickness mode appears at around 3.9 MHz at $k_1=0$ rad/m and follows the dispersion curve of S_3 mode until $k_1=6000$ rad/m. Two other dispersion curves can be observed at lower frequencies: S_0 and S_1. Part of the previous

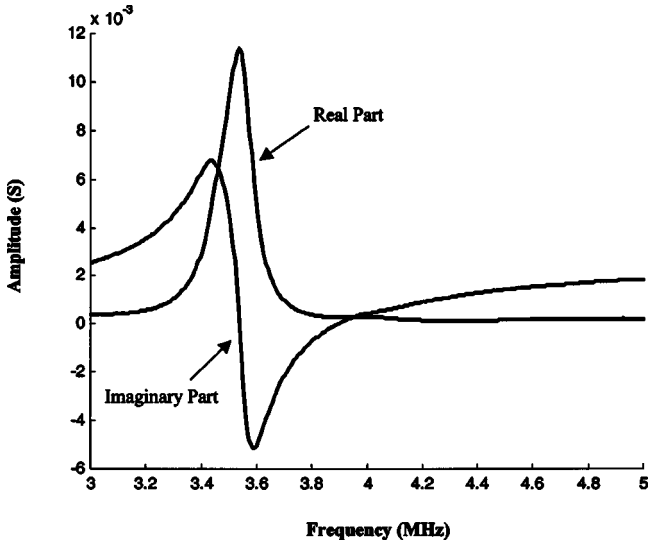


FIG. 10. Real and imaginary parts of experimental electrical admittance of an array element.

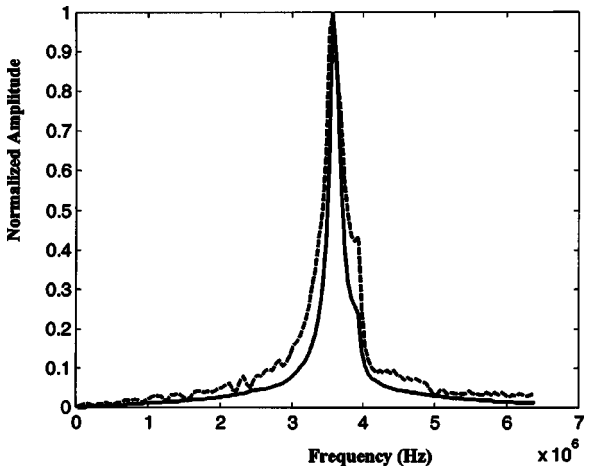


FIG. 11. Curve versus frequency extract from $U(\omega,k_1)$ diagram for $k_1 = 0$ rad/m.

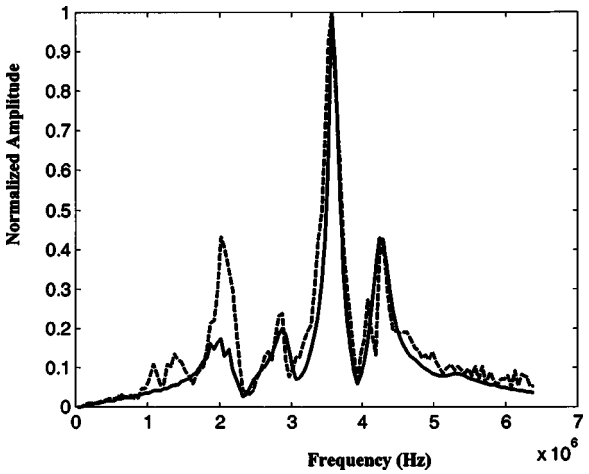


FIG. 13. Curve versus frequency extract from $U(\omega,k_1)$ diagram for $k_1 = 7363$ rad/m.

diagram is presented in Fig. 9 with a greater amplitude scale, i.e., the zone located between wave numbers $k_1 = 0$ rad/m and $k_1 = 80\,000$ rad/m and between frequencies $f = 0$ MHz and $f = 4$ MHz. It can be seen that the experimental dispersion curve of the S_0 mode reached a maximum for $k_1 = \pm 16\,000$ rad/m and decreased symmetrically, with lower amplitude around the lines $k_1 = \pm 16\,000$ rad/m. The dispersion curve of S_0 mode seemed periodic, with a period equal to 2*16 000 rad/m. This was due to the periodicity of the composite structure in which the S_0 Lamb wave underwent Bragg diffraction: The value of $k_1 = 16\,000$ rad/m defines the first Brillouin zone limit:[22] when the wave number approaches the first Brillouin zone limit, the position of the S_0 Lamb wave dispersion curve is modified due to reflections inside the microstructure. This limit defines the wave number range in which the piezocomposite material can be considered as homogeneous equivalent material. Note that the homogenization limits were expressed in terms of useful frequency range and defined by the frequency of lateral modes.[23]

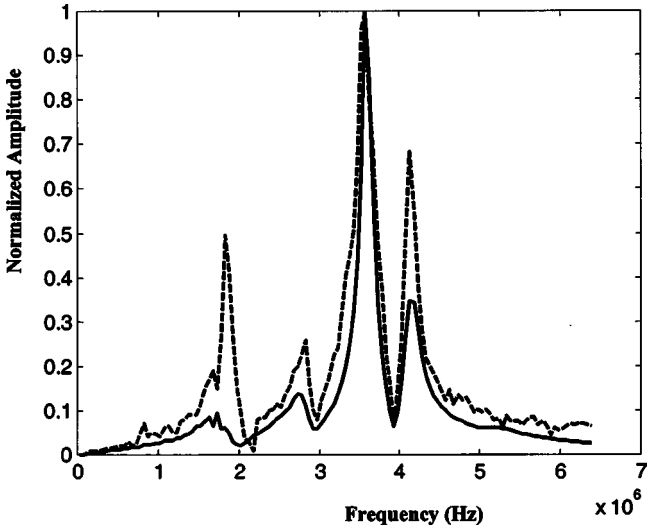


FIG. 12. Curve versus frequency extract from $U(\omega,k_1)$ diagram for $k_1 = 6136$ rad/m.

B. Comparison between theory and experiment

Equivalent electroacoustic characteristics, dielectric and mechanical losses are required to simulate the behavior of the piezocomposite plate. The equivalent thickness mode characteristics were therefore first measured from a piezocomposite sample of circular shape (i.e., e_{33}, complex values of c_{33}^E and ϵ_{33}^S). Other parameters were evaluated theoretically using the homogenization model described by Levassort.[24] This first set of equivalent data was used to initialize the input parameters of our model. The theoretical and experimental (ω,k_1) diagrams were then adjusted to extract the equivalent material parameters. Several curve versus frequency comparisons on the experimental (ω,k_1) diagrams were chosen (for $k_1 = 0$, 6136, and 7363 rad/m), and then adjusted until the experimental and theoretical curves became similar. For losses, default values of low percentages were introduced and not fitted here. The three final curve versus frequency comparisons are presented in Figs. 11–13, and the equivalent electroacoustic constants deduced from experimental diagrams on the composite array are reported in Table I and compared to the initial values.

Experimental and theoretical $U(\omega,k_1)$ diagrams are shown in Figs. 14 and 15 with a level contour representation. The amplitude of displacement was normalized by the maxi-

TABLE I. Comparison between exprimental and theoretical (i.e. homogenization) material constants of the piezocomposite material (50% ceramic volume fraction).

Constant	Theoretical parameters	Experimental parameters
C_{11}^E (10^{10} Pa)	1.473	1.25
C_{12}^E (10^{10} Pa)	0.897	0.748
C_{13}^E (10^{10} Pa)	0.932	0.927
C_{33}^E (10^{10} Pa)	3.33	3.7
C_{44}^E (10^{10} Pa)	0.296	0.5
e_{13} (C/m)	−0.37	−1
e_{33} (C/m)	13.87	9.5
e_{15} (C/m)	0.003	0.003
$\epsilon_{11}^S/\epsilon_0$	7.91	7.81
$\epsilon_{33}^S/\epsilon_0$	751	683
ρ (kg/m^3)	4303	4020
k_t	0.68	0.53

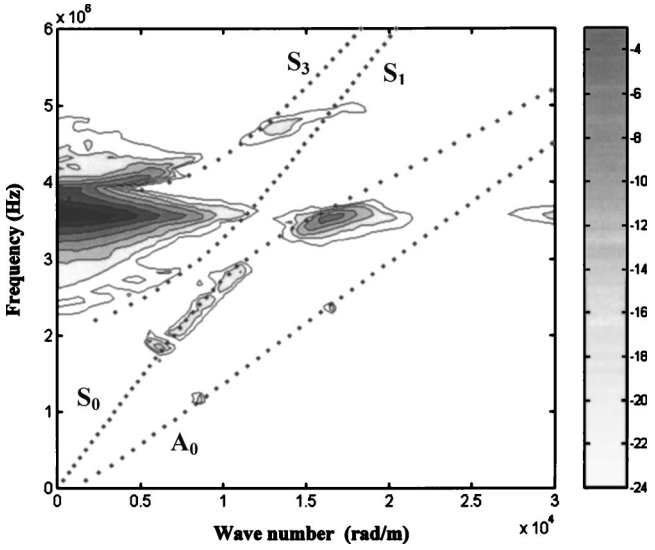


FIG. 14. Level contour representation of the experimental displacement diagram $U(\omega,k_1)$ (in dB). Theoretical dispersion curves are superimposed using dotted lines to identify Lamb modes.

mum and is shown in a logarithmic scale. Contour levels between 0 and −24 dB for experimental diagram, between 0 and −30 dB for theoretical diagram, are presented, by steps of −3 dB. As already observed, the maximum amplitude was located at 3.5 MHz for $k_1=0$ rad/m. The simulation showed the same behavior. It can be seen that the positions of S_3 and S_0 Lamb mode dispersion curves were well predicted, while the wave vector was smaller than the first Brillouin zone limit. The theoretical amplitude of S_0 mode was smaller than the experimental value. The transient displacement field calculated from the theoretical $U(\omega,k_1)$ diagram is shown in Fig. 16, with an absolute value of the displacement, where the amplitude of the voltage excitation has been taken into account. The theory is in accordance with the findings. The impulse displacement response over the excited element lasted 4 to 5 microseconds. The predicted maximum amplitude was equal to 0.1 nm and the experimental value was 0.13 nm. The frequency response of the displacement at the center of the excited element is shown in Fig. 17, while Fig. 18 gives the displacement profile obtained just after the excitation time. The theoretical and experimental widths of the displacement profiles were equal up to −20 dB. Finally, the theoretical and experimental electrical impedance of the element are presented in Fig. 19. The good agreement between theory and experimental findings indicates the relevance of this approach to describing the response of the plate and to determining the equivalent electromechanical tensor. Moreover, these findings validate assumptions made regarding the homogeneity of piezocomposite behavior for wave numbers smaller than the first Brillouin zone limit.

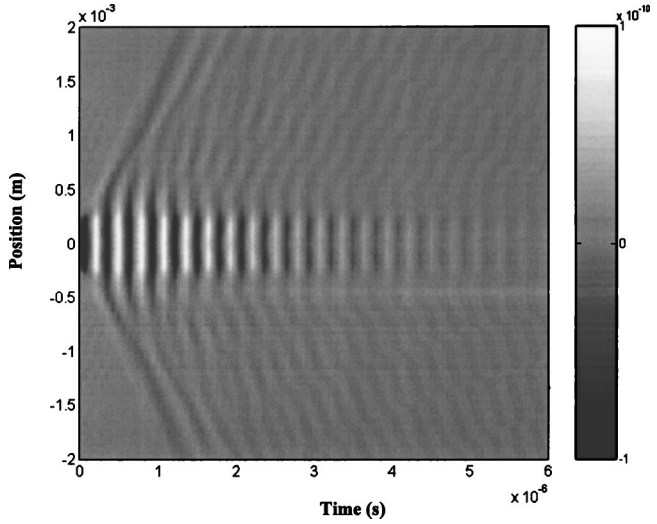


FIG. 16. Theoretical displacement diagram (in meters).

C. Response of the plate covered with an electrode array

In this part, the second test sample with 32 electrodes was investigated in two operating conditions, i.e., with all

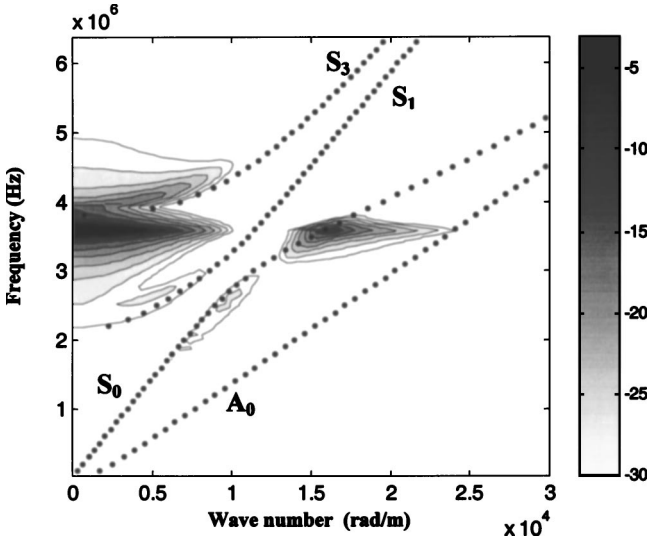


FIG. 15. Level contour representation of the theoretical displacement diagram $U(\omega,k_1)$ (in dB). Theoretical dispersion curves are superimposed using dotted lines to identify Lamb modes.

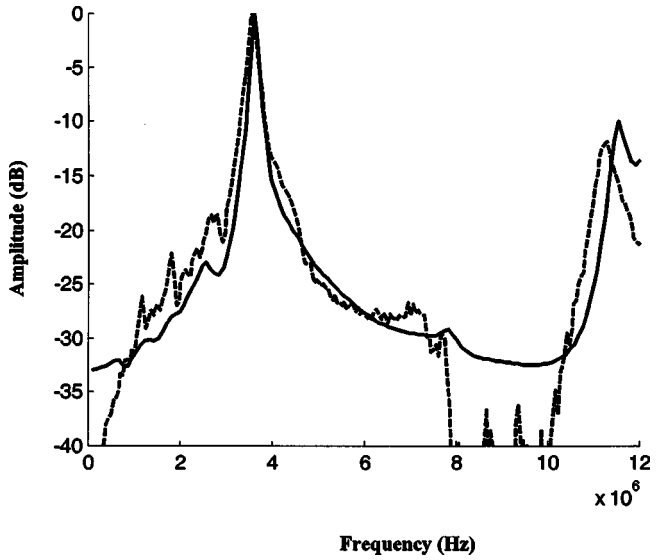


FIG. 17. Experimental (dashed) and theoretical (continuous) displacement frequency response obtained at the center of the element.

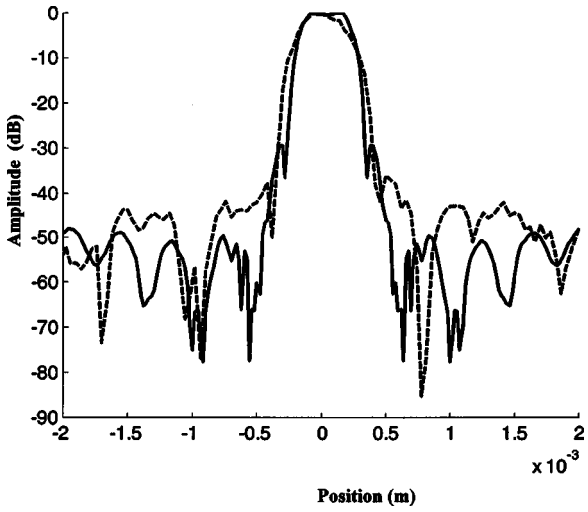


FIG. 18. Experimental (dashed) and theoretical (continuous) displacement profiles just after excitation time.

neighboring elements of the excited element in free voltage and with all neighboring elements grounded. The two experimental (ω,k_1) diagrams are shown in Figs. 20 and 21. Theoretical dispersion curves of A_0, S_0, S_1, and S_3 modes calculated for a plate whose excited surfaces were free and grounded, respectively, have also been plotted. For the first one (Fig. 20), it can be seen that compared with the response of the single-element device (Fig. 14), the maximum displacement amplitude always occurred at $k_1=0$, around 3.5 MHz, while the amplitudes of S_0 and S_1 modes seemed smaller. The beginning of the S_3 mode dispersion curve has also been observed with similar amplitude. Figure 21 shows that when the neighboring elements were grounded, S_0 and S_1 modes practically disappeared and that only the S_3 Lamb mode existed. The maximum amplitude was then located at 3.3 MHz, i.e., the cutoff frequency of S_3 mode when the two surfaces were electrically shorted. This explains modifications observed in the shape of the (ω,k_1) diagram. It is interesting to note that the theoretical diagram presented in Fig. 2 shows the same behavior but with several differences

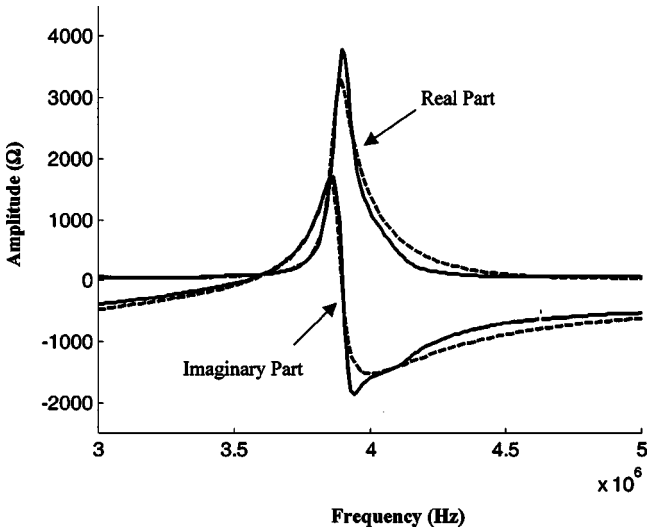


FIG. 19. Experimental (dashed) and theoretical (continuous) real and imaginary parts of electrical impedance.

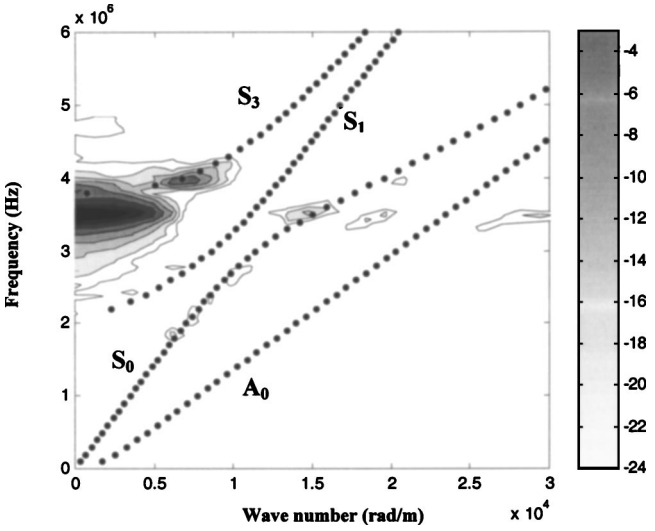


FIG. 20. Level contour (in dB) representation of experimental displacement diagram $U(\omega,k_1)$ where the neighboring elements of the excited are free voltage.

in the amplitude levels of the S_0 and S_1 modes. This suggests that the response of the plate with grounded neighboring elements can be reasonably modeled using uniform electrical boundary conditions [Eq. (19)]. Nevertheless, the free-voltage condition is more difficult to simulate and requires taking into account the neighboring electrodes in the formulation of the electrical boundary conditions.

IV. CONCLUSION

The purpose of this study was to provide an analytical model to predict the electromechanical response of a bare piezocomposite array submitted to an electrical source. The main hypothesis was to model the composite structure as a homogeneous medium. The approach developed for this purpose was based on the propagation theory of waves in a multilayered structure. Modeling of the electrical source was

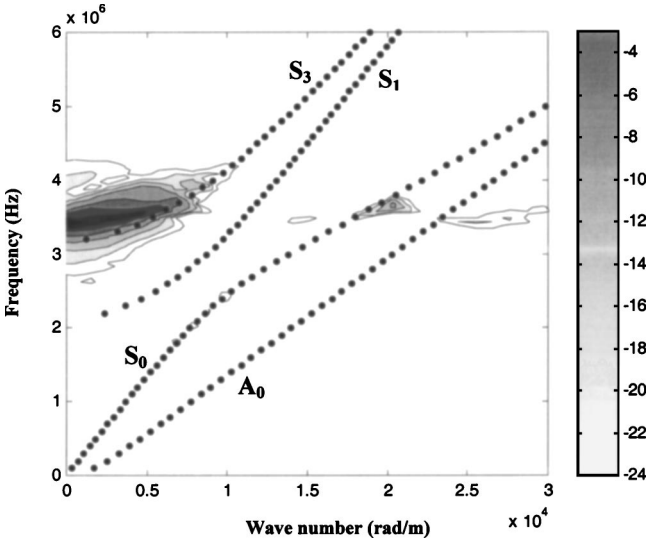


FIG. 21. Level contour (in dB) representation of experimental displacement diagram $U(\omega,k_1)$ where the neighboring elements of the excited are grounded.

performed using Milsom's method in order to take into account the nonuniformity of the electrical boundary conditions.

Limits in terms of wave number and frequency were defined from experimental measurements. Within these limits, the piezocomposite material can be considered as a homogeneous medium, and the model correctly predicts the transient and harmonic displacements on the surface of the array. The presence of several Lamb waves of different amplitudes was observed, the two main ones being S_0 and S_3 Lamb modes. This model can be used to evaluate equivalent piezocomposite parameters, adjusting theoretical and experimental (ω,k_1) diagrams of the displacement. Fitting techniques using both electrical cross-coupling measurements and mechanical displacements will be developed in future studies.

Measurements were performed on a plate covered with a 32-element array. The first case, when all neighboring elements of the excited element were grounded, was well predicted assuming uniform voltage excitation conditions. The electrical boundary conditions must be included for free-voltage elements. The same methods developed to take into account the interdigital electrodes in SAW devices must be used. Work is in progress and will be the subject of a future paper.

Finally, the model will be extended to predict the electroacoustic response of a complete array structure vibrating in water, the aim being to obtain this response at all angles and thus the elementary radiation pattern.

[1] W. A. Smith and B. A. Auld, "Modeling 1–3 composite piezoelectrics: thickness-mode oscillations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**(1), 40–47 (1991).

[2] O. T. Von Ramm and S. W. Smith, "Beam steering with linear array," IEEE Trans. Biomed. Eng. **30**, 438–452 (1983).

[3] D. H. Turnbull and S. Foster, "Beam steering with pulsed two-dimensional transducer arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**(4), 320–333 (1991).

[4] B. Delannoy, C. Bruneel, F. Haine, and R. Torguet, "Anomalous behavior in the radiation pattern of piezoelectric transducers induced by parasitic Lamb wave generation," J. Appl. Phys. **51**(7), 3942–3948 (1980).

[5] ANSI/IEEE std 176-1987, "IEEE Standard on Piezoelectricity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**(5), 717–773 (1996).

[6] N. Lamberti and M. Pappalardo, "A general approximated two-dimensional model for piezoelectric array elements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(2), 243–252 (1995).

[7] S. Sherrit, H. D. Wiederick, and B. K. Mukherjee, "The complete matrix of the piezoelectric, dielectric and elastic material constants of 1–3 piezoelectric ceramic/polymer composites," Proc. SPIE **3241**, 327–337 (1997).

[8] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990).

[9] J. J. Ditri and J. L. Rose, "Excitation of guided waves in generally anisotropic layers using finite sources," J. Appl. Mech. **61**, 330–338 (1994).

[10] A. T. De Hoop and J. H. M. T. Van der Hijden, "Seismic waves generated by an impulsive point source in a solid/fluid configuration with a plane boundary," Geophysics **50**(7), 1083–1090 (1985).

[11] Y. Liu, C. Wang, and C. Ying, "Transient waves in a piezoelectric half-space generated by line-loaded surface sources," IEEE International Ultrasonics Symposium, 377–381 (1991).

[12] W. T. Thomson, "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89–93 (1950).

[13] N. A. Haskell, "Dispersion of surface waves on multilayred media," Bull. Seismol. Soc. Am. **43**, 17–34 (1953).

[14] L. Knopoff, "A matrix method for elastic wave problems," Bull. Seismol. Soc. Am. **54**, 431–438 (1964).

[15] S. G. Joshi and Y. Jin, "Excitation of ultrasonic Lamb waves in piezoelectric plates," J. Appl. Phys. **69**(12), 8018–8024 (1991).

[16] R. F. Milsom, N. H. C. Reilly, and M. Redwood, "Analysis of generation and detection of surface and bulk acoustic waves by interdigital transducers," IEEE Trans. Sonics Ultrason. **24**(3), 147–166 (1977).

[17] D. Royer and E. Dieulesaint, *Ondes Élastiques dand les Solides* (Masson, Paris, 1996).

[18] F. Teston, G. Feuillard, L. Tessier, and M. Lethiecq, "Mass sensitivity of acoustic plate mode in liquids," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**(5), 1266–1272 (1998).

[19] D. Certon, N. Felix, E. Lacaze, F. Teston, and F. Patat, "Investigation of cross-coupling in 1–3 piezocomposite arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(1), 85–92 (2001).

[20] N. Felix, D. Certon, L. Ratsimandresi, M. Lethiecq, and F. Patat, "1D ultrasound array: Performances evaluation and characterization by laser interferometry," IEEE International Ultrasonics Symposium (2000).

[21] D. Alleyne and P. Cawley, "A two-dimensional Fourier transform method for the measurement of propagating multimode signals," J. Acoust. Soc. Am. **89**(3), 1159–1168 (1991).

[22] Y. J. Wang, "Waves and vibrations in elastic superlattice composites," Stanford University, 1986.

[23] T. R. Gururaja, W. A. Schulze, L. E. Cross, R. E. Newham, B. A. Auld, and Y. J. Wang, "Piezoelectric composite material for ultrasonic transducer applications. I. Resonant modes of vibration of PZT rod–polymer composites," IEEE Trans. Sonics Ultrason. **32**, 481–498 (1985).

[24] F. Levassort, M. Lethiecq, D. Certon, and F. Patat, "A matrix method for modeling electroelastic moduli of 0–3 piezocomposites," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**(2), 445–452 (1997).

The energy method for analyzing the piezoelectric electroacoustic transducers

Boris Aronov
Acoustic Research Laboratory, Advanced Technology and Manufacturing Center and Department of Electrical and Computer Engineering, the University of Massachusetts, Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300 and BTECH Acoustics, 1445 Wampanoag Trail, Suite 115, East Providence, Rhode Island 02915

(Received 21 February 2004; revised 1 August 2004; accepted 13 August 2004)

The energy method of calculating the properties of piezoelectric electroacoustic transducers is considered. The Lagrangian of an electroacoustic transducer as a system performing multiple energy conversions is introduced. The Euler equations describing operation of a transducer with many mechanical degrees of freedom are derived from the least action principle. The corresponding multicontour equivalent circuit of the transducer is introduced. For the particular case of a transducer having one mechanical degree of freedom the governing equations are obtained by applying the Energy Conservation Law, and equivalent circuit with one mechanical branch also introduced. Application of the energy method is illustrated with examples of the pulsating spherical transducer as the one degree of freedom system and the multimode cylindrical transducer comprised of circular rings as the system with multiple degrees of freedom. Advantages of the method for application with electroacoustic transducers are summarized. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1802536]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.38.Pf [AJZ] Pages: 210–220

I. INTRODUCTION

Calculation of electroacoustic transducers constitutes a complex coupled electromechanoacoustic problem. In terms of a scientific approach this is a multidisciplinary problem due to involvement of multiple energy conversions, and its solution has to synthesize results of different theoretical disciplines such as electrostatics, electromechanical conversion, vibration of mechanical systems, and acoustic radiation. In terms of practical applications electroacoustic transducers should be considered as a part of the transmit or receive system, which requires their representation in the form of an equivalent electrical circuit that adequately reflects the properties of the transducer as a member of the system. Thus, the solution to the transducer problem has to be described in the generalized coordinates such as real electrical currents and voltages or their mechanical analogs. Their exist two ways to approach this problem: by the ''Newtonian'' and ''Lagrangian'' methods.

In the case of the Newtonian approach the partial differential equations of the elastic body vibration have to be solved under electrical boundary conditions defined by the piezoelectric equations of state and the mechanical boundary conditions defined in the course of simultaneous solving of the associated acoustic radiation problem. In this approach the solution of the problem is described in terms of the physical (geometrical) coordinates and the basic variables are displacement and force.

In the case of the Lagrangian (analytical mechanics) or energy approach, the variational principles have to be applied to the transducer as a complex electromechanoacoustic system undergoing multiple energy conversions. A solution to the problem is described by the Lagrange-type equations in generalized coordinates. The basic variables are displacement and energy.

The first approach is most widely practiced in application to piezoelectric elctromechanical systems. The most cited books in the field[1–3] and the vast majority of papers related to analysis of piezoelectric transducers including many recent publications in the Journal of the Acoustical Society of America employ this approach. Notably, most of these papers are related to the electromechanical transducers alone and do not consider the acoustical loading effects at all or include these effects by observation in the final results. An inherent hardship of this approach is that description of the interaction of the transducer subsystems may, in general, require different geometrical coordinates. Namely, the crystallographic coordinates for electric field calculation, the coordinate system compatible with mechanical system boundary conditions, and the coordinate system appropriate for describing the generated acoustic field can be different. Also a separate procedure is required in order to interpret the obtained results in the form of an equivalent electromechanical circuit, which in principle must be described in generalized coordinates.

The energy approach for solving piezoelectric transducers problems is not as common, though it has certain advantages in the analysis of complex systems with multiple energy conversions such as electroacoustical transducers. Thus, the solutions afforded by the energy methods are inherently described in the generalized coordinates, and complications arising from a mismatch of geometrical coordinate systems are avoided. The resulting equations may be directly interpreted in terms of the equivalent electromechanical circuits.

The examples of application of the Lagrange equations to flexural mode piezoelectric transducers was given in Refs.

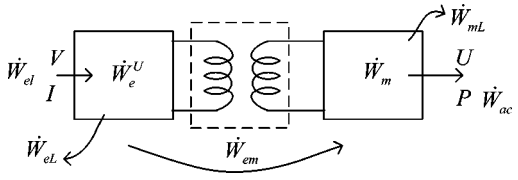


FIG. 1. The block-diagram of a transducer as an energy conversion system in the transmit mode: V and I are the input voltage and current, U and P are the output velocity and sound pressure in the acoustic field.

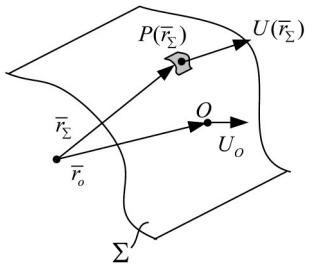


FIG. 2. Radiating surface of the transducer: U_o is the velocity of the reference point O.

4 and 5. The more extended energy analysis of electroacoustic systems was demonstrated in Refs. 6 and 7, which are unfortunately hardly accessible for the majority of the acoustic community.

The main objective of this paper is to present a general energy approach to piezoelectric electroacoustic transducers theory and to illustrate this technique with particular examples.

One of the reasons for the ''underdevelopment'' of the energy methods in piezoelectric transducers theory may be the lack of a convenient expression for the Lagrangian of a piezoelectric electroacoustic system, which in its turn requires an expression for internal energy of a deformed piezoelectric body in a suitable form. The expression for the Lagrangian is introduced in Sec. IV of this paper by employing the results of analysis of the internal energy of a piezoelectric body presented in Ref. 8 and by taking into consideration the energy balances that take place in the course of electroacoustic transducer operation. The block diagram of a transducer as an energy converting system and the energy balances in the course of the transducer operation are considered in Secs. II and III. In Sec. III also an important particular case of electroacoustic transducer having one mechanical degree of freedom is considered and the equivalent electromechanical circuit of the transducer is introduced by employing the energy conservation law. The general case of the electroacoustic transducer with multiple degrees of freedom is considered in Sec. IV. The Lagrangian of the system is introduced and the *Least Action Principle* is applied to obtain the Lagrange-type equations, that describe the transducer operation in generalized coordinates. Multicontour equivalent circuit of the transducer is introduced, for which these equations constitute the Kirchhoff equations. In Sec. V the Lagrange equations and the multicontour equivalent circuit are used to analyze the electroacoustic multimode ring transducer as an example of application of the energy method.

II. BLOCK DIAGRAM OF THE ELECTROACOUSTIC TRANSDUCTION

The operation of an electroacoustic transducer involves multiple energy conversions. This is illustrated with the block diagram of electroacoustic transduction in the transmit mode shown in Fig. 1. The energies and energy fluxes (powers) involved in the transduction are denoted as W and $\dot{W}$ with subscripts corresponding to the particular energy form. Thus, $W_{\rm el}$ is the total electric energy supplied to the transducer; W_e^U is the part of the electrical energy stored on the electrical side (block 1) of the ''blocked'' transducer (i.e., under the condition that the transducer can not vibrate); $W_{\rm eL}$ is the energy of electrical losses; $W_{\rm em}$ is the part of supplied electrical energy, which is transformed into the mechanical energy W_m by means of an employed mechanism of electromechanical conversion; W_m is the energy of vibration of the transducer mechanical system (block 3); $W_{\rm mL}$ is the energy of mechanical losses; $W_{\rm ac}$ is the acoustical energy radiated by the vibrating transducer surface. The mechanism of electromechanical conversion is represented in Fig. 1 by the ideal electromechanical transformer (block 2) of turns ratio n.

The energy of mechanical system vibration W_m may be represented as

$$W_m = W_{\rm kin} + W_{\rm pot}^V, \tag{1}$$

where $W_{\rm kin}$ is the kinetic energy and $W_{\rm pot}^V$ is the potential energy of the mechanical system at constant electrical input voltage.

Arrows in Fig. 1 show the directions of the spatial energy flow: The incoming fluxes are directed inside the blocks and are considered as conditionally positive, the outgoing fluxes are directed outside and are considered as conditionally negative.

Assuming the harmonic time dependence of transducer operation and omitting the time depending factor $e^{j\omega t}$, we will use the complex representation of power denoted as $\bar{\dot{W}}$. Thus, in particular we have

$$\bar{\dot{W}}_{\rm el} = IV^*, \tag{2}$$

where V and I are the complex amplitudes of applied voltage and current through the transducer input, and $(*)$ is the sign of the complex conjugate.

The acoustic power radiated by a transducer can be found by integrating the acoustic power density $P(\bar{r}_\Sigma)V^*(\bar{r}_\Sigma)d\Sigma$ over the transducer's radiating surface Σ, as shown in Fig. 2.

In the complex form the acoustic radiating power is

$$\bar{\dot{W}}_{\rm ac} = \int_\Sigma P(\bar{r}_\Sigma)U^*(\bar{r}_\Sigma)d\Sigma = U_o^* \int_\Sigma P(\bar{r}_\Sigma)\,\theta(\bar{r}_\Sigma)d\Sigma. \tag{3}$$

The distribution of velocity over the transducer surface is introduced in formula (3) as

$$U(\bar{r}_\Sigma) = U_o\,\theta(\bar{r}_\Sigma), \tag{4}$$

where $\bar{r}_\Sigma$ is the radius vector defining the points on the transducer surface and U_o is the velocity of the reference point O having radius vector $\bar{r}_o$. Function $\theta(\bar{r}_\Sigma)$ is the normalized

distribution of velocity on the surface. In general $\theta(\bar{r}_\Sigma)$ may be frequency dependent.

The sound pressure in the radiated acoustic field $P(\bar{r})$ and on the transducer surface in particular $P(\bar{r}_\Sigma)$, has to be determined by means of the acoustic radiation theory so far as distribution of velocity on the transducer surface is known. Strictly speaking, the calculation of sound field is not normally a part of the transducer design problem, though the electromechanical transducer does not exist as an electroacoustic device without an acoustic load. The acoustic power radiated can be represented in the form

$$\bar{\dot{W}}_{\mathrm{ac}} = Z_{\mathrm{ac}} |U_o|^2, \tag{5}$$

which can be considered as the definition for the transducer radiation impedance Z_{ac}. Equating expressions (3) and (5) we obtain

$$Z_{\mathrm{ac}} = \frac{1}{U_o} \cdot \int_\Sigma P(\bar{r}_\Sigma)\, \theta(\bar{r}_\Sigma)\, d\Sigma. \tag{6}$$

The sound pressure generated by a vibrating surface may be represented in general as

$$P(\bar{r},\omega) = \frac{\rho c}{r} U_o e^{-j(kr-\pi/2)} \chi(\bar{r},\omega), \tag{7}$$

where function $\chi(\bar{r},\omega)$ depends on the radiating surface configuration and mode of vibration.

We will refer to this function as the diffraction function of the transducer in the transmit mode. In the far field this function becomes independent of r and represents the directionality of transducer. For instance, for uniformly vibrating (pulsating) spherical radiator of radius a

$$P(\bar{r},\omega) = \frac{\rho c}{r} e^{-j(kr-\pi/2)} U_o \frac{ka^2}{1+jka} e^{-ka} \tag{8}$$

(see Ref. 9), and the diffraction function is

$$\chi = \frac{(ka)^2}{1+jka}. \tag{9}$$

Let us introduce for an arbitrary vibrating surface the concept of the ''referred volume velocity,'' U_{Vr}, defined as

$$U_{\tilde{V}r} = U_o S_\Sigma, \tag{10}$$

where S_Σ is the total area of the radiating surface. This quantity does not depend on the mode of vibration in contrast to the ''real'' volume velocity, or to the source strength, which is defined as

$$U_{\tilde{V}} = U_o \int \theta(\bar{r})\, d\Sigma = U_o S_{\mathrm{av}},$$

(Here and further symbol $\tilde{V}$ stands for a volume, S_{av} is introduced as the average area of a vibrating surface). Thus, for the oscillating sphere $U_{\tilde{V}r} = U_o 4\pi a^2$,

whereas $U_{\tilde{V}} = 0$. In the case that the wave size of the sphere is small $(ka \to 0)$, from expressions (8) and (10) it follows that sound pressure generated by the small sphere is

$$P_0 = \frac{\rho c}{r} e^{-j(kr-\pi/2)} U_o ka^2 = \frac{\rho c}{r} e^{-j(kr-\pi/2)} U_{\tilde{V}r} \frac{1}{2\lambda}. \tag{11}$$

We define the ratio of the sound pressure generated by an arbitrary transducer to the sound pressure generated by small pulsating sphere having the same referred volume velocity, as the diffraction coefficient of the transducer in the transmit mode, $k_{\mathrm{dif}.t}$. Using expressions (7) and (10), we obtain

$$\frac{P(\bar{r},\omega)}{P_0(U_{\tilde{V}r})} = \frac{2\lambda \chi(\bar{r},\omega)}{S_\Sigma} = k_{\mathrm{dif}.t}. \tag{12}$$

In the case of the spherical transducer

$$k_{\mathrm{dif}.t} = \frac{1}{1+jka} e^{jka}. \tag{13}$$

III. ENERGY BALANCES AND THE EQUIVALENT ELECTROMECHANICAL CIRCUIT OF A TRANSDUCER WITH SINGLE MECHANICAL DEGREE OF FREEDOM

By applying the energy conservation law the following balances of energy fluxes may be formulated for the electrical and mechanical transducer subsystems. The energy balance associated with block 1 (electrical subsystem of the transducer), may be represented as

$$\bar{\dot{W}}_{\mathrm{el}} = \bar{\dot{W}}_e^U + \bar{\dot{W}}_{\mathrm{eL}} + \bar{\dot{W}}_{\mathrm{em}}. \tag{14}$$

The energy balance associated with block 3 in Fig. 1 (mechanical subsystem of the transducer), may be represented as $\bar{\dot{W}}_m = \bar{\dot{W}}_{\mathrm{em}} - \bar{\dot{W}}_{\mathrm{mL}} - \bar{\dot{W}}_{\mathrm{ac}}$ or, also taking into account equality (1),

$$\bar{\dot{W}}_{\mathrm{em}} = \bar{\dot{W}}_{\mathrm{kin}} + \bar{\dot{W}}_{\mathrm{pot}}^V + \bar{\dot{W}}_{\mathrm{mL}} + \bar{\dot{W}}_{\mathrm{ac}}. \tag{15}$$

The energy balances (14) and (15) are valid for any electroacoustic transducer. So far as the piezoelectric transducers are concerned, all of the transducer mechanical system body or at least part of it is made of piezoelectric ceramic. In this case the energies W_e^U, W_{pot}^V, and W_{em} are the components of internal energy of the piezoelectric body. The internal energy of an arbitrary deformed piezoelectric body is analyzed in Ref. 8, where the rigorous definitions of the energy components W_e^S, W_m^E, and W_{em} are given. Thus, the energies W_e^S and W_m^E are equivalent to W_e^U and W_{pot}^V in our notations and they will be used further, as only piezoelectric transducers will be considered. As is shown in the previous paper,[8] the electrical and mechanical variables in the expression of internal energy can be considered as separable in most practical cases of transducer designs, and the components W_e^S and W_m^E can be calculated separately, as if it would be for a passive dielectric body having ''blocked'' dielectric constants (at $S=0$) and ''short-circuited'' elastic constants (at $E=0$). After the electric field and strain distribution are found in the body with prescribed dielectric and elastic constants of piezoelectric material, the energy W_{em} can be calculated as well by formulas given previously. Now we suppose the following:

(1) The electric field in the blocked transducer body is known (which has to be obtained by a solution of the electrostatic problem, because the dimensions of the mechanical systems in the frequency range of underwater

applications are much smaller than the electromagnetic wavelengths), and W_e^S and $\dot{\overline{W}}_e^S$ are expressed as

$$W_e^S = \frac{V^2 C_e^S}{2}, \quad \dot{\overline{W}}_e^S = j\omega C_e^S V V^*, \tag{16}$$

where C_e^S is the capacitance of the blocked transducer.

(2) The distribution of displacement in the mechanical system of the transducer considered as passive is known and is expressed as

$$\xi(\overline{r}_\Sigma, \omega) = \xi_0 \theta(\overline{r}_\Sigma). \tag{17}$$

With the displacements in the mechanical system known, the energies W_{kin} and W_m^E can be determined by integrating the energy densities over the transducer body and finally represented as $W_{\text{kin}} = \dot{\xi}_o^2 M_{\text{eqv}}/2$ and $W_m^E = \dot{\xi}_o^2/2C_m^E = \dot{\xi}_o^2 K_m^E/2$, where M_{eqv} and $C_m^E(K_m^E)$ are the equivalent mass and compliance (rigidity) of the transducer, or in the complex form

$$\dot{\overline{W}}_{\text{kin}} = j\omega M_{\text{eqv}} U_o U_o^*, \quad \dot{\overline{W}}_m^E = \frac{U_o U_o^*}{j\omega C_m^E}. \tag{18}$$

The electromechanical energy W_{em} can be represented as $W_{\text{em}} = v\ \xi_o\ n/2$, where n is the electromechanical transformation coefficient (see Ref. 8), or in the complex form as

$$\dot{\overline{W}}_{\text{em}} = V U_o^* n = V^* U_o n. \tag{19}$$

In this relation term Vn can be interpreted as the electromechanical force F_{em} applied to the mechanical system of the transducer. After substituting in Eq. (1) the expressions for energies (2), (16), (19), and $\dot{\overline{W}}_{\text{eL}}$, which can be represented as

$$\dot{\overline{W}}_{\text{eL}} = \frac{V V^*}{R_{\text{eL}}}, \tag{20}$$

we arrive at

$$I = \left(j\omega C_e^S + \frac{1}{R_{\text{eL}}} \right) V + U_o n. \tag{21}$$

After substituting the expressions for energies (5), (17), and (19) and for the energy of mechanical losses $\dot{\overline{W}}_{\text{mL}}$, which can be represented as

$$\dot{\overline{W}}_{\text{mL}} = r_{\text{mL}} U_0 U_0^*, \tag{22}$$

into formula (15) we arrive at

$$Vn = \left(j\omega M_{\text{eqv}} + \frac{1}{j\omega C_m^E} + r_{\text{mL}} + Z_{\text{ac}} \right) U_o. \tag{23}$$

Equations (21) and (23) relate the input quantities V and I and velocity of the reference point U_o and thus they describe the operation of the transducer at each frequency (given that the distribution of velocity on the transducer surface and the associated sound field are known). These equations can be considered as the Kirchhoff equations for the electrical and mechanical contours of the electromechanical circuit shown in Fig. 3, which makes the circuit their equivalent representation. The parameters of the mechanical branch in this circuit have to be considered as frequency dependent,

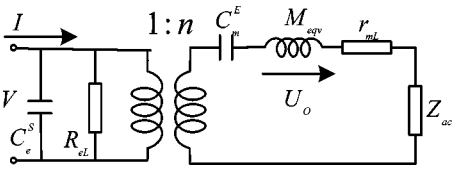


FIG. 3. The equivalent electromechanical circuit of electroacoustic transducer: U_o is the velocity of the reference point on the transducer surface.

unless the mode of vibration $\theta(\overline{r}_\Sigma)$ does not change in the operational frequency range that the transducer is used. Fortunately, the mechanical systems of transducers very often can be considered as vibrating with fixed velocity distribution in the operational frequency range, in which case only the velocity of the reference point needs to be determined in order to solve the vibration problem. Such systems can be considered as having one mechanical degree of freedom. For example, the following piezoelectric mechanical systems of wide spread use can be considered as having one mechanical degree of freedom:

- Uniformly loaded thin piezoelectric ceramic rings and spheres under uniform electrical excitation (fixed velocity distribution in this case follows from the symmetry considerations);
- Mechanical systems of transducers operating in the vicinity of their resonant frequency (usually for projectors), in which case the mode of vibration is the normal resonant mode of the mechanical system;
- Mechanical systems of transducers operating below the first resonant frequency (usually for receivers), in which case the velocity distribution may be approximated with sufficient accuracy by the fundamental mode of vibration or by the deflection curve that is determined under the condition of uniform static loading of the mechanical system.

For transducers with one mechanical degree of freedom in a prescribed operational frequency range the parameters of the equivalent circuit in Fig. 3 may be considered frequency independent, which greatly simplifies its application as a tool for calculation of the transducer operation.

In order to illustrate this technique consider the example of the transducer with a mechanical system in the shape of a thin spherical shell made of piezoelectric ceramic, fully electroded on the inner and outer surfaces, and poled in the radial direction, as shown in Fig. 4. An alternative approach to calculating of the piezoelectric spherical shell electrome-

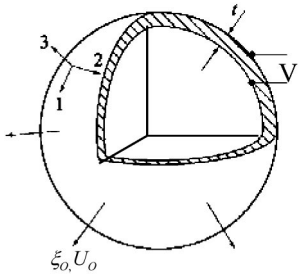


FIG. 4. Illustration of a spherical transducer.

chanical parameters is demonstrated in Ref. 2. The standard notations of the terms of piezoelectric equations as they are defined in Ref. 2 will be used further.

Under the assumption that $t \ll 2a$, where a is the average radius of the shell, and that the inner and outer surfaces of the shell are free of normal stresses, it can be concluded that in the body of the sphere $T_3=0$ and $E_3=V/t$. Due to symmetry the stresses $T_1=T_2=T_{1,2}$ and all the shear stresses are zero, $T_4=T_5=T_6=0$. The "working" deformations are $S_1=S_2=S_{1,2}$. The deformation $S_{1,2}$ can be found as $S_{1,2}=[4\pi(a+\xi_o)^2-4\pi a^2]/4\pi a^2=\xi_o/a$. The piezoelectric equations with T and E_3 as independent variables in this case become

$$S_{1,2}=(s_{11}^E+s_{12}^E)T_{1,2}+d_{31}E_3, \tag{24}$$

$$D_3=2d_{31}T_{1,2}+\varepsilon_{33}^T E_3. \tag{25}$$

From these equations we obtain

$$D_3=\frac{2d_{31}}{s_{11}^E+s_{12}^E}S_{1,2}+\varepsilon_{33}^{S_{1,2}}E_3, \tag{26}$$

where $\varepsilon_{33}^{S_{1,2}}=\varepsilon_{33}^T(1-k_p^2)$ is the dielectric constant of a piezoelectric element "blocked" in the directions of deformation S_1 and S_2. The capacitance of the spherical shell blocked in directions 1 and 2 is $C_e^{S_{1,2}}=4\pi a^2\varepsilon_{33}^T(1-k_p^2)/t$.

The equivalent parameters for the spherical shell have to be calculated by considering its energy status. The kinetic energy of vibration and the equivalent mass are

$$W_{\rm km}=\tfrac{1}{2}\dot{\xi}y_o^2 4\pi a^2 t\rho=\tfrac{1}{2}\dot{\xi}_o^2 M_{\rm eqv}, \quad M_{\rm eqv}=4\pi a^2 t\rho. \tag{27}$$

The potential energy of the shell at $E_3=0$ is

$$W_m^E=\frac{1}{2}\int_V (S_1T_1+S_2T_2)d\tilde{V}=\frac{1}{2}8\pi a^2 t S_{1,2}T_{1,2}.$$

After substituting $T_{1,2}$ from Eq. (18) at $E_3=0$ and $S_{1,2}=\xi_o/a$ we arrive at

$$W_m^E=\frac{1}{2}\xi_o^2\frac{8\pi t}{s_{11}^E+s_{12}^E}=\frac{1}{2}\xi_o^2 K_m^E.$$

Thus, the equivalent rigidity of the spherical shell is

$$K_m^E=\frac{1}{C_m^E}=\frac{8\pi t}{s_{11}^E+s_{12}^E}. \tag{28}$$

The resonant frequency will now be obtained as

$$f_{ra}=\frac{1}{2\pi}\sqrt{\frac{K_m^E}{M_{\rm eqv}}}=\frac{1}{2\pi a\sqrt{\rho s_{11}^E}}\sqrt{\frac{2}{1+\sigma_1^E}}, \tag{29}$$

where the analog of the Poisson's ratio for the piezoelectric ceramic is denoted $\sigma_1^E=-s_{12}^E/s_{11}^E$. The electromechanical energy for the pulsating sphere will be found as

$$W_{\rm em}=\frac{1}{2}\int_{\tilde{V}}\frac{2d_{31}}{s_{11}^E d_{31}+s_{12}^E}(S_1+S_2)E_3 d\tilde{V}=\frac{1}{2}n\xi_o V,$$

and after substituting $S_1=S_2=\xi_o/a$ and $E_3=V/t$, we arrive at

$$n=8\pi d_{31}/(s_{11}^E+s_{12}^E). \tag{30}$$

After the capacitance and the equivalent compliance of the sphere are determined, the resistances of the electrical and mechanical losses may be represented as

$$R_{\rm eL}=\frac{1}{\omega C_e^S\tan\delta_e}=\frac{Q_e}{\omega C_e^S}, \quad r_{\rm mL}=\omega C_m^E\tan\delta_m=\frac{Q_m}{\omega C_m^E}, \tag{31}$$

where Q_e and Q_m are the electrical and mechanical quality factors of the piezoelectric material. Thus, all the equivalent parameters of a spherical resonator are determined. It remains to introduce the sound field related parameters of the spherical projector in order to complete the equivalent circuit of the transducer as an electromechanical device and to provide a link between the transducer surface velocity $U_o=\dot{\xi}_o$ and the generated acoustic field. These parameters are the radiation impedance $Z_{\rm ac}$ and the diffraction function $\chi(\omega,\overline{r})$ or the diffraction coefficient $k_{\rm dif.t}$, which are defined by formulas (6), (7), and (12). For the spherical projector the diffraction function and the diffraction coefficient are given by formulas (12) and (13). And the radiation impedance will be obtained from formula (6) at $\theta(\overline{r_\Sigma})=1$ as

$$Z_{\rm ac}=4\pi a^2(\rho c)_w\left[\frac{(ka)^2}{1+(ka)^2}+j\frac{ka}{1+(ka)^2}\right]. \tag{32}$$

The data obtained in this example are sufficient to fulfill the calculation of a spherical projector. The velocity U_o may be found from the equivalent circuit of Fig. 3 and after this the sound pressure in acoustic field can be determined from formula (8).

IV. TRANSDUCERS WITH MULTIPLE MECHANICAL DEGREES OF FREEDOM, THE MULTI-CONTOUR EQUIVALENT CIRCUIT

In general, the mechanical system of an electroacoustic transducer has to be considered as the system with multiple degrees of freedom. In other words, the displacement distribution in the mechanical system may be represented as a function of a number of independent variables (generalized coordinates). Thus, if the actual displacement distribution $\xi(\overline{r_\Sigma},\omega)$ is expressed as

$$\xi(\overline{r_\Sigma},\omega)=\sum_{i=1}^{N}\xi_i(\omega)\cdot\theta_i(\overline{r_\Sigma}), \tag{33}$$

where $\theta_i(\overline{r_\Sigma})$ is a set of linearly independent functions satisfying the boundary conditions for the mechanical system, then the quantities $\xi_i(\omega)$ can be considered as the generalized coordinates for the system. In this case all the energies involved will be expressed through the generalized coordinates ξ_i and the generalized velocities $\dot{\xi}_i$. Thus, for example, the kinetic energy can be represented as

$$W_{\text{kin}}=\frac{1}{2}\int_{\tilde{V}}\rho\dot{\xi}(\overline{r_\Sigma},\omega)^2 d\tilde{V}$$
$$=\frac{1}{2}\int_{\hat{V}}\rho\left[\sum_1^N \xi_i\theta_i(\overline{r_\Sigma})\right]^2 d\tilde{V}$$
$$=\frac{1}{2}\sum_{i=1}^N\sum_{l=1}^N M_{il}\dot{\xi}_i\dot{\xi}_l,\quad M_{il}=\frac{1}{\dot{\xi}_i}\frac{\partial W_{\text{kin}}}{\partial\dot{\xi}_l},\tag{34}$$

where M_{il} can be considered as the equivalent masses. The example of representation of all the other energies will be given in Sec. V.

Equations of motion for a system with multiple degrees of freedom can be derived from the variational *Least Action Principle*.[10] According to this principle the equations of motion of the system can be found from the following condition of minimizing the function represented by the integral of the *Lagrangian L* of the system taken between the fixed initial (t_0) and final (t_1) moments of time.

$$\delta\int_{t_0}^{t_1} L(\xi_1,\xi_2,\ldots,\xi_N;\dot{\xi}_1,\dot{\xi}_2,\ldots,\dot{\xi}_N)dt=0.\tag{35}$$

The Lagrangian is a function of the generalized coordinates of the system and their derivatives with respect to time). The Lagrangian has to be determined for the particular system. The equations of motion may be obtained from the condition (35) by means of the calculus of variations and they are known as the Euler equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\xi}_i}\right)-\frac{\partial L}{\partial\xi_i}=0,\quad \text{for}\quad i=1,2,N.\tag{36}$$

In order to apply Eq. (36) to the mechanical system of the electroacoustic transducer, a proper Lagrangian has to be used. Based on the energy balance expressed by Eq. (15), we suggest this function to be

$$L=W_{\text{kin}}-W_m^E+(W_{\text{em}}-W_{\text{mL}}-W_{\text{ac}}).\tag{37}$$

When choosing the function (37), the analogy with the classical dynamics is used, in which case for the passive mechanical systems the Lagrangian is known as

$$L=W_{\text{kin}}-W_{\text{pot}}+W_e.\tag{38}$$

As for this case the equations of the type of Eq. (36) are commonly called the Lagrange equations, we will also retain this name for Eq. (36), when they are related to the electromechanoacoustical systems. In Eq. (38) the term W_e is the work done by external forces. The terms in parenthesis in Eq. (37) also represent energies due to the external actions for transducer operating in the transmit mode. When determining the signs of spatial fluxes of energy, the rule is applied according to which the incoming energy flux W_{em} is considered to be positive and the outgoing energy fluxes W_{mL} and W_{ac} are considered to be negative. After substituting the Lagrangian in the form of relation (37) into Eq. (36) and after transferring to the complex form, we arrive at the equations for an electroacoustic transducer mechanical system vibration in the form

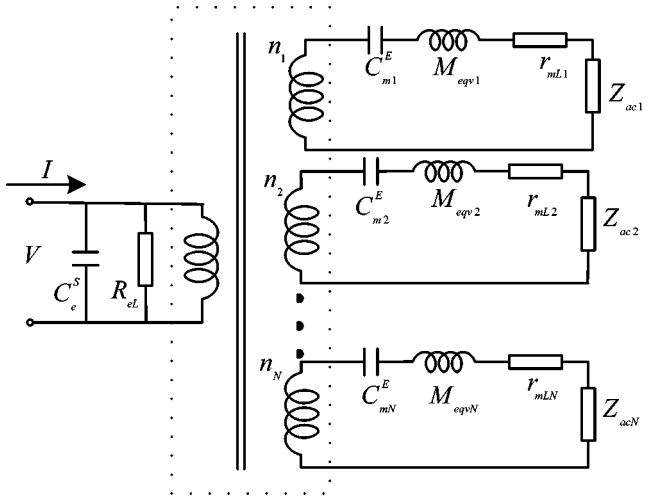


FIG. 5. The equivalent electromechanical circuit of an electroacoustic transducer with multiple degrees of freedom.

$$(j\omega M_{\text{eqv}i}+1/j\omega C_{mi}^E+r_{\text{mL}i}+Z_{\text{aci}})U_i=Vn_i,$$
$$\text{for}\quad i=1,2,N,\tag{39}$$

where the equivalent parameters with number i are attributed to the generalized velocity U_i. The equation for electrical subsystem may be obtained from the energy balance (7) in the same way as Eq. (21) was obtained, but in this case expression (19) for $\overline{\dot{W}}_{\text{em}}$ has to be replaced by

$$\overline{\dot{W}}_{\text{em}}=V^*\sum_1^N U_i n_i,\tag{40}$$

taking into account the representation of the displacement (33). In the result the following equation will be obtained:

$$I=\left(j\omega C_e^S+\frac{1}{R_{\text{eL}}}\right)V+\sum_1^N U_i n_i.\tag{41}$$

Equations (39) and (41) can be considered as the Kirchhoff equations for the related equivalent electromechanical circuit of the transducer with multiple mechanical degrees of freedom, which is represented in Fig. 5 as a set of mechanical contours, each corresponding to a generalized velocity and coupled to the electrical side of the transducer by ideal electromechanical transformers of the turns ratio n_i.

The important question is how to choose a suitable set of the supporting functions $\theta_i(\overline{r})$ in order to represent the displacement distribution in the form of Eq. (33). The clarity of interpretation of results and computational problems in solving Eqs. (39) depend completely on the choice of the set of supporting functions. Equations (39) have the simplest form, if the functions $\theta_i(\overline{r})$ are the normal modes of vibration of the real transducer mechanical system and under the real loading conditions. In this case the functions $\theta_i(\overline{r})$ are orthogonal and Eqs. (39) and the contours in Fig. 5 are independent. The parameters $M_{\text{eqv}i}$, C_{mi}^E, n_i, and Z_{aci} are in this case the modal parameters calculated for a particular normal mode $\theta_i(\overline{r})$.

Although very convenient for interpretation of the Lagrange equations, the choice of normal modes under real loading conditions has some significant disadvantages. First,

it is hard to obtain the normal modes under real loading conditions, because both the mechanical system vibration and acoustic radiation problems are involved. Second, the evaluation of equivalent parameters should be repeated each time the load conditions change. At the same time it is known that the modes of vibration of the typical mechanical systems of underwater electroacoustic transducers do not change significantly under the action of real acoustical loads.

Therefore, it is reasonable to formulate the problem in terms of *in vacuo* normal modes of the mechanical system. In this case Eqs. (39) may become coupled because of acoustical loading. Formally it will result in Z_{aci} becoming

$$Z_{\mathrm{aci}} = Z_{\mathrm{acii}} + \sum_{i \neq l} z_{\mathrm{acil}} \cdot \frac{U_l}{U_i}, \tag{42}$$

where Z_{acii} is the modal self-radiation impedance and z_{acil} are the mutual intermodal radiation impedances. For a particular transducer type and loading conditions (for example, in the case of the application of different baffles) the self- and inter-modal impedances can be calculated separately by means of the theory of acoustic radiation. The resulting acoustic coupling between the mechanical contours of the equivalent circuit in Fig. 5 [between Eqs. (39)] can be taken into account by introducing Z_{aci} in the form of expression (42).

One more complication in deriving the normal modes of the transducer mechanical system is related to anisotropy of the piezoelectric material. The normal modes of the mechanical system of the same configuration and under the same boundary conditions, strictly speaking, may be different, if it is made of different piezoelectric ceramic compositions. Further simplification can be achieved, if to employ as the set of the supporting functions for a real anisotropic mechanical system of a transducer, the normal modes of analogous mechanical system but made of isotropic passive material. This may result in some additional coupling between Eqs. (39) and between the contours of the equivalent circuit in Fig. 5. But it appears that the effect of this coupling on the parameters calculated on the basis of *in vacuo* normal modes of mechanical systems made of isotropic material is small enough, at least so far as the modern piezoelectric ceramic compositions are concerned. In any event, the contribution of this effect can be easily estimated and taken into account in the form of a correction to the main part of a solution.

Summarizing the above considerations the outline for the application of the energy method for solving electroacoustic transducer problems may be formulated as follows:

(1) Determination of the electric field in the piezoelectric body of the transducer by solving the electrostatic problem for the body as a passive dielectric with ''blocked'' dielectric constant ε_{ii}^{S}.
(2) Determination of the *in vacuo* normal modes of an isotropic passive mechanical system analogous to the real mechanical system of the transducer made of piezoelectric material and under the same boundary conditions, or determination of the deflection curve (mode of vibration) for the said passive mechanical system in the case that the system can be considered as having one degree of freedom in the frequency range under consideration.
(3) Calculation of the components of the internal energy of the transducer body (W_e^S, W_m^E, W_{em}, ΔW) by formulas given in Ref. 8 using the known normal modes (or deflection curve) and the characteristics of the real active material of the transducer, and determination of the modal equivalent transducer parameters.
(4) Calculation of the field generated by the transducer with a known mode of velocity distribution by means of the theory of the acoustic radiation, and determination of the acoustical load Z_{aci} including the modal self-radiation impedances Z_{acii} and intermodal mutual impedances z_{acil}.

To this end the problem can be considered completed (at least to the first approximation). The equivalent circuits in Figs. 3 and in 5 with thus determined circuit elements may be used for calculation of the transducer's electroacoustic parameters. In conclusion, an estimation has to be made of the effects of coupling between Eqs. (39) due to nonorthogonality of the set of supporting functions.

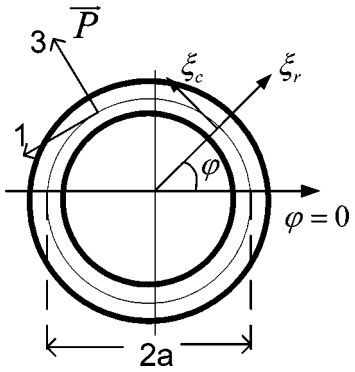


FIG. 6. Illustration of the extensional vibrations of a general type in the plane of a ring: ξ_r and ξ_c are the displacements in the radial (axis 3) and circumferential (axis 1) directions.

V. MULTIMODE GYLINDRICAL TRANSDUCER COMPOSED OF PIEZOELECTRIC CERAMIC RINGS

In order to illustrate the multiple degrees of freedom transducer calculation, consider an example of multimode cylindrical transducer employing the extensional vibrations of the piezoelectric ceramic rings. The transducer with a mechanical system in the shape of the thin piezoelectric ceramic rings is a typical example of a transducer with one mechanical degree of freedom under assumption of uniform excitation and uniform acoustical loading. If to reject these assumptions, then the symmetry considerations are not valid and the ring may vibrate in a general extensional mode with displacements taking place in the radial and circumferential directions as illustrated in Fig. 6.

The ring is assumed to be thin ($t \ll 2a$) and of small height h ($h \ll 2a$). Under these assumptions it can be considered that all the stresses in the ring other than the circumferential stress T_1 are equal to zero, and the piezoelectric equations for the case become

$$S_1 = s_{11}^E T_1 + d_{31} E_3, \tag{43}$$

$$D_3 = d_{31} T_1 + \varepsilon_{33}^T E_3. \tag{44}$$

It is known[11] that so far as the extensional vibrations of the ring are concerned the strain S_1 in the circumferential direction is related to the displacements ξ_r and ξ_c as follows:

$$S_1 = \frac{\xi_r}{a} + \frac{1}{a}\frac{\partial \xi_c}{\partial \varphi}, \tag{45}$$

and the following relation between the displacements is fulfilled:

$$\frac{\partial \xi_r}{\partial \varphi} + \xi_c = 0. \tag{46}$$

An arbitrary distribution of the radial displacement on the ring surface can be represented as

$$\xi_r(\varphi, \omega) = \sum_{i=0}^{\infty} \xi_i(\omega) \cos i\varphi, \tag{47}$$

where ξ_i is the radial displacement of the reference point at $\varphi = 0$.

According to Eqs. (46) and (47),

$$\xi_c(\varphi, \omega) = \sum_{i=1}^{\infty} i\xi_i \sin i\varphi. \tag{48}$$

Taking into account expressions (47) and (48) for ξ_r and ξ_c, S_1 from formula (45) becomes

$$S_1 = \frac{1}{a}\sum_{i}^{\infty} \xi_i (1 + i^2) \cos i\varphi. \tag{49}$$

Applying the piezoelectric Eqs. (43) at $E_3 = 0$ we obtain

$$T_1^E = \frac{1}{s_{11}^E} S_1 = \frac{1}{a \cdot s_{11}^E} \sum_{i=0}^{\infty} \xi_i (1 + i^2) \cos i\varphi. \tag{50}$$

Now the energy status and all the equivalent parameters of the ring can be determined.

The kinetic energy of the ring is

$$\begin{aligned} W_{\text{kin}} &= \frac{1}{2}\int_{\tilde{V}} \rho(\dot{\xi}_r^2 + \dot{\xi}_c^2) d\tilde{V} = \frac{\rho t h a}{2} \int_0^{2\pi} \left[\left(\sum_{i=0}^{\infty} \dot{\xi}_i \cos i\varphi \right)^2 \right. \\ &\quad \left. + \left(\sum_{i=1}^{\infty} i\dot{\xi}_i \sin i\varphi \right)^2 \right] d\varphi \\ &= \frac{1}{2} M \left[\dot{\xi}_0^2 + \frac{1}{2} \sum_{i=1}^{\infty} \xi_i^2 (1 + i^2) \right], \end{aligned} \tag{51}$$

where M is the mass of the ring. When evaluating the integral in relation (51), the orthogonality of the functions $\theta_i(\varphi) = \cos i\varphi$ on the interval 0 to 2π was used. According to Eq. (34) we obtain

$$M_{\text{eqvo}} = M, M_{\text{eqvi}} = M(1 + i^2)/2. \tag{52}$$

The potential energy of the ring at $E_3 = 0$ and equivalent rigidities, K_{mi}^E, can be calculated as

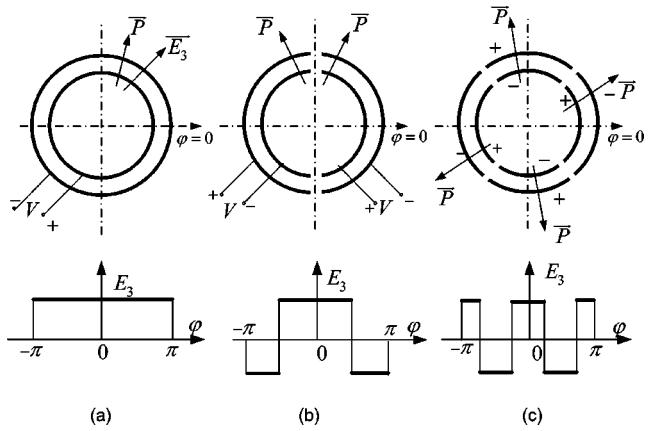


FIG. 7. Electric field configurations for excitation of different modes of ring extensional vibrations.

$$\begin{aligned} W_m^E &= \frac{1}{2}\int_{\tilde{V}} S_1 T_1^E d\tilde{V} \\ &= \frac{th}{2as_{11}^E} \int_0^{2\pi} \left[\sum_{i=0}^{\infty} \xi_i (1 + i^2) \cos i\varphi \right]^2 d\varphi \\ &= \frac{1}{2}\frac{2\pi th}{as_{11}^E} \left[\xi_0^2 + \frac{1}{2}\sum_{i=1}^{\infty} \xi_i (1 + i^2) \right] \end{aligned}$$

and

$$K_{mi}^E = \frac{1}{\xi_i}\frac{\partial W_m^E}{\partial \xi_i} \text{ or } K_{m0}^E = \frac{2\pi th}{as_{11}^E}, \quad K_{mi}^E = \frac{\pi th}{as_{11}^E}(1 + i^2)^2. \tag{53}$$

The electromechanical energy and electromechanical transformation coefficients n_i have to be calculated as follows. We denote the electric field in the ring as a function of φ, and noting $E_3(\varphi) = (V/t)\,\theta_E(\varphi)$, then

$$\begin{aligned} W_{\text{em}} &= \frac{1}{2}\int_{\tilde{V}} \frac{d_{31}}{s_{11}^E} S_1 E_3 d\tilde{V} \\ &= \frac{1}{2}\frac{d_{31} t h}{s_{11}^E}\frac{v}{t}\int_0^{2\pi} \theta_E(\varphi) \sum_{i=0}^{\infty} \xi_i (1 + i^2) \cos i\varphi d\varphi \\ &= \frac{1}{2} v \sum_{1}^{N} \xi_i n_i, \end{aligned}$$

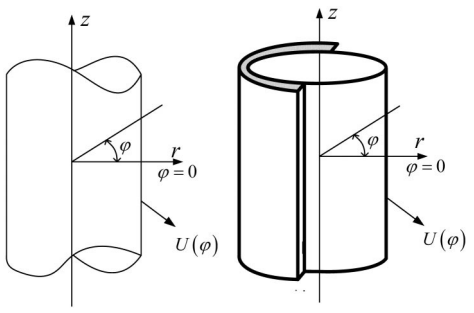


FIG. 8. A cylindrical radiator: (a) omnidirectional, uniform acoustic load; (b) unidirectional, with rigid baffle.

$$\bar{W}_{\mathrm{em}}=\frac{d_{31}h}{s_{11}^{E}}V^{*}\int_{0}^{2\pi}\theta_{E}(\varphi)\left[\sum_{i=0}^{\infty}U_{i}(1+i^{2})\cos i\varphi\right]d\varphi$$
$$=V^{*}\sum_{1}^{N}U_{i}n_{i}. \tag{54}$$

Thus,

$$n_{i}=\frac{1}{V^{*}}\frac{\partial\bar{\dot{W}}_{\mathrm{em}}}{\partial U_{i}}$$
$$=\frac{d_{31}h}{s_{11}^{E}}(1+i^{2})\int_{0}^{2\pi}\theta_{E}(\varphi)\cos i\varphi d\varphi$$
$$=\frac{2\pi d_{31}h}{s_{11}^{E}}(1+i^{2})\alpha_{i}, \tag{55}$$

where

$$\alpha_{i}=\frac{1}{2\pi}\int_{0}^{2\pi}\theta_{E}(\varphi)\cos i\varphi d\varphi \tag{56}$$

are the Fourier coefficients for the function $\theta_E(\varphi)$. Thus, the values of the modal transformation coefficients n_i depend on the configuration of electrodes on the ring surfaces, which is described by the function $\theta_E(\varphi)$. Several variants of the configuration of electrodes and the corresponding plots of functions $\theta_E(\varphi)$ are represented in Fig. 7.

Variant (a): $\theta_E(\varphi)=1$, uniform electrode, $\alpha_0=1$, $\alpha_i=0$ for $i\neq 0$. The only electromechanically active mode of vibration is the breathing mode $\theta_0(\varphi)=1$, and $n_0=2\pi d_{31}h/s_{11}^{E}$.

Variant (b): The electrodes are separated into two parts and the parts are connected electromechanically in opposite. (Electromechanically positive means that the electrical field and polarization are in the same direction, while electromechanically negative means that the electrical field and polarization are in the opposite directions.) Therefore,

$$\theta_{E}(\varphi)=\begin{cases}1 & |\varphi|<\dfrac{\pi}{2}\\ -1 & \dfrac{\pi}{2}<|\varphi|<\pi\end{cases}\quad\text{and}\quad\alpha_{0}=0,\quad\alpha_{\mathrm{i}}=(-1)^{m+1}\frac{2}{i\pi},\quad i=2m-1,\quad m=1,2,...,.$$

The first electromechanically active mode of vibration is $\theta_1(\varphi)=\cos\varphi$ and $n_1=8d_{31}h/s_{11}^{E}$.

Variant (c): The electrode is separated into four parts and the neighboring parts are electromechanically connected in opposite. We will obtain $\alpha_0=0$, $\alpha_1=0$, $\alpha_i=(-1)^{m+1}2/i\pi$ for $i=2(2m-1)$, $m=1,2,\ldots,$. The first active mode is $\theta_2(\varphi)=\cos 2\varphi$ and $n_2=10d_{31}h/s_{11}^{E}$.

In the same manner the modal electromechanical coefficients can be determined for an arbitrary electrodes configuration. The capacitance, $C_e^{S}=C_e^{S_1}$, of the transducer has to be calculated in a regular way using the dielectric constant $\varepsilon_{33}^{S_1}=\varepsilon_{33}^{T}(1-k_{31}^{2})$.

The acoustic load for the transducer depends on the boundary conditions on the radiating surface. After substituting the distribution of velocity $U(\bar{r}_{\Sigma})$ in the form of the series (47) under the integral in Eq. (3) we will obtain the acoustic radiated power per unit length for an infinitely long cylinder

$$\bar{\dot{W}}_{\mathrm{ac}}=a\int_{0}^{2\pi}P^{*}(a,\varphi)\left(\sum_{i=0}^{\infty}U_{i}\cos i\varphi\right)d\varphi. \tag{57}$$

The sound pressure radiated by a long cylinder having an arbitrary distribution of velocity in the circumferential direction is considered in Ref. 9. In any case it can be represented in the form

$$P(a,\varphi)=\sum_{l=0}^{\infty}A_{l}\cos l\varphi. \tag{58}$$

After substituting the sound pressure into Eq. (57) and integrating, we will obtain

$$\bar{\dot{W}}_{\mathrm{ac}}=\sum_{i=0}^{\infty}Z_{\mathrm{ac}i}U_{i}U_{i}^{*}. \tag{59}$$

In the case that $U(\varphi)=U_i\cos i\varphi$ at $0\leqslant\varphi\leqslant 2\pi$ [i.e., no baffles are applied as shown in Fig. 8(a)] the modal impedances $Z_{\mathrm{ac}i}$ are

$$Z_{\mathrm{ac}0}=2\pi aj\omega\rho\frac{H_{0}^{(2)}(ka)}{[H_{0}^{(2)}(kr)]'_{r=a}},$$
$$Z_{\mathrm{ac}i}=\pi a\times j\omega\rho\frac{H_{i}^{(2)}(ka)}{[H_{i}^{(2)}(kr)]'_{r=a}},\quad i=1,2,..., \tag{60}$$

where $H_0^{(2)}(kr)$ are the Hankel functions.

Thus all the parameters of the equivalent multicontour circuit in Fig. 5 are determined and the mechanical contours prove to be independent.

Now consider the acoustical load in the case that a part of the cylindrical surface at $|\varphi|>\pi/2$ is covered with a rigid baffle, as shown in Fig. 8(b). In this case the integration in relation (57) has to be fulfilled over the open part of cylindrical surface, and the result of the integration can be represented in the form

$$\bar{\dot{W}}_{\mathrm{ac}}=\sum_{i=0}^{\infty}Z_{\mathrm{ac}ii}U_{i}U_{i}^{*}+\sum_{i\neq l}^{\infty}z_{\mathrm{ac}il}U_{i}U_{l}^{*}. \tag{61}$$

The mutual terms corresponding to $i\neq l$ are included, be-

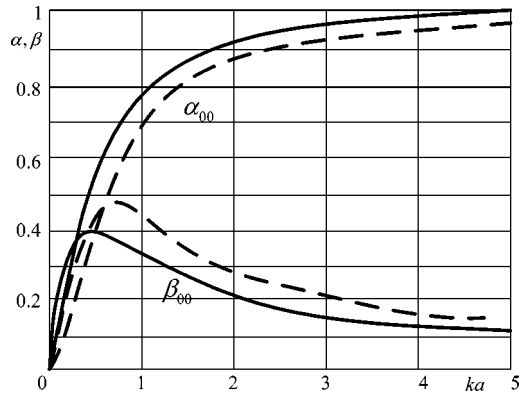


FIG. 9. Nondimensional coefficients of the radiation impedances of a cylindrical transducer in 0 mode of vibration (solid line: 0 mode, without baffle, $S_{\text{eff}}=2\pi a$; dashed line: 0 mode, with 180° baffle, $S_{\text{eff}}=\pi a$).

cause the supporting functions $\theta_i(\varphi)=\cos i\varphi$ are not orthogonal in the interval from $\pi/2$ to $-\pi/2$.

The acoustic load Z_{aci}, related to Eq. (39) with number i and to the corresponding contour of the equivalent circuit in Fig. 5, will be obtained by formula (48). And using relation (61) it can be represented in the form of Eq. (42). The term $z_{\text{aci}l}\cdot U_l/U_i$ in Eq. (42) is the coupled radiation impedance between equations number i and number l and the corresponding contours of the equivalent electromechanical circuit of the transducer. It is convenient to represent the self and mutual radiation impedances in the form $Z_{\text{aci}i}=(\rho c)_w S_{\text{eff}\,i}(\alpha_{ii}+j\beta_{ii})$ and $z_{\text{aci}l}=(\rho c)_w S_{\text{eff}\,i}(\alpha_{il}+j\beta_{il})$, where α and β are nondimensional coefficients, and

$$S_{\text{eff}\,i}=a\int_0^{2\pi}\theta_i^2(\varphi)d\varphi$$

will be defined as the effective surface area per unit length of a cylinder.

The nondimensional coefficients α and β for the self radiation and mutual radiation impedances for the most commonly used 0 and 1 modes of cylindrical transducer without baffle and with 180° rigid baffle coverage are represented in Figs. 9–11. (These results were obtained by using the technique similar to those applied in Ref. 9 for calculation of radiation of a cylindrical surface in the case that distribution

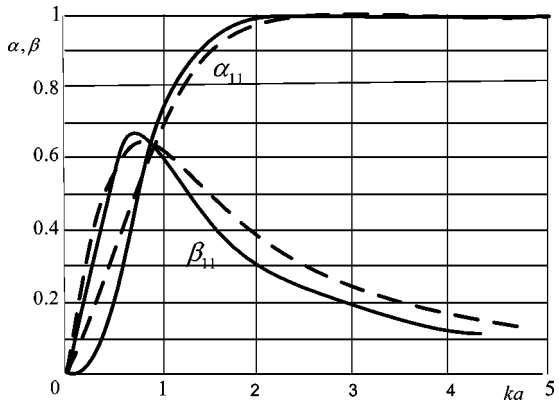


FIG. 10. Nondimensional coefficients of the radiation impedances of a cylindrical transducer in 1 mode of vibration (solid line: 1 mode, without baffle, $S_{\text{eff}}=\pi a$; dashed line: 1 mode, with 180° baffle, $S_{\text{eff}}=\pi a/2$).

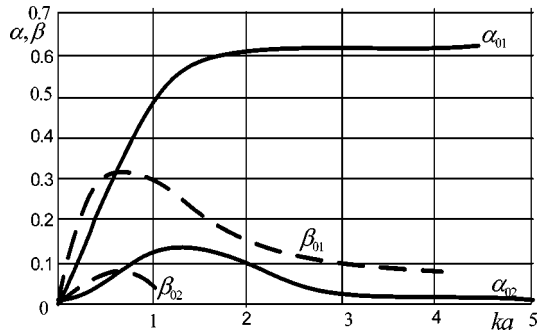


FIG. 11. Nondimensional coefficients α_{oi} and β_{oi} for the mutial radiation impedances between zero and other modes of a cylindrical transducer with 180° rigid baffle, $S_{\text{eff}}=\pi a/2$.

of velocity by circumference is not uniform.)

Note that even in the case of uniform electromechanical excitation, and therefore only 0 order mode being electromechanically active, all the other modes, strictly speaking, are involved in the baffled cylindrical transducer operation because of the coupling through the acoustical load. But practically only the contribution of the closest mode may be of significant importance, because the higher the mode (the larger the number l) the smaller the relative velocity U_l/U_0 in the operational frequency range and the smaller the coupled impedance $z_{\text{ac}0l}U_l/U_0$. In addition the mutual radiation impedance $z_{\text{ac}0l}$ between modes drops, when number l increases. For the mutual impedances z_{01} and z_{02} this is illustrated in Fig. 11.

In order to govern the frequency response of the transducers by underlining or suppressing of certain modes of vibration the technique described in Ref. 12 can be used. An example of an application of this technique to the first order cylindrical transducer is presented in Ref. 13.

The alternative approach to calculate the cylindrical transducer properties by solving the partial differential equations of the problem was demonstrated for the transducers operating in the dipole mode in Ref. 14 and for the baffled transducers with compliant baffles in Ref. 15.

VI. CONCLUSION

In conclusion several merits of the energy method approach to the electroacoustic piezoelectric transducers problems may be summarized as follows.

First, it is shown that by applying the described energy method the analysis of electroacoustical transducers may be reduced to the solving of separate independent problems (those of electrostatics, electromechanical conversion, vibration of mechanical systems, and acoustic radiation) by means of their corresponding theoretical disciplines and to synthesizing the results obtained in the Lagrange equations describing the transducer operation in generalized coordinates. These equations may be considered simultaneously as the Kirchhoff equations for a multicontour electromechanical equivalent circuit of the transducer. In this sense it has to be recognized that there is no need to consider the problem of calculation of the electroacoustic or electromechanical transducers as a special electroelastic problem as it may appear

from the viewpoint of the Newtonian (partial differential equations) approach.

The results related to electric fields in piezoelectric elements, to vibration of passive mechanical systems having the same shape as the mechanical systems of the piezoelectric transducers, and to radiation of sound by those systems in most cases can be taken as previously solved in electrostatics and in the theories of mechanical vibration and acoustic radiation, respectively. The only special issues are those related to the theory of electromechanical transformation or, in terms of the described energy approach, to the calculation of the electromechanical energy W_{em} and of the energy of electrical interaction between elements of the deformed piezoelectric body of the transducer ΔW. But both of the energies may be calculted in a straightforward way by formulas presented previosly,[8] so far as the electric field and mechanical vibration are known.

Second, employing the amplitudes of the normal modes of mechanical system vibration as the generalized coordinates makes interpretation of the transducer analysis especially physical clear.

The piezoelectric elements, of which mechanical systems of transducers are built, usually have a simple geometry and therefore the configuration of the electric field is basically simple (with the possible exception of special cases of tangentially polarized piezoelements with stripped electrodes, or partly electroded elements). In combination with known strain distributions for particular normal modes this makes it easy to optimize the electromechanical coupling for the chosen modes, as it was described previously,[12] and thus to govern the frequency response of the transducer in a predictable way. This feature is especially advantageous for designing a transducer with desirable operational properties. This problem may be defined as the ''reverse'' transducer problem. The problem of calculating parameters of a known transducer design being considered as the ''direct'' transducer problem. The physical transparency of the energy approach allows for the development of physical and enginering intuition rather than relying on numerical calculations.

Lastly, in the case that the normal modes for the analogous passive mechanical system are not known, the typical procedure for application of direct methods (such as the Rayleigh method for one degree of freedom systems or the Ritz method for multiple degrees of freedom systems) can be used in a straightforward way in the framework of the described technique. The difference will be that instead of the normal modes some set of linear independent trial functions satisfying the boundary conditions for the problem and suitable for calculations have to be used to represent the displacement distribution. The resulting equations in this case do not have a physical meaning and their equivalent multicontour interpretation is not applicable. They just form a set of equations for numerical calculations, which have to be fulfilled by using common computational methods.

ACKNOWLEDGMENTS

The author wishes to thank Dr. David A. Brown for his cooperation and assistance in revising and preparing the paper for publication. This work was supported in part by ONR 321SS Lindberg and BTECH Acoustics.

[1] W. P. Mason, *Electromechanical Transducers and Wave Filters*, 2nd ed. (Van Nostrand, Princeton, New Jersey, 1948).
[2] D. A. Berlincourt, D. R. Curran, and H. Jaffe, in *Piezoelectric and Piezomagnetic Materials and their Functions in Transducers*, in Physical Acoustics, edited by W. P. Mason, Vol. IA, (Academic Press, New York, 1964).
[3] R. Holland and E. P. EerNisse, *Design of Resonant Piezoelectric Devices* (MIT Press, Cambridge: 1969).
[4] F. J. Rosato, ''Lagrange equations applied to flexural mode transducers,'' J. Acoust. Soc. Am. **57**, 1397–1401 (1975).
[5] B. S. Aronov and L. B. Nikitin, ''Calculation of flexural modes of piezoceramic plates,'' Sov. Phys. Acoust. **27**(5), 382–387 (1981).
[6] T. Hayasaka, *Electroacoustics* (Mir, Moskow, 1982), (translated in Russian from Japanese).
[7] B. S. Aronov, *Piezoceramic Electromechanical Transducers* (Energoatomizdat, Leningrad, 1990), (in Russian).
[8] B. S. Aronov, ''Energy analysis of a piezoelectric body under nonuniform deformation,'' J. Acoust. Soc. Am. **113**, 2638–2646 (2003).
[9] P. M. Morse, *Vibration and Sound*, reprint of the 2nd edition by the Acoustical Society of America, Woodbury, NY, 1981.
[10] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[11] A. E. H. Love, *Mathematical Theory of Elasticity*, 4th ed. (Dover Reprint, New York, 1927).
[12] B. S. Aronov, ''On the optimization of the effective electromechanical coupling coefficients of a piezoelectric body,'' J. Acoust. Soc. Am. **114**(2), 792–800 (2003).
[13] B. S. Aronov, ''Calculation of first-order cylindrical piezoceramic receivers,'' Sov. Phys. Acoust. **34**(5), 466–470 (1988).
[14] R. S. Gordon, J. L. Butler, and L. Prasad, ''Equivalent circuit of a ceramic ring transducer operated in the dipole mode,'' J. Acoust. Soc. Am. **58**, 1311–1314 (1975).
[15] D. B. Dianov, I. M. Zadirienko, and A. G. Kuz'menko, ''Frequency response and directivity characteristics of a cylindrical transducer in a compliant baffle,'' Sov. Phys. Acoust. **19**(5), March–April (1974).

Annoyance with aircraft noise in local recreational areas and the recreationists' noise situation at home[a)]

Norun Hjertager Krog[b)] and Bo Engdahl[c)]
Norwegian Institute of Public Health, Division of Environmental Medicine, P.O. Box 4404 Nydalen, 0403 Oslo, Norway

(Received 10 June 2002; revised 7 July 2004; accepted 27 July 2004)

Few socioacoustic studies have examined the effect of noise on outdoor recreationists. Most studies concentrate on one setting of the everyday life of a noise-exposed population, which mainly has been the residential setting. This article relates annoyance with aircraft noise in outdoor recreational areas to the recreationists' noise situation at home. In conjunction with the relocation of the main airport of Norway in 1998, field studies were conducted before and after the change in one area near the old airport (1930 survey respondents), and one area near the new airport (1001 survey respondents). Multivariate linear regression analyses of the relationship between annoyance and aircraft noise exposure (L_{Aeq} for the aircraft events) in the recreational areas were conducted, controlled for noise annoyance at home, or aircraft noise exposure at home, the situation (before/after the change), context- and demographic variables. People more highly annoyed at home tended to be more annoyed than others while in the recreational areas. A significant effect of aircraft noise exposure at home on annoyance in the recreational setting was not found. More research is warranted regarding the relationship between noise exposure at home and outdoor recreational demands. [DOI: 10.1121/1.1795732]

PACS numbers: 43.50.Lj, 43.50.Qp [DWK] Pages: 221–231

I. INTRODUCTION

Although concern about adverse effects of noise on outdoor recreationists has grown (International Recreational Noise Symposium, 1998; FICAN Symposium on the Preservation of Natural Quiet, 1999), relatively few studies have been conducted that combine survey data with acoustic measurements from outdoor recreational areas (Anderson *et al.*, 1993; Fidell *et al.*, 1996; Fleming *et al.*, 1998; Miller *et al.*, 1999; Krog *et al.*, 2000). Some studies are social surveys without acoustical measurements of noise exposure (Cessford, 1998; McDonald *et al.*, 1995). Sutton (1999) does not have acoustic measurements, but combines survey data with estimates of number of aircraft overflights at different times of the day. Other studies have focused exclusively on describing the soundscapes of natural areas, without measuring the reactions of the visitors (Krause, 1999; Downing *et al.*, 1999). The main noise source of concern has been different kinds of aircraft activity, mainly scenic flights. The kinds of areas that have been studied are typically areas to which people travel for a distance to visit.

The relocation of the main airport of Norway on the 8th of October, 1998, offered the opportunity to conduct studies in local outdoor recreational areas affected by either a decrease or an increase in aircraft noise exposure. The study presented in this article is part of a larger study containing both a field- and laboratory experiment (Aasvang and Engdahl, 2004), field surveys with simultaneous collection of acoustical data (Krog and Engdahl, 2004; this article), and telephone surveys. The analyses that are presented in this article combine survey data and acoustical measurements from two different local outdoor recreational areas. In addition to information about the recreationists' reactions to the aircraft noise in the recreational area, the survey also offered information about their perception of noise at home. Most studies of the relationship between noise exposure and annoyance have examined one setting in isolation from other settings. But, people live their lives in different arenas that together form their everyday lives. Little is known about how the combined strains from noise in different everyday settings influence how well people cope with noise in any given setting. Both the old and the new main airport were situated close to both recreational and residential areas. At least some of the visitors to the outdoor recreational areas were likely to be affected by the air traffic, and the changes in air traffic, at home. It is reasonable to assume that the perception of aircraft noise in the recreational areas partly depends on how annoyed the recreationists are by aircraft noise and other noise sources in other settings of their daily lives. Recreation is in a sense restoration, and the recreational context could be seen as complementary to other contexts. Studies on people's motivation for visiting outdoor recreational areas indicate that sound is both part of what people seek relief from, and what they actively seek to experience in visiting outdoor recreational areas (Driver *et al.*, 1987; Vaagbø, 1993). The reasons that were judged as rather important or very important by the highest proportions of the respondents in the study by Vaagbø were: ''To experience the silence and peace of nature'' (87 percent), ''to get out in nature, away from noise and pollution'' (85 percent), and ''to get away from hustle and bustle'' (79 percent).

[a)]Portions of this work were presented at the Internoise 2003 Congress, Jeju Island, Korea.
[b)]Electronic mail: norun.krog@fhi.no
[c)]Current affiliation: BREKKE & STRAND akustikk as, Hovfaret 17, 0275 Oslo, Norway.

In this article the relationship between annoyance with aircraft noise in the recreational areas and annoyance with noise at home is first examined. Annoyance with noise at home is specified as annoyance with noise from aircraft, road traffic, and neighbor. It was hypothesized that people who were highly annoyed by noise at home would be more in need of relief from noise than others, and consequently less tolerant of aircraft noise in the outdoor recreational areas. Second, the article investigates if there is a difference in annoyance with aircraft noise in the recreational areas between visitors who have their home address within the noise zones of an airport and visitors who are less exposed to aircraft noise at home. To collect detailed information about the visitors' exposure to different noise sources at home was not possible within the scope of the project. But, address information given in the survey questionnaire made this simple classification of home addresses possible. It was hypothesized that people who were living within the noise zones of an airport would be more in need of relief from aircraft noise than others, and consequently less tolerant of aircraft noise in the outdoor recreational areas.

In addition to the noise situation at home, other characteristics of the visitors and the recreational context itself may influence reactions to aircraft noise in the recreational areas. By ''recreational context'' we mean the situation of being in an outdoor recreational area, as opposed to for example being at home, or at work. Kariel (1990) found that backcountry hikers experienced technical sounds considerably more annoying than highway-oriented campers. Anderson *et al.* (1993) found people on short hikes to be more annoyed by aircraft noise than people at overlooks. First-time visitors were found to be less annoyed by the same aircraft noise exposure than people who had visited the area before (Anderson *et al.*, 1993). Tarrant *et al.* (1995) found attitudes and recreation motives to be strongly related to visitor evaluations of aircraft overflights in wilderness areas. In the study of the effects of military jets on hikers in a mountain area, Krog *et al.* (2000) found the visitor reaction to be influenced by the duration of the hike.

Dose–response relationships are used not only to describe an environmental situation, but also to predict the effect of changes in exposure. The common assumption that estimates on the grounds of a stable state situation could be used in the prediction of effects of change has been criticized (Flindell and Porter, 2000; Guski, 2000). Some studies from residential areas indicate an ''overreaction'' to changes in noise exposure, compared to the reaction that is predicted on the basis of dose–response relationships from a steady-state situation (Langdon and Griffiths, 1982; Brown *et al.*, 1985; Horonjeff and Robert, 1997; Fields *et al.*, 2000; Fidell *et al.*, 2002; Fidell and Pearsons, 2003). Whether or not the overreaction effect occurs outside a daily exposure context cannot be inferred from these studies.

The purpose of the analysis of another article on the same study (Krog and Engdahl, 2004) was to examine how characteristics of the recreational situation affected annoyance with aircraft noise during the visit, and, especially, how the dose–response relationship was influenced by an abrupt change in noise levels. The other context variables that were examined were reasons for participation in outdoor recreation, the duration of the visit, and whether the area was the recreationist's primary recreational area or not. The results showed a strong effect of the situation of change itself, beyond what was expected on the grounds of the actual changes in noise exposure levels. In addition, an effect of the reasons for participation in outdoor recreation and the duration of the visit was found in both study areas.

Although reviews of studies from the residential setting have found little effect of demographic variables on noise annoyance (Job, 1988; Fields, 1993; Miedema and Vos, 1999), demographic variables possibly may influence the relationship between noise annoyance in the recreational setting and at home. The visitor characteristics age (Anderson *et al.*, 1993; Vorkinn, 1999; Krog *et al.*, 2000) and gender (Anderson *et al.*, 1993; Fleming *et al.*, 1998; Vorkinn, 1999) have been found to influence annoyance in an outdoor recreational setting. Another personal characteristic that may influence noise annoyance both in the recreational area and at home is noise sensitivity. In his meta analysis of factors influencing annoyance, Fields (1993) found evidence in studies from residential areas that noise annoyance is related to noise sensitivity. In a preliminary analysis of some of the data from the study that is presented in this article (Krog and Engdahl, 1999), noise sensitivity was indicated to influence noise annoyance in the recreational setting.

Both the previous article on the same study (Krog and Engdahl, 2004) and this article focus on how contextual variables influence noise annoyance in outdoor recreational areas. The unique contribution of this article is to expand on this contextual analysis of noise annoyance. That is, the possible relationship between the experience of noise in two different contexts is explored. We investigate how annoyance in the recreational area may not only be influenced by characteristics of the recreational situation itself, but also how the annoyance with this short-term exposure may be influenced by the long-term noise situation at home.

In addition to variables describing either noise annoyance or exposure at home, demographic variables are added to the analyses in this article. Demographic variables were added because it was assumed that age, gender, and occupational status might affect both exposure at home through time spent at home, recreational needs, and thus reactions to noise. Finally, noise sensitivity was added to the analyses, to examine whether a relationship between noise annoyance in the recreational area and at home could be explained by noise sensitivity influencing the reaction in both contexts.

II. MATERIALS AND METHODS

A. Site selection

The outdoor recreational areas selected for the study were Bygdøy near the old main airport, and Romeriksåsen near the new main airport. Since comparison between the areas was not a main issue, it was not regarded important that the areas be totally comparable in all aspects. The main point was that they were areas experiencing opposite changes in aircraft noise exposure. The areas were selected because of their location relative to the airports, and because they are

TABLE I. Median aircraft noise exposure, number of respondents, and response rates for each area and data collection period.

	Bygdøy		Romeriksåsen	
	1998	1999	1998	1999
Median exposure, $L_{\mathrm{Aeq,air}}$	69 dB A	47 dB A	44 dB A	45 dB A
Single respondents	303	268	88	207
Respondents in groups: 2	428	484	166	348
3–4	200	146	37	105
5–9	34	67	5	45
Total number of respondents	965	965	296	705
Response rate	62%	52%	88%	77%

much used by the local communities for hiking and related activities. Bygdøy is an area of about 2.6 km^2, containing both a popular beach and a small forest. Romeriksåsen is a larger forest area of about 7600 km^2. There are several small lakes in the area, and a network of forest roads and paths. The roads are generally not open for car traffic. There are also some private and public cabins.

Aircraft sound was dominated by sound from jet aircraft in both areas. There was also occasional traffic with smaller aircraft and helicopters.

B. Procedure

Noise measurements with simultaneous collection of survey data were carried out at one site at Bygdøy and three different sites in Romeriksåsen. The questionnaire was tested in the field prior to the data collection. Field studies at Bygdøy were conducted during 6 weekend days in the period from the 18th of April to the 10th of May, 1998 (t_1), and again during 7 weekend days in the period from the 17th of April to the 2nd of May, 1999 (t_2). The first field study in Romeriksåsen was carried out during 5 weekend days from the 5th to the 26th of September, 1998 (t_1), and the second during 10 weekend days in the period from the 21st of August to the 26th of September, 1999 (t_2). The seasonal difference (spring/autumn) between the study periods in the two areas was not regarded as important, since the temperatures and weather conditions were quite similar for the chosen study periods. We chose times of the year for the surveys with relatively homogeneous activity within each area, and between the areas. The survey periods were chosen to be in the peak season in each area for hiking and related activities. The bathing season at Bygdøy (approximately from June to August) was deliberately avoided, since bathing is an activity that is likely to be less sensitive to noise than hiking in the forest. Table I summarizes some information regarding the subsamples: the median noise exposure during the visits to the area, the number of respondents and group sizes, and the response rates. A number of respondents were not included in the analyses because of missing values on one or more of the analyzed variables (Tables II, III, and V). The item nonresponse rates for the analyzed variables varied between zero and 5 percent. The highest nonresponse rates were for gender and the index variables.

The survey data were collected in the period from about 12 a.m. to 5 p.m. each day. Sometimes the data collection lasted a bit longer, sometimes shorter. The variation was due to weather conditions and the amount of visitors on the various study days. Data could not be collected in rainy weather, for several reasons. First, the sound measurement equipment could not be used when it was raining. In addition, people tended to disappear from the areas very quickly as soon as it started to rain. It would have been difficult getting people to sit down and fill out a questionnaire, and the questionnaires would have become wet without a shelter.

The survey data were collected at the trail close to a parking lot, or at parking lots. The study was masked as a general study about outdoor recreational areas in the region. Whether alone or in a group, *all* hikers aged 18 or older were asked to fill out a questionnaire about today's visit while on their way out of the area. An additional inclusion criterion was that the respondents should have stayed in the area for at least 15 min. Six of the respondents in the Bygdøy sample and one in the Romeriksåsen sample were excluded from the analysis because it appeared that they had stayed less than 15 min in the area. Subjects belonging to groups were assigned a specific group number to identify their group membership.

Only people who had not participated in the study before the change were invited to participate in the second survey. Otherwise, the conditions for answering the questionnaire would not be equal, since the participants of the first survey would already be familiar with the study and its purpose. Self-completion of questionnaires was chosen to avoid potential respondents getting to know the content of the questionnaire before answering it. This could have been a problem with interviews at the site. The respondents were not allowed to talk to each other while completing the questionnaire.

Acoustic data were collected at one location only for each study site at a place hidden from the visitors, which is three different locations in Romeriksåsen and one at Bygdøy. The locations were at positions central to the walking area and away from local sound sources like running water, roads, or parking lots.

An observer was responsible for categorizing sound sources as either aircraft, human, or natural. Sound source identification was based on perceived dominant source. When the respondents completed the questionnaire as they left the area, they attempted to recall the hour and minute at which they had entered the area. The field staff recorded the time when the questionnaire was handed in. Individual sound-exposure doses were estimated by combining the sound-exposure levels from each aircraft sound event with

the time of visit of each respondent into A-weighted equivalent aircraft sound exposure levels ($L_{\mathrm{Aeq,air}}$).

The accuracy of the noise dose has not been estimated, but it is believed to be similar in the two areas. In Romeriksåsen the visitors covered a large area (the 95%-confidence interval of reported walking distance was 2–15 km), but the air traffic was dominated by high-altitude overflights that makes the sound levels even. At the smaller area of Bygdøy the visitors were never further away from the recording location than about 1 km, but, being close to the airport the flights were mainly low altitude with larger area level deviations.

The acoustical measurements are described in more detail elsewhere (Engdahl, 2001).

C. Analyses

The data from the two study areas were analyzed separately. The relationship between annoyance and aircraft noise exposure in the two area samples was illustrated in a previous article on the same study by two tables showing the percentages choosing each annoyance response in 5-decibel noise groups (Krog and Engdahl, 2004). Multivariate linear regression was chosen for the analyses that are presented in this article. Linear regression has the advantage over logistic regression analysis that the information contained in the total measurement scale is kept. Because the dependent variable was a category scale, and not a continuous variable, the suitability of linear models was tested by comparing the results to the results of multivariate monotonic regression models. The program GOLDMineR from SPSS was used in these analyses. The monotonic regression models mostly did not prove significantly better, nor did they give a different result than the linear regression models. The improvement in the explained variance was in no cases more than a few percent. Since the linear regression analysis also has the advantage of offering standardized coefficients, which facilitates the direct comparison of effect sizes, the linear models were chosen.

Since *all* visitors were invited to participate, whether they were alone or in a group, the data could not be expected to be totally independent, however. Although the questionnaires were filled out individually, people in a group shared the same exposure and might share other experiences and attitudes as well. This means that the standard errors may be underestimated in an ordinary multivariate linear regression analysis, and thus that the conclusions regarding the significance of the results may be wrong. To test for the possible dependence in the data, linear mixed-effect models were tested that account for intraclass correlation among subjects belonging to the same group. The data were analyzed by linear mixed models including random effects using the procedure MIXED in SPSS 11.0. In a linear mixed model, the outcome, noise annoyance, at the individual level is modeled in terms of both individual and cluster level variables, while concurrently estimating and adjusting for the amount of intraclass correlation present in the data. The term ''fixed effects'' is used to denote factors with levels that are deliberately arranged, rather than randomly sampled from an infinite population of possible levels. Factors with randomly sampled levels are called random effects. The effects of noise exposure, the contextual and the demographic variables were modeled as fixed effects, while group was introduced as a random effect in the model.

The validity of the mixed model analysis of the Bygdøy data is uncertain, because the iteration terminated without reaching convergence. With this caution, we use the model as our best estimate of how group membership influenced the results. The results from the mixed-model analyses indicated that there was a significant random effect of group membership in both datasets ($p<0.001$).[1] A comparison of the results from the mixed-model analysis with the results from the multivariate linear regression analysis indicated, however, that the estimates and the sampling errors were only marginally different in the two analyses. The significance of the other variables in the model was not altered by accounting for group membership. Since the more complex model, the mixed model, did not alter the main results, the simpler multivariate linear regression model is used to represent the results.

D. Variables

1. Analyses of the relationship between noise annoyance in the recreational area and at home

The first analyses examined the relationship between annoyance with aircraft noise in the recreational areas and annoyance with noise at home. Variables describing the recreationists' individual aircraft noise exposure during the visit, the recreational context, and demographic background were also included in the model.

The dependent variable in the analyses was ''aircraft noise annoyance'' during the present visit to the area. The question about annoyance with ''the sound from aircraft'' was embedded in a series of questions about annoyance with possible adverse conditions in the area. The wording of the question was: ''If you still think of today's visit to Bygdøy/Romeriksåsen. Have you been annoyed by any of the following conditions during your visit?'' Annoyance was measured on a four-category scale: ''very annoyed,'' ''rather annoyed,'' ''slightly annoyed,'' or ''not annoyed.'' The numerical coding for this and the other categorical and dichotomous variables used in the analyses are given in Tables II, III, and V.

The same four-category scale was used for the variables measuring noise annoyance at home. The wording of the question was: ''To what degree are you annoyed by noise at home?'' Annoyance at home was further specified as ''noise from neighbor,'' ''aircraft noise,'' and ''road traffic noise'' in the questionnaire, and was put into the analyses as three separate variables.

A-weighted equivalent aircraft sound levels (''$L_{\mathrm{Aeq,air}}$'') was used to describe the recreationists' noise exposure during the visit.

Variables describing the recreational context were the ''situation,'' the ''duration of the visit,'' if the area was the visitor's ''primary recreational area'' or not, and two additive indices describing reasons for participation in outdoor recreation in general, related to experiential qualities of the area.

The situation variable was a dichotomous variable indicating in which situation, before or after the airport change,

the response was given. The variable was included to test whether the studies from before and after the airport change showed different responses to the *same* exposure levels.

The duration of the visit was measured in minutes. The variable was based on the respondents' recall of the hour and minute at which they had entered the area, and the recordings of the field staff of the time when the questionnaire was handed in. The respondents were asked to note ''the date and as accurately as possible the time'' when they came to the outdoor recreational area at Bygdøy/Romeriksåsen. They were given an example, to explain the level of accuracy that was desired: ''Example: If you know that you started the hike at 11.05, or 11.17, then write that, do not put it in round figures.'' The respondents were allowed to discuss this question with others in order to get the information as accurate as possible.

If the area was the visitor's primary recreational area or not was based on answers to the question: ''If you think of the last twelve months, is Bygdøy/this part of Romeriksåsen the outdoor recreational area that you have visited most frequently, or did you visit other areas more frequently?'' The response categories were ''most frequently Bygdøy/this part of Romeriksåsen,'' ''most frequently other areas,'' or ''not sure.'' In the analyses the category ''not sure'' was combined with ''most frequently other areas.''

The variables composing the indices describing reasons for participation in outdoor recreation were adopted from Vaagbø (1993). The respondents were asked to indicate whether a series of possible reasons for participation were ''very important,'' ''rather important,'' or ''not very important'' to them. The variables were coded from 1 (''not very important'' to 3 (''very important''). Two indices were constructed. The items were combined that describe either visual or acoustic aspects of the nature experience as reasons for participation in outdoor recreational activities. The index ''reasons for participation: the visual nature experience'' was composed by adding together the values of the variables ''how important is the experience of birds and animals, flowers and trees'' and ''how important is the appreciation of nature.'' The index ''reasons for participation: the acoustic nature experience/mental relaxation'' combined the variables ''how important is the experience of the silence and peace of nature,'' ''how important is it to get away from noise and pollution,'' and ''how important is it to get away from hustle and bustle.'' To improve the interpretability of the index scores, the scores were transformed to standard deviation units.

The demographic variables included in the analyses were ''age,'' ''gender,'' and ''occupational status.'' Age was included as a continuous variable in the main analyses. But, to test for a possible nonlinear effect of age, two further age variables were tested. Age was categorized in three groups: 18–39 years, 40–66 years, and 67 years or older. Two dummy variables were generated and tested with the age group 40–66 years as the reference category. Further, a variable ''square of age'' was constructed and tested, for the same purpose.

''Occupational status'' was categorized in four groups in the questionnaire: ''working,'' ''student, pupil,'' ''living on Social Security, pensioner,'' or ''full-time housewife.'' A dichotomous variable was constructed by collapsing the categories ''living on Social Security, pensioner'' and ''full-time housewife'' into the category ''not occupied,'' and the categories ''working'' and ''student, pupil'' into the category ''occupied.'' The reason for combining the categories in this way was the assumption that people living on Social Security, pensioners, and full-time housewives are spending more time at home than students, pupils, and people working outside the home. This difference in time spent at home may represent a difference in the relevance of noise exposure at home in the two groups, as well as a difference in the recreational needs.

Noise sensitivity was included in an additional analysis to test the alternative explanation of a relationship between annoyance at home and annoyance in the recreational area that people who are sensitive to noise are more annoyed than others in both contexts. Noise sensitivity was measured by one question that asked the respondents to evaluate how easily they were disturbed by noise on a 10-point scale, where 1 was ''gets hardly disturbed,'' and 10 was ''gets very easily disturbed.'' The wording of the question was: ''Do you consider yourself a person who easily gets disturbed or irritated by sounds, or are you not easily disturbed or irritated? We ask you to rate yourself on a scale from 1 to 10, where 1 indicates that it takes very much before you get disturbed, and 10 that you are very easily disturbed.''

2. Analyses of the influence of aircraft noise exposure at home on annoyance in the recreational area

In addition to the measures of subjective experience of noise at home, a variable was constructed categorizing exposure to aircraft noise at home. On the basis of information given in the questionnaire about home address, the respondents were assigned to categories of aircraft noise exposure at home. The categories corresponded to the four noise zones around the airports relevant to the time of each study, plus a category for those living outside the noise zones. The noise zones are further described in Table IV.

For the Bygdøy sample from t_1, which was the only subsample that contained people living within noise zones, a multivariate linear regression analysis was conducted, where the exposure variable noise zone replaced the variables describing annoyance with noise at home. Since data from only one point in time were analyzed, the variable ''situation'' was not included in this analysis. Otherwise, the model was the same as in the first two analyses.

III. RESULTS

The results of the multivariate linear regression analyses containing variables describing noise annoyance at home are shown in Tables II (Bygdøy) and III (Romeriksåsen).

A. The relationship between annoyance in the recreational area and at home: Results from Bygdøy

Table II indicates that there was a significant relationship between degree of annoyance with aircraft noise during

TABLE II. Aircraft noise annoyance [four categories: 1 (not annoyed)–4 (very annoyed)] dependent on $L_{Aeq,air}$ at today's visit to the area, context variables, noise annoyance at home, and demographic variables. Bygdøy t_1 and t_2 ($n=1658$).

Bygdøy ($R^2=0.572$)	Unit[a]	Unst. B[b]	SE[c]	Beta[d]	SE[e]	t	p[f]
Constant		2.958	0.281			10.52	
Situation	0-before/ 1-after	−1.357	0.065	−0.596	0.033	−20.87	i
$L_{Aeq,air}$	dB A	0.012	0.003	0.112	0.033	3.99	i
Reasons for participation: the visual nature experience	SD[j]	0.032	0.022	0.028	0.036	1.42	
Reasons for participation: the acoustic nature experience/mental relaxation	SD	0.073	0.022	0.063	0.036	3.32	i
Duration of the visit	Minutes	0.001	0.000	0.050	0.031	3.08	h
Primary recreational area	0-no/1-yes	−0.059	0.041	−0.024	0.032	−1.45	
Annoyed by aircraft noise at home	4 categories, 1 (not)– 4 (very)	0.207	0.026	0.143	0.034	8.07	i
Annoyed by road traffic noise at home	4 categories, 1 (not)– 4 (very)	0.062	0.022	0.048	0.033	2.75	h
Annoyed by noise from neighbor	4 categories 1 (not)– 4 (very)	−0.050	0.028	−0.031	0.033	−1.79	
Age	Year	0.001	0.002	0.014	0.037	0.69	
Gender	0-male/ 1-female	−0.085	0.038	−0.037	0.032	−2.23	g
Occupational status	0-not occ./ 1-occ.	−0.107	0.064	−0.032	0.037	−1.68	

[a]Variable unit.
[b]Unstandardized beta.
[c]Standard error of the unstandardized beta.
[d]Standardized beta.
[e]Standard error of the standardized beta.
[f]Multivariate linear regression analysis.
[g]$p<0.05$.
[h]$p<0.01$.
[i]$p<0.001$.
[j]Standard deviation units.

the visit to the recreational area at Bygdøy, near the old main airport, and degree of annoyance with aircraft noise and road traffic noise at the visitor's home. Visitors who were annoyed by aircraft noise or road traffic noise at home tended to be more annoyed by aircraft noise in the recreational area than visitors who were less annoyed at home. Comparing the size of the betas, annoyance with aircraft noise during recreation was, however, not indicated to be as strongly related to annoyance with road traffic noise at home as to annoyance with the same source at home. There was no significant relationship between annoyance with aircraft noise during the visit to the recreational area and annoyance with noise from neighbors.

Annoyance with aircraft noise at home was indicated to be the second most important variable in the model, more important than the noise exposure during the visit, measured in $L_{Aeq,air}$. The variable that was indicated to most strongly affect annoyance during the visit was the ''situation'' in which the response was given, before or after the airport change. Given the same noise levels, people tended to be less annoyed after the old airport was closed down than before the moving of the airport. Other context variables with a significant positive effect on annoyance were ''reasons for participation: the acoustic nature experience/mental relaxation,'' and the duration of the visit.

There were no significant effects of the demographic variables, except a weak effect of gender: Women tended to be less annoyed than men. In reanalyzing data from residential studies, Miedema and Vos (1999) found a curvilinear effect of age, people in the youngest and the oldest age groups being more annoyed than people in the middle age categories. An analysis with dummy variables for age groups was conducted, to test if there was a nonlinear effect of age in the data of this study, but no such effect was found, and the linear variable was kept. Another possible reason why an age effect was not found may be that both noise annoyance during recreation and at home were in the same way influenced by age. To test for this possibility, $L_{Aeq,air}$ and age, and then square of age, was entered in a first step of the linear regression analysis, before including the variables concerning noise annoyance at home. However, no significant effect of age or square of age was found.

The complete model explained 57% of the variance in the data.

B. The relationship between annoyance in the recreational area and at home: Results from Romeriksåsen

In the results from Romeriksåsen, near the new main airport (Table III), the effect of annoyance with aircraft noise

TABLE III. Aircraft noise annoyance [four categories: 1 (not annoyed)–4 (very annoyed)] dependent on $L_{\mathrm{Aeq,air}}$ at today's visit to the area, context variables, noise annoyance at home, and demographic variables. Romeriksåsen t_1 and t_2 ($n=849$).

Romeriksåsen ($R^2=0.200$)	Unit[a]	Unst. B[b]	SE[c]	Beta[d]	SE[e]	t	p[f]
Constant		−0.553	0.377			−1.47	
Situation	0-before/ 1-after	0.348	0.066	0.176	0.028	5.30	i
$L_{\mathrm{Aeq,air}}$	dB A	0.017	0.008	0.073	0.028	2.23	g
Reasons for participation: the visual nature experience	SD[j]	0.016	0.033	0.017	0.019	0.48	
Reasons for participation: the acoustic nature experience/mental relaxation	SD	0.064	0.031	0.072	0.019	2.04	g
Duration of the visit	Minutes	0.001	0.000	0.098	0.016	3.10	h
Primary recreational area	0-no/1-yes	0.044	0.062	0.022	0.017	0.70	
Annoyed by aircraft noise at home	4 categories, 1 (not)–4 (very)	0.370	0.038	0.329	0.018	9.83	i
Annoyed by road traffic noise at home	4 categories, 1 (not)–4 (very)	−0.008	0.035	−0.008	0.017	−0.23	
Annoyed by noise from neighbor	4 categories 1 (not)–4 (very)	0.031	0.047	0.021	0.017	0.65	
Age	Year	−0.001	0.002	−0.016	0.021	−0.46	
Gender	0-male/ 1-female	−0.228	0.057	−0.128	0.016	−4.00	i
Occupational status	0-not occ./ 1-occ.	0.205	0.088	0.083	0.020	2.34	g

[a]Variable unit.
[b]Unstandardized beta.
[c]Standard error of the unstandardized beta.
[d]Standardized beta.
[e]Standard error of the standardized beta.
[f]Multivariate linear regression analysis.
[g]$p<0.05$.
[h]$p<0.01$.
[i]$p<0.001$.
[j]Standard deviation units.

at home was indicated to be the strongest effect in the model, comparing the relative size of the betas. No significant effect was found of annoyance with road traffic noise at home, or noise from neighbors. $L_{\mathrm{Aeq,air}}$ was indicated to be the variable that affected the level of annoyance during recreation least among the significant variables in the model.

The situation was indicated to be the second most influential variable in the results from Romeriksåsen. Context variables beside the situation that significantly influenced annoyance were the index variable ''reasons for participation: the acoustic nature experience/mental relaxation'' and the duration of the visit. People for whom the acoustic nature experience was important tended to be more annoyed by aircraft noise than those for whom this was less important. People tended to be more annoyed on visits of longer duration.

Two demographic variables significantly influenced annoyance. The results indicated that there were a relatively strong effect of gender and a weaker effect of occupational status. As was found at Bygdøy, women tended to be less annoyed than men. People with an occupation tended to be more annoyed by aircraft noise while recreating in Romeriksåsen than people who did not have an occupation. The same additional analysis was done regarding the age variable as for the Bygdøy sample, but no significant effect was found.

The effect sizes (betas) were estimated to be somewhat different in the two datasets. However, basing the test on the standard errors of the beta values, only the estimated effects of the following variables were significantly different between the two samples: annoyance with aircraft noise at home ($p<0.0001$), gender ($p<0.05$), and occupational status ($p<0.01$). The difference between the betas of the situation variable was not tested, since the variable represents a different frame of reference in the two areas. At Bygdøy the variable represents the contextual effect of a decrease in noise levels, while the context is an increase in noise levels in Romeriksåsen.

The model explained 21% of the variance in the data.

C. The effect of noise sensitivity

It was hypothesized that people who are annoyed by noise at home may be more easily annoyed than others in the recreational areas, because they are more in need of rest from a noisy environment. An alternative explanation of the relationship between annoyance in the recreational area and annoyance at home is that people who are sensitive to noise are more annoyed than others in both contexts. The alternative explanation was tested by adding the noise sensitivity variable to the multivariate regression model in Tables II and III.

The inclusion of noise sensitivity meant little to the effects of noise annoyance at home. A significant effect of

TABLE IV. Aircraft noise exposure at home, Bygdøy t_1. Noise zones, EFN (a Norwegian aircraft sound metric based on 24-hour time-averaged sound-exposure level with a 10-dB nighttime weighting from 0 to 10 dB on shoulder hours, and a 3-dB penalty on Sundays).

Noise zone (EFN)	Frequency	Percent
<55 dB	726	75
Zone I (55–60 dB)	153	16
Zone II (60–65 dB)	37	4
Zone III (65–70 dB)	25	3
Zone IV (>70 dB)	0	0
Missing	24	2
Total	965	100

noise sensitivity was found in both areas (Bygdøy: beta $=0.086$, SE $=0.009$, $t=5.15$; Romeriksåsen: beta $=0.095$, SE $=0.014$, $t=2.97$). But, in both cases the effect of noise sensitivity was indicated to be smaller than the effect of being annoyed by aircraft noise at home. Adjusted R^2 just increased from 0.572 to 0.579 (Bygdøy), and from 0.200 to 0.204 (Romeriksåsen). There were only minor changes in the effects of being annoyed at home by aircraft noise, and, at Bygdøy, road traffic noise. (The beta for annoyance with aircraft noise at home was altered from 0.143 to 0.140 in the Bygdøy model, and from 0.329 to 0.313 in the Romeriksåsen results. The effect of annoyance with road traffic noise at home was altered from beta $=0.048$ to beta $=0.039$ in the Bygdøy results.) The small negative effect of annoyance with noise from neighbor in the Bygdøy data was slightly increased to be significant in the analysis with noise sensitivity (beta $=-0.041$, SE $=0.028$, $t=-2.35$).

D. The relationship between annoyance in the recreational area and exposure to aircraft noise at home

Table IV shows the distribution of home addresses relative to the old main airport in the sample from Bygdøy at t_1. The majority of the sample was living outside the noise zones. Smaller proportions of the respondents belonged in the noise zones 1 through 3, the largest number of these in the outermost noise zone.

The results of the analysis including noise zone are presented in Table V. No significant effect of noise zone at home was found. The model explained 5 percent of the variance in the data.

IV. DISCUSSION

A. Summary of results

Annoyance with aircraft noise in the recreational areas was related to annoyance with noise at home, but the relationship was not independent of the noise source. A relationship between annoyance with aircraft noise in the recreational area and annoyance with aircraft noise at home was found in both datasets. Comparing the size of the betas, the effect was larger in Romeriksåsen, at the new airport, than at Bygdøy, at the old airport. No consistent effect was found of annoyance with noise from neighbors. At Bygdøy there was a smaller significant effect of annoyance with road traffic noise at home. No significant effect of aircraft noise exposure at home was found.

Beside $L_{\mathrm{Aeq,air}}$ for the visit, annoyance was also influenced by the situation (before/after the change), ''reasons for

TABLE V. Aircraft noise annoyance [four categories: 1 (not annoyed)–4 (very annoyed)] dependent on $L_{\mathrm{Aeq,air}}$ at today's visit to the area, context and demographic variables, and noise zone at home. Bygdøy, t_1 ($n=817$).

Bygdøy t_1 ($R^2=0.05$)	Unit[a]	Unst. B[b]	SE[c]	Beta[d]	SE[e]	t	p[f]
Constant		2.058	0.422			4.88	
$L_{\mathrm{Aeq,air}}$	dB A	0.007	0.005	0.047	0.034	1.39	
Reasons for participation: the visual nature experience	SD[i]	0.055	0.043	0.054	0.042	1.26	
Reasons for participation: the acoustic nature experience/mental relaxation	SD	0.144	0.042	0.142	0.041	3.46	h
Duration of the visit	Minutes	0.003	0.001	0.115	0.038	3.33	h
Primary recreational area	0-no/1-yes	−0.065	0.081	−0.030	0.037	−0.81	
Year of birth	Year	0.003	0.003	0.046	0.046	1.08	
Gender	0-male/ 1-female	−0.159	0.072	−0.078	0.035	−2.19	g
Occupational status	0-not occ./ 1-occ.	−0.103	0.118	−0.036	0.041	−0.88	
Noise zone at home	5 categories, 0 (<55 dB) 4 (>70 dB)	0.046	0.055	0.030	0.020	0.83	

[a]Variable unit.
[b]Unstandardized beta.
[c]Standard error of the unstandardized beta.
[d]Standardized beta.
[e]Standard error of the standardized beta.
[f]Multivariate linear regression analysis.
[g]$p<0.05$.
[h]$p<0.001$.
[i]Standard deviation units.

participation: the acoustic nature experience/mental relaxation,'' the duration of the visit, gender, and noise sensitivity in both datasets. In addition, an effect was found of occupational status in the data from Romeriksåsen. Except for annoyance with aircraft noise at home, gender, and occupational status, the betas were not significantly different between the two samples.

B. The effect of noise annoyance at home

If the relationship between annoyance in the recreational area and noise annoyance at home was just a methodological scaling artifact, one would expect a similar result for the different same-scale variables, which was not found. Thus, the relationship between noise annoyance in the recreational area and noise annoyance at home seems to be more than just the tendency to answer similar questions in a similar manner.

It was hypothesized that people who are annoyed by noise at home may be more easily annoyed than others in the recreational areas, because they are more in need of rest from a noisy environment. An alternative explanation of the relationship between annoyance at home and annoyance in the recreational area is that people who are sensitive to noise are more annoyed than others in both contexts. The results did not support the hypothesis that the relationship between noise annoyance in the recreational area and noise annoyance at home can be explained by underlying noise sensitivity.

The analyses combine data from before and after a change in noise exposure levels. The larger impact of annoyance with aircraft noise annoyance at home on the visitors' tolerance in Romeriksåsen than at Bygdøy may have to do with the direction of change and the following deterioration in environmental qualities both at home and in the recreational area. That is, the effect of aircraft noise annoyance at home is assumed to depend on the situation (before/after the change). The assumption is tested by conducting one additional multivariate regression analysis for each area sample, adding the interaction term of the situation and annoyance with aircraft noise at home. Regarding Bygdøy, the interaction term of the situation and annoyance with road traffic noise at home is also tested. All interactions tested are significant ($p<0.05$). At Bygdøy the effects of being annoyed by aircraft noise and road traffic noise at home is weaker after the airport was closed down than before the change. In Romeriksåsen the effect of being annoyed by aircraft noise at home is larger after the change than before the change. The effect of the interaction is stronger in Romeriksåsen than at Bygdøy.

C. The effect of noise exposure at home

The number of respondents in this study living within the noise zones was limited, and may be the reason why a significant effect of the variable noise zone at home could not be demonstrated. The variation in noise exposure at home was also probably too small to find effects. The indicated effect of noise annoyance at home warrants more research on this topic including more detailed data on noise exposure at home.

D. The effect of variables describing the recreational context

The effect of the context variables was the issue of another article on the same study, and was thoroughly discussed there (Krog and Engdahl, 2004). Here, we will comment mainly on the deviations between the results in this article and the previous analyses.

Regarding the results from Bygdøy, a small significant effect of the variable ''reasons for participation: the visual nature experience'' found in the earlier analysis was not significant in the extended analysis. The effect of the ''situation'' of change was somewhat diminished in the extended analysis. The effects of the other significant variables in the model were about the same or only slightly changed by the inclusion of the new variables.

Regarding Romeriksåsen, a significant effect of the variable ''primary recreational area'' was found in the analysis of the previous article. This effect was diminished and not significant in the extended analysis. In addition, the effects of the variables ''situation'' and ''reasons for participation: the acoustic nature experience/mental relaxation'' were diminished. The effects of the variables $L_{\mathrm{Aeq,air}}$ and the duration of the visit were about the same as in the previous article.

The diminished effects indicate that some of the effect of these variables found in the previous analyses was related to the variables included in the analyses of this article. It may, for instance, be that the effect of ''reasons for participation: the acoustic nature experience/mental relaxation'' and the effect of the area being the visitor's ''primary recreational area'' were related both to noise annoyance at home and during recreation. The effect of the ''situation'' variable indicates an ''overreaction,'' or a relief/grief reaction, to the change in noise exposure levels in the recreational areas. That is, the visitors tended to be less annoyed at the same noise levels at Bygdøy after the moving of the airport than before, and vice versa in Romeriksåsen. The size of the ''situation'' effect may depend on the initial noise levels, as well as the character and amount of change. It may also depend on the time past since the change at the time of the second survey. However, the effect indicates that data collected in a stable state situation cannot properly predict the effect of a change in noise levels. One would have to add the effect of the ''situation'' of experiencing the change itself. The reason for the diminished effect of the ''situation'' variable, found in both areas when new variables were included, may be that the moving of the airport influenced the noise exposure both in the recreational area and at the visitor's home. Although few respondents lived within the noise zones of the airports, the changes in the air traffic are noticeable in a much wider area around the airports. The relief/grief effect found in the recreational areas may have been increased by relief/grief because of changed noise exposure at home. This was adjusted for when noise annoyance at home was included in the model. But, this also means that the effect of noise annoyance at home found in this study may depend on the special changing noise environments in which the data were collected. The results should be interpreted with care, perhaps as an initial explorative analysis of how the experience of noise in a recreational situation may

be influenced by the visitors' experience of noise at home. Conclusions about the relationship in a stable state situation cannot be drawn. The data on noise exposure at home were also too sparse in the present study.

E. The effect of demographic variables

Studies that review large data materials from the residential setting have found demographic variables to have little influence on noise annoyance (Job, 1988; Fields, 1993; Miedema and Vos, 1999). The overall results of the analyses of data from the outdoor recreational setting in this article are in accordance with this finding. But, two demographic variables were found to influence annoyance.

Women were found to be less annoyed by aircraft noise than men in both local outdoor recreational areas studied. This is the opposite of the gender effect that has been hypothesized for the residential setting (Fields, 1993). Neither in the meta analysis of residential data by Fields (1993), nor in the secondary analysis of original field data sets by Miedema and Vos (1999) was a gender difference demonstrated, however. The gender effect found in the present study points to the importance of not generalizing from one type of setting to another. The recreational setting offers a different set of conditions and circumstances for reactions to noise than the residential setting.

In a Norwegian survey (Vorkinn, 1999) men found it more important than women to avoid noise in the most untouched natural area type, while women tended to find it slightly more important than men in some other recreational area types. The questions did not, however, presuppose actual experience, but were attitudinal questions posed at a general population sample. A moderate gender effect was found in the initial exploratory analyses conducted by Anderson *et al.* (1993) of their data from National Parks in the U.S. It is, however, not stated what the effect was. Fleming *et al.* (1998) found a significant gender effect similar to the one in this study for one of their study sites in Bryce Canyon National Park, but not for the other. The nonsignificant effect at the other site was, however, in the same direction. No gender effect was found in the Krog *et al.* study of effects of noise from jet fighters in a mountain area in Norway (Krog *et al.*, 2000).

An explanation of a gender difference may be that outdoor recreation has a different meaning for men than women. The significant positive effect of having an occupation found in Romeriksåsen may be interpreted in a similar manner as the gender effect. People with an occupation may have a different need for quiet relaxation than others, a need that is related to their job situation. The effect of occupation was, however, the opposite at Bygdøy (although not significant), and no firm conclusions can be drawn regarding this variable.

Contradictory to our findings, an effect of age was found in the studies by Anderson *et al.* (1993), Vorkinn (1999), and Krog *et al.* (2000). In Anderson's study, age was not included in any but the initial analyses, and the direction of the moderate effect found was not specified. In the survey by Vorkinn (1999) no age difference was found regarding attitudes to noise in the untouched area types, but regarding other area types there was a weak tendency that the wish to avoid noise increased with increasing age. In the Krog *et al.* (2000) study of noise from jet fighters a relatively strong effect of age was found, in the same direction.

V. CONCLUDING REMARKS

Noise research has mainly been focused on only one setting of the everyday life of a noise-exposed population at a time, which has been the residential setting. The results of this study indicate that annoyance with noise in one setting may be related to the experience of noise in another setting of the local environment. While the residential setting can be seen as the basic arena in the local environment, the recreational areas are arenas for restoration from the strains of everyday life on other arenas. The results indicate that people who were annoyed by aircraft noise at home tended to be more annoyed by this same noise source in outdoor recreational areas than were other recreationists. The analysis in this article only contained a rough estimate of aircraft noise exposure at the recreationists' homes, that is, categorization in noise zones. The number of people in the samples living within any of the noise zones was low. The data did not contain information about exposure at home from sources other than aircraft. Further investigations are warranted regarding the relationship between noise exposure at home and requirements regarding the soundscapes of outdoor recreational areas.

ACKNOWLEDGMENTS

We would like to thank Eyolf Osmundsen for initiating the study and Kristian Tambs for his contributions to the planning of the study. We thank the members of the reference group of the study for useful discussions in the early phases: Birgitta Berglund, Idar Granøien, Ronny Klæboe, and Kåre Liasjø. The content of this article is, however, the responsibility of the authors alone. We would also like to express our gratitude to the recreationists who participated in the study, and the staff who conscientiously collected and coded the data. Finally, we would like to thank the anonymous reviewers for valuable suggestions in their reviews of this article. The study was supported by The Research Council of Norway and the Norwegian Air Traffic and Airport Management.

[1]The *p* value indicates the probability of obtaining the result given that the null hypothesis (no relationship between the variables) was true. The smaller the *p* value, the more likely it is that the null hypothesis is not true. Conventionally the null hypothesis is rejected if the *p* value is less than 0.05, which means that it is less than 5 percent change that the null hypothesis is true, given the obtained result.

Aasvang, G. M., and Engdahl, B. (**2004**). ''Subjective responses to aircraft noise in an outdoor recreational setting: A combined field and laboratory study,'' J. Sound Vib. **276,** 981–996.

Anderson, G. S., Horonjeff, R. D., Menge, C. W., Miller, N. P., Robert, W. E., Rossano, C., Sanchez, G., Baumgartner, R. M., and McDonald, C. (**1993**). ''Dose-response relationships derived from data collected at Grand Canyon, Haleakala and Hawaii Volcanoes National Parks,'' HMMH Report No. 290940.14; NPOA Report No. 93-6, National Park Service, Denver, CO.

Brown, A. L., Hall, A., and Kyle-Little, J. (**1985**). ''Response to a reduction in traffic noise exposure,'' J. Sound Vib. **98**, 235–246.

Cessford, G. (**1998**). ''Visitor satisfactions, impact perceptions, and attitudes toward management options on the Milford Track,'' Report No. 87, Department of Conservation, Wellington, New Zealand.

Downing, M., Hobbs, C., and Stusnick, E. (**1999**). ''Measurement of the natural soundscapes in south Florida national parks,'' in FICAN Symposium on the Preservation of Natural Quiet. 138th Meeting of the Acoustical Society of America. 3 November, 1999. Available online at ⟨http://www.fican.org/download/downing.pdf⟩.

Driver, B. L., Nash, R., and Haas, G. (**1987**). ''Wilderness benefits: A state-of-knowledge review,'' in *Proceedings, National Wilderness Research Conference: Issues, State of Knowledge, Future Directions*, USDA Forest Service General Technical Report INT-220, Fort Collins CO, pp. 294–319.

Engdahl, B. (**2001**). ''Flystøy i rekreasjonsområdet på Bygdøy og Romeriksåsen. Analyse av lydopptak i forbindelse med en feltundersøkelse om effekt av flystøy i rekreasjons- og friluftsområder'' (''Aircraft noise in the outdoor recreational areas at Bygdøy and in Romeriksåsen. Analysis of sound recordings from field studies of effects of aircraft noise in recreational areas''), Rapport 2001:2, Folkehelsa, Oslo, Norway.

FICAN Symposium on the Preservation of Natural Quiet. 138th Meeting of the Acoustical Society of America. 3 November, 1999, (**1999**). Proceedings available online at ⟨http://www.fican.org/pages/sympos.html⟩.

Fidell, S., and Pearsons, K. (**2003**). ''Sensitivity to prospective transportation noise exposure,'' Noise Control Eng. J. **51**, 106–113.

Fidell, S., Silvati, L., and Haboly, E. (**2002**). ''Social survey of community response to a step change in aircraft noise exposure,'' J. Acoust. Soc. Am. **111**, 200–209.

Fidell, S., Silvati, L., Howe, R., Pearsons, K. S., Tabachnick, B., Knopf, R. C., Gramann, J., and Buchanan, T. (**1996**). ''Effects of aircraft overflights on wilderness recreationists,'' J. Acoust. Soc. Am. **100**, 2909–2918.

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., Ehrlich, G. E., and Zador, P. (**2000**). ''Theory and Design Tools for Studies of Reactions to Abrupt Changes in Noise Exposure,'' NASA Contractor Report CR-2000-210280, NASA Langley Research Center, Hampton, VA.

Fleming, G. G., Roof, C. J., Rapoza, A. S., Read, D. R., Webster, J. C., Liebman, P. C., Valihura, P. J., Lewis, K. C., Schomer, P. D., Plante, J. A., and Draper, J. A. (**1998**). ''Development of Noise Dose/Visitor Response Relationships for the National Parks Overflight Rule: Bryce Canyon National Park Study,'' FAA-AEE-98-01/DOT-VNTSC-FAA-98-6, U.S. Department of Transportation, Federal Aviation Administration.

Flindell, I., and Porter, N. (**2000**). ''The implications of context-based assessment for noise management,'' in Proceedings Internoise 2000, Nice, France.

Guski, R. (**2000**). ''Aspects contributing to global annoyance judgments in field interviews,'' in Proceedings Internoise 2000, Nice, France.

Horonjeff, R. D., and Robert, W. E. (**1997**). ''Attitudinal responses to changes in noise exposure in residential communities,'' NASA Contractor Report 97-205813, NASA Langley Research Center, Hampton, VA.

International Recreational Noise Symposium (**1999**). ''Effects on man and on the environment. Queenstown, New Zealand, 1998,'' Noise Control Eng. J. **47**, 78–162.

Job, R. F. S. (**1988**). ''Community response to noise: A review of factors influencing the relationship between noise exposure and reaction,'' J. Acoust. Soc. Am. **83**, 991–1001.

Kariel, H. G. (**1990**). ''Factors affecting response to noise in outdoor recreational environments,'' Can. Geogr. Geogr. Can. **34**, 142–149.

Krause, B. (**1999**). ''Loss of natural soundscapes within the Americas,'' in FICAN Symposium on the Preservation of Natural Quiet. 138th Meeting of the Acoustical Society of America. 3 November, 1999. Available online at ⟨http://www.fican.org/download/krause.pdf⟩.

Krog, N. H., and Engdahl, B. (**1999**). ''Aircraft noise in recreational areas: Effects on visitors' experience and well-being,'' Noise Control Eng. J. **47**, 147–149.

Krog, N. H., and Engdahl, B. (**2004**). ''Annoyance with aircraft noise in local recreational areas, contingent on changes in exposure and other context variables,'' J. Acoust. Soc. Am. **116**, 323–333.

Krog, N. H., Aasvang, G. M., Osmundsen E., and Engdahl, B. (**2000**). ''Effects of noise from military jets on hikers in a recreational area'' in Proceedings Internoise 2000, Nice, France.

Langdon, F. J., and Griffiths, I. D. (**1982**). ''Subjective effects of traffic noise exposure, II: Comparisons of noise indices, response scales, and the effects of changes in noise levels,'' J. Sound Vib. **83**, 171–180.

McDonald, C. D., Baumgarten, R. M., and Iachan, R. (**1995**). ''Aircraft management studies: National Park Service Visitors Survey,'' HMMH Report No. 290940.12; NPOA Report No. 94-2, National Park Service, U.S. Department of the Interior.

Miedema, H. M., and Vos, E. M. (**1999**). ''Demographic and attitudinal factors that modify annoyance from transport noise,'' J. Acoust. Soc. Am. **105**, 3336–3344.

Miller, N. P., Anderson, G. S., Horonjeff, R. D., and Thompson, R. H., (**1999**). ''Mitigating the effects of military aircraft overflights on recreational users of parks,'' AFRL-HE-WP-TR-2000-0034, Lexington, MA, Harris, Miller, Miller and Hanson, Inc.

Sutton, S. T. (**1999**). ''Aircraft noise impacts: A case study in the glacier region of the west coast of New Zealand,'' Noise Control Eng. J. **47**, 87–90.

Tarrant, M. A., Haas, G. E., and Manfredo, M. J. (**1995**). ''Factors affecting visitor evaluations of aircraft overflights of wilderness areas,'' Soc. Natur. Resour. **8**, 351–360.

Vaagbø, O. (**1993**). ''Den norske turkulturen: Friluftslivets år 1993,'' (''The Norwegian tradition of outdoor life: The year of outdoor life 1993'') (FRIFO).

Vorkinn, M. (**1999**). ''Synspunkter på støy i friluftsområder'' (''Opinions on noise in outdoor recreational areas''), Særtrykk (offprint), Østlandsforskning, Lillehammer. Available from Østlandsforskning, Serviceboks, 2626 Lillehammer.

Evaluation of stage acoustics in Seoul Arts Center Concert Hall by measuring stage support

Jin Yong Jeon[a)]
School of Architectural Engineering, Hanyang University, Seoul 133-791, Korea

Michael Barron[b)]
Department of Architecture and Civil Engineering, University of Bath, Bath, Somerset BA2 7AY, United Kingdom

(Received 25 October 2002; revised 14 July 2004; accepted 10 October 2004)

Stage acoustics is an important characteristic for concert halls, both for the acoustic quality on stage and for the audience. However, relatively little research has been conducted into the question. This study was based on the investigation of an actual concert hall stage, that of the Seoul Arts Center Concert Hall in Korea. The stage acoustics was evaluated in the actual hall, and with two models: a 1:25 scale model and a computer model. The study was based on the stage support parameter ST1 proposed by Gade as a measure of support for individual performers [Acustica **65**, 193–203 (1989)]. The variation of support was measured on the empty stage of the actual hall and in the two models. The effect of musicians on stage, the effect of moving the orchestra, the effect of ceiling height and of stage-wall profile were also investigated. Conclusions are drawn both relating to the Seoul Concert Hall stage and stages in general. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1829258]

PACS numbers: 43.55.Br, 43.55.Gx, 43.55.Ka [MK] Pages: 232–239

I. INTRODUCTION

The acoustic conditions for performers are clearly an important aspect in the acoustic design of concert halls, yet there are many issues associated with stage acoustics remaining to be resolved. For the performer it is important to hear his/her own sound and to hear the sound of colleagues in order to maintain ensemble. The acoustic character of the stage surround may well also influence the blend of the orchestral sound for the audience.

When performing, the musician hears the direct sound from his or her instrument followed by a sound reflection off the floor. Later, the performer receives direct sound from surrounding performers, as well as reflections off their instruments, chairs, or the floor. Stronger reflections may come from stage walls, the ceiling, or suspended elements. This sequence of reflections is important for the performer to maintain the rhythm, intonation, balance, and timbre of their sound.

The complexity of the problem of acoustic conditions for performers was revealed in an early study by Barron in the Gulbenkian Great Hall, Lisbon.[1] In that hall many major changes in stage configuration were not considered significant by the musicians, though all performers liked a strong overhead reflection. Marshall, Gottlob, and Alrutz[2] found evidence that, for small groups, reflections should not be too early and that high frequencies were more important than low frequencies for ensemble playing. Meyer has made many investigations into the necessary conditions for ensemble playing; a recent reference[3] summarizes some of his findings. O'Keefe[4] has provided evidence of the need for good line of sight between performers.

A major advance was made in 1989 by Gade,[5,6] who conducted experiments with players in two anechoic chambers linked electronically. He subsequently extended his survey to behavior in real halls. Of several proposed measures relating to performing conditions, stage support (ST1) emerges as the most robust. This measure is concerned with support of the performer's own sound and assumes that sound returning to the performer delayed between 20 and 100 ms is beneficial. Measured values of ST1 have been found to correspond with musicians' assessments of ease of playing.[7] An interesting implementation on a real stage of Gade's principles has been reported by Rindel.[8]

Though considerable progress has been made on the question of conditions for performers during the last quarter century, many issues remain unresolved. In particular, much more could be known about the implications of the geometrical shape of the stage and its size. In this paper, it was decided to use Gade's proposed support measure ST1 to investigate an individual concert hall stage. The auditorium chosen was the concert hall in the Seoul Arts Center. Measurements were made in the real hall, in a 1:25 scale model of the stage region and with a computer simulation model.

II. THE SEOUL ARTS CENTER CONCERT HALL AND ITS MODELS

Seoul Arts Center (SAC) Concert Hall was completed in 1988, designed by Kim Suk Chul and Associates. For the acoustics of the hall, Rob Harris (Arup Acoustics, UK) carried out a review of the acoustical design at the scheme design stage. Jeff Charles (Bickerdike Allen Partners, UK) also

a)Author to whom correspondence should be addressed. Electronic mail: jyjeon@hanyang.ac.kr
b)Electronic mail: m.barron@bath.ac.uk

 0001-4966/2005/117(1)/232/8/$22.50

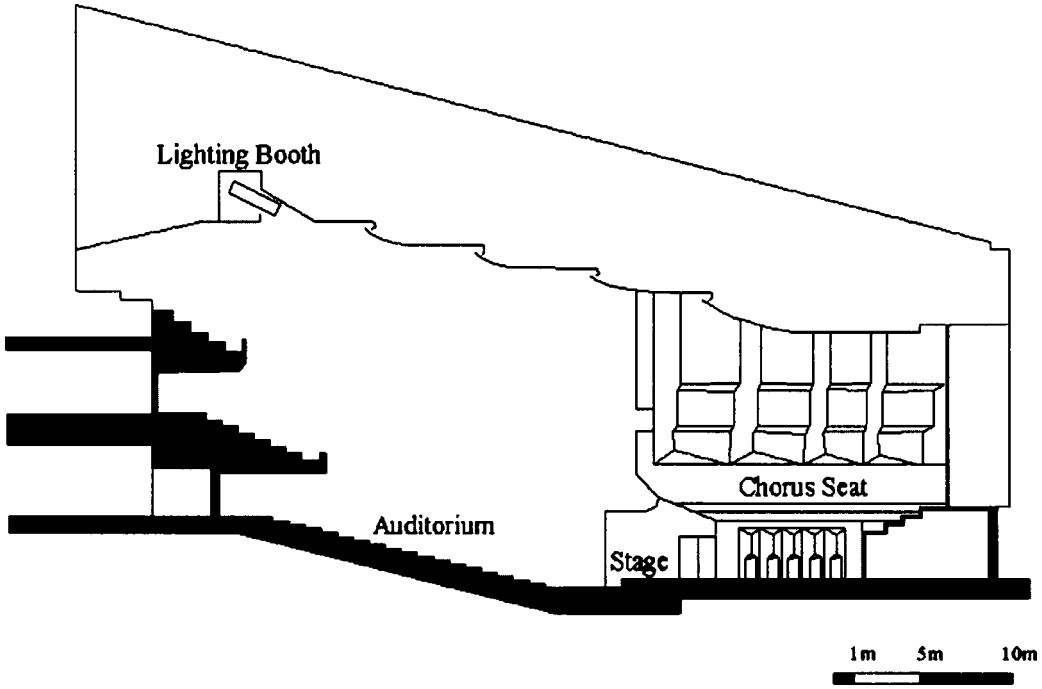


FIG. 1. Long section of the Seoul Arts Center Concert Hall.

provided some advice to this stage. The SAC Concert Hall is the principal concert hall for the city of Seoul. See Figs. 1 and 2.

The plan form of the hall is essentially fan-shaped. It has a total capacity of 2596, seated at three levels. Principal dimensions are length 42 m, width 33 m, and height 14–15 m, with an auditorium volume of 23 300 m^3. The midfrequency occupied reverberation time is 1.9 s.

A distinctive feature of the SAC concert hall is the size of the stage. With an area of 270 m^2 and an average width of 22 m, the stage is large even by contemporary standards. From the concert halls listed in Beranek's book,[9] the average stage area in older halls (Amsterdam Concertgebouw, Boston Symphony Hall, and Vienna Musikvereinssaal) is 158 m^2, whereas modern halls after 1962 have an average area of around 200 m^2. An area of 190 m^2 is generally considered adequate for a 100-piece orchestra.[7] No reflectors have been placed above the SAC stage, leaving a clear distance of 14–15 m to the ceiling.

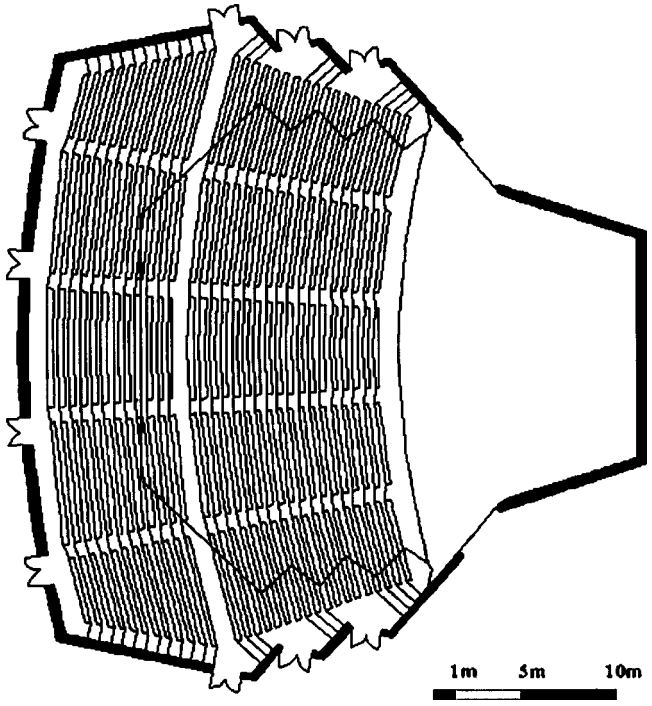


FIG. 2. Plan of the Seoul Arts Center Concert Hall.

The walls of the stage contain triangular diffusers, which were intended to contribute to sound diffusion and provide support for musicians. The diffusers are constructed of wood veneer on 6-mm plasterboard.

A. Subjective characteristics

The SAC concert hall stage is not particularly popular with musicians who play on it, as they find that they have difficulty in hearing themselves and their colleagues. When an orchestra rehearses for the first time on this stage, they find they require more time to adapt than on other stages and tend to comment on the difficulties they have hearing neighboring players as well as the players across the stage. Domestic orchestral musicians comment on the stage when they compare it with stages of other similar size auditoria in other countries; they find the SAC stage ''not drier'' but that ''it is harder to adapt to the acoustics of the stage.'' Conductors find a time delay in the reflected sound of their orchestra and a lack of a useful blend. Choir members sitting on the seats behind the stage complain of the low level of orchestral sound.

Contrary to comments by ensembles or orchestras, there are few controversial remarks from soloists. For instance, soprano Jesse Norman appeared to be satisfied with the fan-type, centrally protruding stage and mentioned that it kept her close to the audience. The violinist Gidon Cremer said

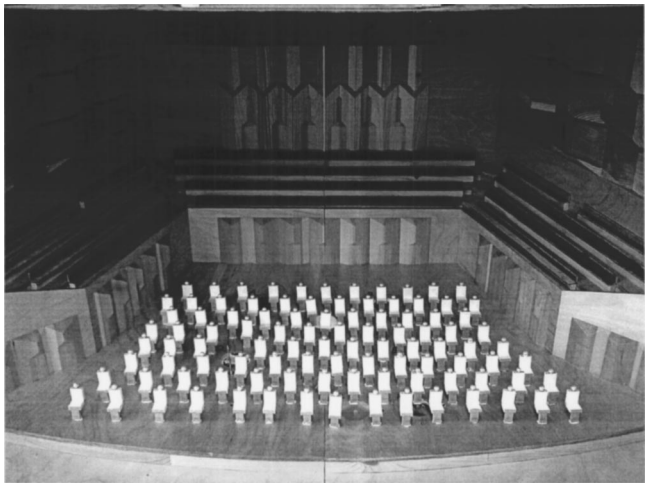


FIG. 3. 1:25 scale model of the Seoul Arts Center Concert Hall stage with 100 model musicians.

that he felt comfortable on the stage and that he preferred to use the front region of the stage. Regarding the hall itself, Beranek mentioned in a casual comment on the acoustics of the hall that the measured bass ratio during performance was relatively higher than he expected.

B. Scale model of the stage region

A 1:25 scale model of the Seoul Arts Center Concert Hall stage was built at Bath University of 9-mm varnished timber (see Fig. 3). The model included correctly modeled seating[10] in the choir region behind the stage (though in this particular hall the choir area probably has very little direct influence on stage support). Because measurements only involved the early sound, no air drying or nitrogen atmosphere was needed for the model measurements.

One complication with model tests using a spark source is that propagation is nonlinear at positions close to the source;[11] in other words, propagation deviates from the inverse square law with distance. For the 4-mJ source used for these measurements, the signal is nonlinear up to a distance of about 80 mm from the source. Since for the measurement of support the microphone is located 40 mm from the source, we can expect the direct sound to be influenced by nonlinearity. Measurements were therefore made to establish the magnitude of this effect by measuring at two distances from the source: 40 and 120 mm. The measured correction values to the direct sound were -1.5, -0.5, -1.0, 1.1 dB at 250, 500, 1000, 2000 Hz. These corrections have been applied to all model measurements of ST1.

Seated model people (also included in Fig. 3) were developed with absorption that matched real people accurately. The model people were made of medium-density fiber board (MDF) covered with a thin layer of fabric; a bead was used as a head. The absorption of the model people at a density of 2.3 m^2/person was measured in a model reverberation chamber and is compared with figures for real people,[12] as shown in Table I. One hundred model people were available to act as model musicians (see Fig. 3).

C. Computer model

A computer model of the SAC auditorium was created using the ODEON program (version 3.1). The model consisted of 629 surfaces; in general, elements larger than 0.3 m were reproduced. Raked seating was modeled as inclined surfaces. The diffusing elements around the stage were modeled individually. Tests were made using 6791 rays. The transition order (the maximum number of early specular reflections) was set at 1. Impulse responses 2 s long were calculated and scattering was set according to Lambert's law. For the stage-wall diffusers, the scattering coefficient which is independent of frequency within the present ODEON version was set at 0.3, whereas for flat surfaces it was set at 0.1 following the recommendations of the ODEON programmer.[13,14] A scattering coefficient of 0.7 was used for audience, though this is unlikely to be significant for this study of stage acoustics. The computer model was only used for an unoccupied stage. The reverberation time (RT) at 500 Hz from the computer model was in the range of 1.9–2.4 s (audience area), whereas the RT in the real hall from recent measurements at 31 positions is in the range of 1.9–2.3 s. The average error of 0.05 s is probably due to the difference of the computer model absorption from the real hall absorption. Agreement with other acoustical parameters as is follows: the range of EDT in the computer model was 1.7–2.3 s compared with 1.4–2.4 s in the real hall (average error=0.1 s); the range of C80 in the computer model was -2.6 to $+2.8$ dB compared with -4.3 to $+3.2$ dB (average error=0.7 dB). These differences are basically small and are unlikely in themselves to significantly affect prediction of conditions on stage.

III. STAGE SUPPORT (ST1) AND ITS MEASUREMENT

The support measure ST1 (or ST early) was proposed by Gade.[5,6] It is the ratio expressed in dB of the early energy (20–100 ms) to the direct sound energy (0–10 ms), as shown in Eq. (1). To measure ST1 an omnidirectional source on stage and microphone are placed 1.2 m above the floor and 1 m apart. ST1 is a measure of the reflected sound returning to a performer within 100 ms. The direct sound component will include both the direct sound and the inevitable floor reflection. Higher values of support are generally considered preferable.

$$\mathrm{ST1} = 10 \log \left\{ \int_{20\ \mathrm{ms}}^{100\ \mathrm{ms}} p^2(t)\,dt \Big/ \int_{0\ \mathrm{ms}}^{10\ \mathrm{ms}} p^2(t)\,dt \right\}, \ \mathrm{dB}. \tag{1}$$

TABLE I. Absorption by full-size and 1:25 model performers at a density of 2.3 m^2/person.

Absorption per person (m^2)	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz
Real persons—BBC (Ref. 12)	0.00	0.16	0.43	0.74	0.86
1:25 scale model people	0.11	0.24	0.31	0.61	0.83

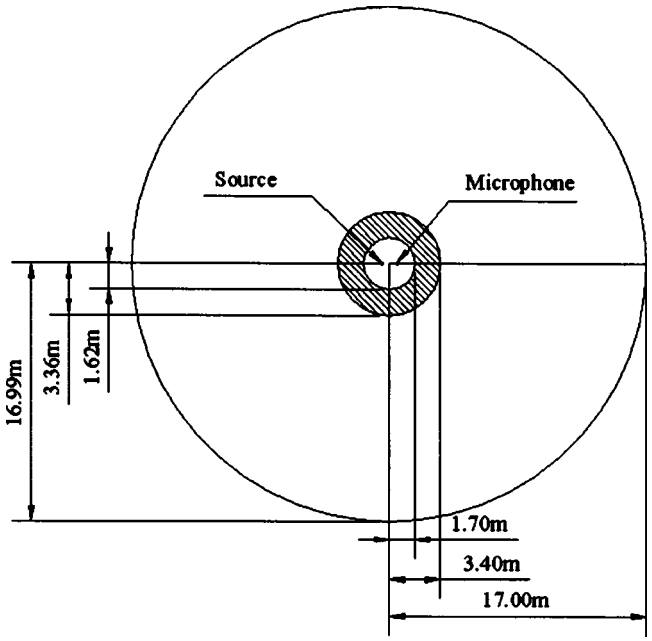


FIG. 4. Significant regions for reflectors for the ST1 support measure (see the text).

Figure 4 shows, in two-dimensional form, the important regions for reflecting surfaces for a measurement of ST1. A reflecting surface passing through the central unshaded ellipse will produce a reflection within 10 ms of the direct sound that will contribute to the direct sound component in Eq. (1). This is not desirable for the measurement, so this region should be clear of chairs, etc. during the measurement. There is, however, an inevitable floor reflection which arrives within this 10-ms period.

Reflections from surfaces in the shaded region in Fig. 4 arrive between 10 and 20 ms of the direct sound and thus have no effect on ST1; however, surfaces in this region may screen sound that would otherwise reach surfaces further away from the source and receiver. Reflections from surfaces in the outer unshaded region arrive between 20 and 100 ms of the direct sound and contribute to the support measure; the region extends from about 3.4 m to 17 m from the source/receiver.

Gade [personal communication, 1999] recommends, for the support measurement, removing objects to at least 2 m away from the transducers and to keep both transducers at least 4 m away from wall surfaces. This last recommendation ensures that reflections from walls arrive later than 20 ms after the direct sound and thus contribute to the numerator in the definition. There is no published difference limen for ST1, but a difference of 2 dB is probably significant for a performer.

ST1 takes values on concert hall stages between about -8 and -18 dB. It is generally measured in octaves between 250 and 2000 Hz. Gade has found a relationship between measured support values on stages and the ''platform volume'' (=platform area×height to the ceiling or reflectors).[7] The SAC concert hall has a platform volume of 3900 m^3, which can be expected to have a mean ST1 value of between -14 and -17 dB.

Beranek[9] also includes measured mean ST1 values for famous auditoria. Among highly ranked halls, Zurich Grosser Tonhallesaal and Lenox Tanglewood Shed, Massachusetts, have mean ST1 values of -12.6 and -12.8 dB, respectively, whereas Vienna's Musikvereinssaal has -13.9 dB. However, the Berlin Philharmonie and Amsterdam Concertgebouw have lower mean ST1 values of -16.8 and -17.8 dB. Beranek's interview with Eugene Ormandy after months of conducting in the latter hall reveals the condition of the stage acoustics in the Amsterdam Concertgebouw: ''There is a jumble of sound and poor orchestral balance. The ceiling over the stage seems to be too high.''

A. Measurement technique in the real hall

Measurements on the empty stage of the SAC hall were made with an omnidirectional dodecahedron loudspeaker at eight positions on stage. Responses were acquired using a PCI-MIO-16E-1 board (National Instruments). A maximum-length sequence (MLS) source signal was used; analysis was by a program written in LABVIEW at Seoul. ST1 measurements reported here at full size, and for models are the mean of the three octave levels between 500 and 2000 Hz.

B. Scale model measurement techniques

In the 1:25 scale model of the stage, a spark source with 4-mJ electrical discharge[11] was used as the source in the scale model with a Brüel & Kjaer 1/8-in. microphone with nose cone as receiver. Microphones were always placed 40 mm from the spark source, equivalent to 1-m full size. Signals were acquired with an EISA A-2000 board (National Instruments) and analyzed by software written at Bath in LABWINDOWS. In general, when referring to results from the scale model, full-size equivalent frequencies and dimensions will be used below.

IV. MEASUREMENT RESULTS

A. Measurements on the empty stage

Support values have been measured with the stage empty in the real hall and the two models. The same eight measurement positions were used, labeled A–H as shown in Fig. 5. Source and receiver heights of 1.2 m were used; the measurement frequencies were 500 Hz, 1 kHz, and 2 kHz. All quoted values of ST1 in this paper are the average of those three values.

As shown in Fig. 5, positions B and E are symmetrical complements of positions A and D. Positions C and F (located on the center line) were selected to investigate the effect of distance from the nearest wall profiles. Positions G and H are close to the audience area.

Figure 5 also includes circles around positions C and D. Plane surfaces tangential to these circles provide reflections back to the microphone with delays of 10, 20, and 50 ms. The significance of the delays 10 and 20 ms was explained above in Sec. III and illustrated in Fig. 4. One notes from Fig. 5 that at position C reflections within 50 ms arrive from three surfaces to contribute to support, whereas at position D one surface will provide major reflections and two other sur-

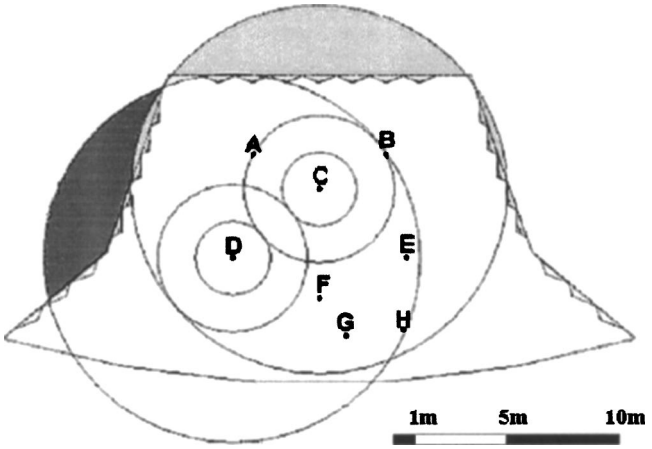


FIG. 5. Measurement positions on stage. For positions C and D, circles are drawn representing regions where reflectors would produce reflections with delays of 10, 20, and 50 ms.

faces may make marginal contributions. Figure 6 shows a comparison of log impulse responses for positions C and D. In light of the comments above and analysis of Fig. 5, it is interesting to note that many reflections delayed about 50 ms arrive at position C but not at position D. Stage support is higher at position C than D (see below).

Measured results of ST1 in the real hall and the two models are plotted for the eight positions in Fig. 7, in all cases with an empty stage. Values for the real hall and the two models follow the same trend with the highest values at positions A and B, lower values for positions further forward on stage, and minimum values at positions F and G. Measured values at the symmetrical positions A and B as well as D and E are similar in each ''model,'' as we would expect.

However, measured values in the real hall are consistently higher than in the models. We ascribe this to reflections from seating, etc. back to the stage for the real hall; the scale model was only of the stage area, while the computer model treated seating as plane surfaces. (Note that in Fig. 5, the largest circles relate to 50-ms delays rather than the 100-ms delay relevant to ST1.)

Agreement in Fig. 7 between the scale and computer models is good except at position G. Subsequent checks on the scale model have shown that there were inaccuracies in the model construction in the ceiling area above the stage

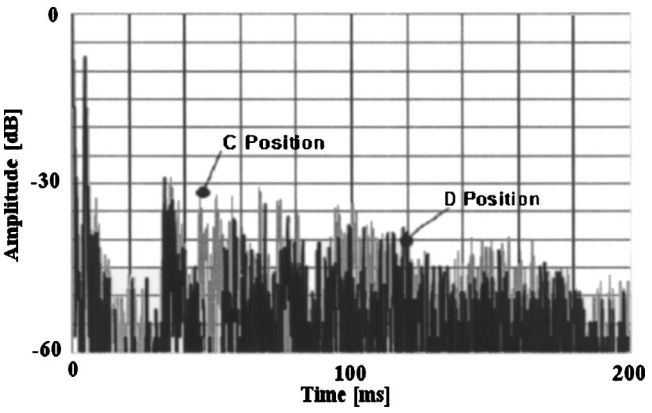


FIG. 6. Log impulse responses at positions C and D.

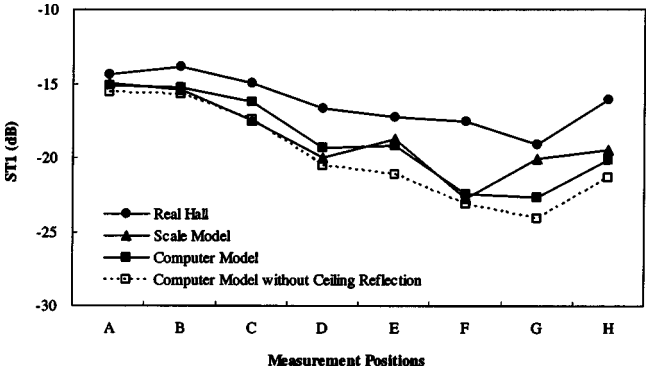


FIG. 7. Measured values of ST1 at different measurement positions (see Fig. 5) for the empty stage in the real hall and scale and computer models.

front. Calculations with the computer model matching either the real hall or the scale model show that ST1 is very sensitive to the geometrical details in this ceiling area, producing variations in line with those observed between the scale and computer models for position G in Fig. 7.

B. Effect of musicians on stage on the measurement of ST1

ST1 was measured in the physical scale model with model musicians on stage. To clarify the influence of musicians on the measurement of support, the model musicians closest to the source were removed to leave different distances clear around the source. Initially 100 model musicians were placed on the stage, as shown in Fig. 3. The musicians closest to the source were removed in 1-m steps, with measurements made when the closest musicians were at 1 to 5 m from the source. In the case of a 1-m distance of musicians from the source, there were musicians adjacent to the measurement microphone, which is also at 1 m from the source. ST1 was measured at eight positions to determine the influence of musicians on measured values of ST1.

The results of this test are shown in Fig. 8, which shows that the highest values for ST1 occurred when the closest musicians were 3–4 m from the source. This result can be explained as follows, with reference to Fig. 4. With musicians 1 m from the source, some diffuse reflections from the closest musicians will probably result in an increase in the

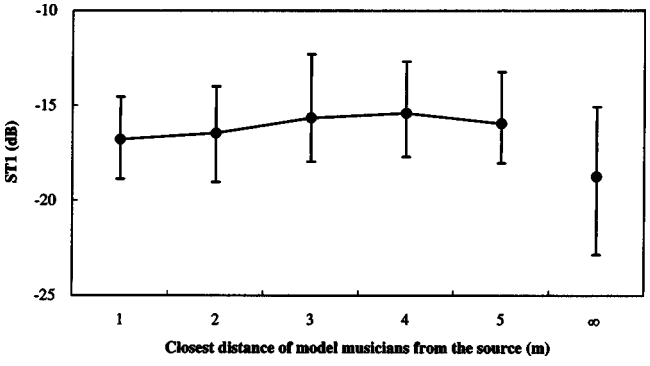


FIG. 8. Measured values of ST1 in the scale model for different distances of the closest model musicians to the source. A ''closest distance'' of ∞ implies an empty stage. The mean and range of values for the eight measurement positions are given.

''direct sound component'' of ST1, reducing its value. With musicians at 2 m, any reflections off the closest musicians fall in the neutral delay zone of 10–20 ms that does not contribute to ST1. The closest musicians will, however, obscure sound from reaching musicians behind them. With the closest musicians at 3 and 4 m, the majority of reflections off the closest musicians will contribute to the numerator of ST1; hence, the higher values. With the closest musicians at 5 m (and beyond), reflections off the closest musicians are weaker due to spherical spreading; hence, the lower values of ST1. With no musicians on stage (closest distance=∞ in Fig. 8), measured support is on average lower than with musicians, regardless of how close they are to the source. One notes also that the spread of ST1 values with an empty stage is larger.

In a real playing situation with an orchestra, the performer will receive diffuse reflections from all neighboring players. For the measurement of support it is necessary to measure the direct sound component, for which the first 10-ms period is allocated. No reflections apart from the floor reflection should be included here. The support measure has a period starting at 20 ms in the numerator, reflecting the fact that very early reflections are not helpful to performers. Chairs and musicians that are close to the performer will, however, obscure some sound from reaching musicians further away. For this reason, Gade's recommendation (Sec. III), which is a minimum requirement that objects should be removed at least 2 m from transducers, appears correct.

The model measurements, shown in Fig. 8, indicate however that support values may increase for minimum distances greater than 2 m. The distance of the nearest chairs is a variable which influences the measured ST1. This suggests for support measurements that the 2 m should not be just a minimum recommended distance but that measurements should always be made with chairs (or musicians) located at a distance of 2 m and beyond.

ST1 measured with model musicians at 2 m from the source was on average 2.3 dB greater than with an empty stage. The presence of musicians obscures reflections from the walls surrounding the stage but provides compensating reflections from the musicians themselves. It is not surprising therefore to discover that the effect of musicians varies considerably with measurement position with a maximum difference of 5.2 dB at position F and a minimum difference of 0.5 dB at position H. A stage occupied with chairs is appropriate to an orchestral performance. Measurements on an empty stage would be appropriate to performance with a small chamber group or soloist.

C. Effect of moving the orchestra

Because of the large 270-m^2 area of the SAC stage, there is flexibility on this stage regarding the location of an orchestra occupying the traditional 190 m^2. Support measurements were made in the scale model with the 100 model musicians occupying the equivalent of 190 m^2 but moved along the direction of the axis of symmetry of the hall with the rear of the orchestra located at three distances from the stage rear wall: 0, 2, and 4 m. With the rear of the orchestra 4 m from the rear wall, the front performers were next to the front of the stage. Measurements were made at all measurement positions for each orchestra location. For each orchestra location some source/receiver positions were outside the region occupied by the orchestra, so measurements were made at all available positions *within* the orchestra area concerned. Thus, for distances of 0 and 2 m for the back of the orchestra from the rear wall, results for measurement positions A–E have been used. For the orchestra position 4 m from the stage back wall, measurement positions D–H were used. For all measurements, the musicians within 0.5 m of the source were removed.

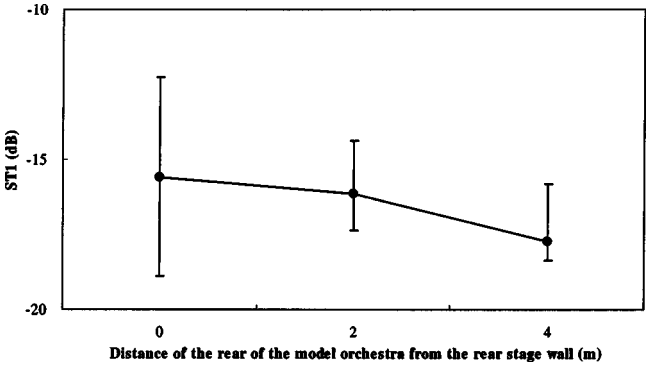


FIG. 9. Mean values of support, ST1, within the orchestra for three positions of the orchestra on stage in the scale model. See the text for measurement positions.

The mean values for support (averaging over 5 points in each case) for the three orchestra positions are plotted in Fig. 9. The location offering the most support is, not surprisingly, the one in which the orchestra has the most surfaces close to it, in other words when it is at the back of the stage.

In reality, different orchestras use the stage differently. Orchestras from the Seoul region that play regularly in the hall generally spread out and use the whole stage. Visiting foreign orchestras often place themselves at the front of the stage, whereas some Korean provincial orchestras play at the rear of the stage.

D. Effect of different ceiling heights above the stage

The ceiling height above the stage in the Seoul Arts Center is 14–15 m, which is large by concert hall standards. No suspended stage reflectors are used in the hall. Investigations were made in the computer model into the acoustic significance of the ceiling reflection and the effect of reducing the height above the stage.

To investigate the effect of the ceiling reflection on ST1, the ceiling surface in the computer model was made fully absorbing. ST1 was calculated at the three octave band frequencies (500–2000 Hz) at eight positions (A–H) on the empty stage. The result of this exercise is included in Fig. 7 as the dashed line. It can be seen that the ceiling reflection makes a small contribution of 1 dB on average to ST1. We would expect this, since the ceiling is within the limiting distance of around 17 m from the source (Fig. 4).

Calculations were also made by computer model of ST1 for lower height ceilings down to 8 m above the stage. The results of this exercise are given in Fig. 10. Reducing the ceiling height from 14 to 8 m increases ST1 by 2 dB on

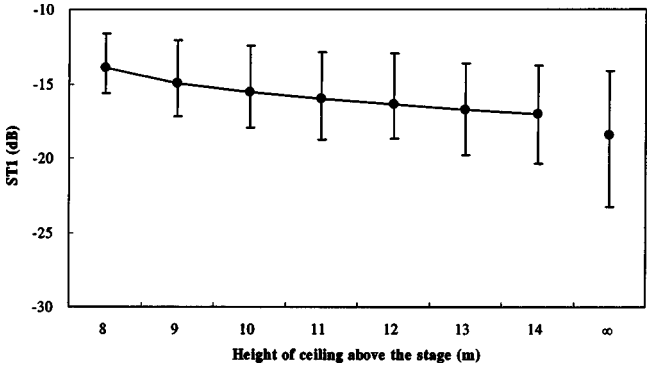


FIG. 10. The effect of ceiling height above the stage on support, ST1, as calculated by computer model. The mean and range of ST1 values were calculated at eight positions. A height of ∞ refers to no ceiling reflection; this condition is also plotted in Fig. 7.

average. The range of ST1 values for the existing 14-m-high ceiling is −15 to −24 dB; with an 8-m-high ceiling the range becomes −15 to −19 dB. Lowering the ceiling is thus of major value for performing positions which are most exposed, in other words those that are furthest from vertical reflecting surfaces.

E. The effect of the stage wall profile

The five wall surfaces surrounding the stage in the SAC Concert Hall contain triangular elements (when viewed in plan). The dimensions of these elements are 1.5×2.8 (width ×height); in plan the apex of these elements is 0.3 m from the vertical wall behind. The intention of these surfaces was that they should provide diffusion. Tests were performed in both the scale model and by computer model to determine the acoustic value of these diffusing surfaces. In the case of the scale model, flat surfaces were placed in front of the triangular elements in order to obscure them. In the computer model, the stage wall surfaces were simply made plane instead of profiled. In both models, the stage was unoccupied by musicians and ST1 was measured at eight positions. The results of these measurements are given in Fig. 11.

Considering first the situation at positions G–H, both scale and computer models predict little change in support with wall profile. At positions A–F, the computer model predicts significant improvements with plane rather than diffusing surfaces. The scale model is, however, likely to reproduce diffraction behavior more accurately than the computer model, and thus in reality only small changes are likely for all the measurement positions.

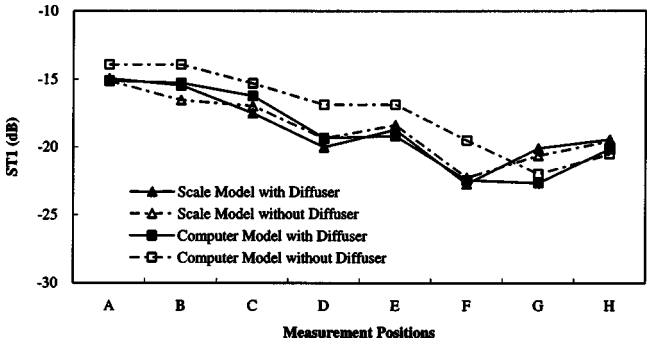


FIG. 11. The effect of profiling of stage walls. Measured values of ST1 in the scale and computer model with and without profiled stage walls. See Fig. 5 for location of measurement positions.

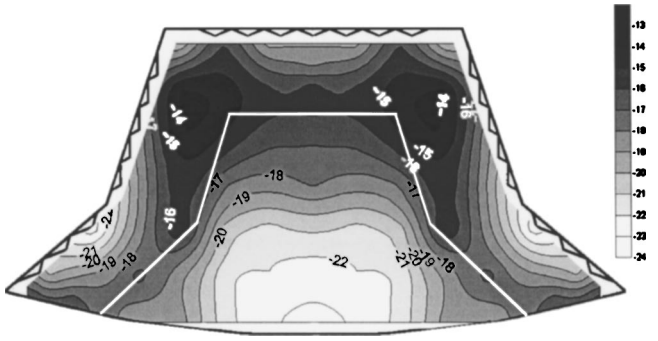


FIG. 12. One-m grid map of ST1 for the empty stage using computer simulation.

Generally, both scale and computer models predict higher support for plane rather than diffusing stage-wall surfaces. The differences are on average 0.0 dB for the scale model and 0.3 dB for the computer model; again, the scale model is likely to be more accurate here.

The reason for the superiority of plane surfaces is obvious in simple geometrical terms, which is relevant at higher frequencies. A triangular diffuser will direct reflections away from the incident direction, whereas a large plane surface will always provide a specular reflection back to the source.[14] The wall profile on the SAC Concert Hall stage may thus be less than ideal for support of the player's own sound, though it may have advantages in terms of hearing sound from other performers. A diffusing surface around the stage may also have benefits for the sound reaching the audience.

F. Variation of support over the stage

Calculation of ST1 has been made by computer at points on a 1-m grid for the empty stage, as shown in the grid map in Fig. 12. The map shows values up to 1.5 m from the stage walls. Though Gade (Sec. III) recommends making measurements at least 4 m from wall surfaces, the additional area clarifies the contour pattern.

The main influences on support for this stage are the stage shape in plan, the diffusing wall profile, and the overhead reflector. The influence of the wall profile was discussed in the previous section, while the effect of the ceiling was investigated in Sec. IV D. In Fig. 12 the highest ST1 values are found at positions around 4 m from both the stage rear wall and side wall. In the front section of the stage, where the stage side walls follow a wide fan, the support drops off sharply as one moves towards the stage front.

V. CONCLUSIONS

A study has been made of the acoustic conditions on an individual concert hall stage using two forms of modeling: scale and computer. The purpose was not to check the relative accuracy of these two forms of modeling but to exploit the particular advantages of each modeling approach. The

study used Gade's support measurement and the opportunity was taken to investigate the effect of occupancy by musicians on measured support.

The support measure ST1 involves measuring energy returning to a receiver 1 m from the source within 20–100 ms of the direct sound. For first-order reflections, this implies surfaces between about 3.4 to 17 m from the source.

Measurements in the scale model with model musicians on stage showed that the presence of musicians can influence measured support. Gade has proposed that musicians (or chairs) should be at a minimum of 2 m from the source or receiver. Measurements reported here suggest that, for consistency, chairs should always be included beyond the 2-m distance.

The stage in the Seoul Arts Center concert hall is both wide and large, with a plan area of 270 m^2. The level of support varies widely from the front to the rear of this stage. Performers at the front of the stage receive poor support. This may well be a characteristic of many concert hall stages and is of course unfortunate, given the importance of the strings for the orchestral sound.

The ceiling in the SAC concert hall is also high above the stage and thus of only minor value for performers. Several measurements lead to the obvious conclusion that good support depends on the presence of sufficiently close surfaces capable of reflecting sound back to performers. The SAC concert hall stage is surrounded by profiled walls which prove not to be very helpful for performers hearing their own sound, but the profiling may well have other acoustic advantages.

This study has shown how a large concert hall stage can compromise the performers' ability to hear themselves and their colleagues. The larger the stage, the more effort is required by designers to ensure that sufficient reflections return to the performers.

ACKNOWLEDGMENTS

The authors wish thanks to Chung Hwa Lee and Byung Kwon Lee for their assistance in conducting this study.

[1] M. Barron, "The Gulbenkian Great Hall, Lisbon II. An acoustic study of a concert hall with variable stage," J. Sound Vib. **59**, 481–502 (1978).
[2] A. H. Marshall, D. Gottlob, and H. Alrutz, "Acoustical conditions preferred for ensemble," J. Acoust. Soc. Am. **64**, 1437–1442 (1978).
[3] J. Meyer, "Influence of communication on stage on the musical quality," 15th ICA Trondheim, 573–576 (1995).
[4] J. O'Keefe, "A preliminary study of reflected sound on stages," 15th ICA Trondheim, 601–604 (1995).
[5] A. C. Gade, "Investigations of musicians' room acoustic conditions in concert halls. I. Method and laboratory experiments," Acustica **65**, 193–203 (1989).
[6] A. C. Gade, "Investigations of musicians' room acoustic conditions in concert halls. II. Field experiments and synthesis of results," Acustica **69**, 249–262 (1989).
[7] M. Barron, *Auditorium Acoustics and Architectural Design* (Chapman & Hall/Routledge, London, 1993).
[8] J. H. Rindel, "Design of new ceiling reflectors for improved ensemble in a concert hall," Appl. Acoust. **34**, 7–17 (1991).
[9] L. Beranek, *Concert and Opera Halls: How They Sound* (Acoust. Soc. Am. through Am. Inst. Physics, New York, 1996).
[10] M. Barron and S. Coleman, "Measurements of the absorption by auditorium seating—A model study," J. Sound Vib. **239**, 573–587 (2001).
[11] M. Barron and C. B. Chinoy, "1:50 scale acoustic models for objective testing of auditoria," Appl. Acoust. **12**, 361–375 (1979).
[12] H. D. Harwood and B. K. F. Lansdowne, "Acoustic Scaling," BBC Research Department, Report No. 1972/3, 1–7 (1972).
[13] C. Lynge, ODEON *Room Acoustics Program Version 3.1, Industrial, Auditorium and Combined Editions Manual* (The Department of Acoustic Technology, Technical University of Denmark, 1998).
[14] P. D'Antonio and T. Cox, "Two decades of sound diffuser design and development. I. Applications and design," J. Audio Eng. Soc. **46**, 955–976 (1998).

Estimation of frequency-averaged loss factors by the power injection and the impulse response decay methods

Brandon C. Bloss and Mohan D. Rao
Mechanical Engineering-Engineering Mechanics Department, Michigan Technological University, Houghton, Michigan 49931

(Received 4 June 2004; revised 31 October 2004; accepted 1 November 2004)

This paper describes a comparison, both analytically and experimentally, between two widely used loss factor estimation techniques frequently used in statistical energy analysis. Analytical models of simple spring/mass/damper systems were created to compare frequency-averaged loss factor values from the single subsystem power injection method and the impulse response decay method. The parameters of the analytical models were varied to study the effects of the total number of modes, amount of damping, location of modes within frequency bands, and the width of the frequency bands on loss factor estimation. The analytical study shows that both methods give accurate loss factor values as long as the damping values remain realistic for linear systems and at least one modal resonance is present in each frequency band. These analytical results were verified experimentally by measuring the loss factors of simple steel plates, with and without damping treatments applied. [DOI: 10.1121/1.1835512]

PACS numbers: 43.55.Wk, 43.40.At. [NX] Pages: 240–249

I. INTRODUCTION

Statistical energy analysis (SEA) models allow the prediction of noise and vibration response of complex, built-up resonant structure, such as an automobile using a computer without having to build and test prototypes. With SEA, the dynamic characteristics, resonant responses, and noise radiation can be approximated from a theoretical aspect using energy flow equations based on the laws of the conservation of energy.[1,2] By breaking up a complex group of components into subsystems, based on the level of modal coupling and various approximations, models can be solved extremely quickly—especially when compared to the time it takes to build and test a prototype structure.

One of the parameters needed to complete a SEA model is the internal loss factor or damping of the individual subsystems. Damping serves to control the steady-state resonant response and to attenuate traveling waves in the structure. Unlike other physical parameters, such as size, shape, and mass, the internal loss factor is difficult to estimate analytically unless the damping is specifically designed to perform a certain way. This creates a need to obtain most damping values experimentally. Damping measurements on systems and structures are rarely straightforward due to the complexity of the dynamic interaction of system joints and geometry. Various experimental techniques can be used for damping estimation, such as frequency domain modal analysis curve-fitting methods, time domain decay-rate methods, and other methods based on energy and wave propagation.[1,3] Each method has its own set of advantages and drawbacks. In this paper, an analytical study and an experimental comparison between two widely used loss factor estimation techniques frequently used in SEA are undertaken. The techniques considered are the decay rate-based impulse response decay method (IRDM) and the energy-based power injection method (PIM). There is a need to understand the differences between these two techniques and the effects of these differences on hybrid SEA models to improve their prediction capabilities.

A. Impulse response decay method (IRDM)

The IRDM and other decay rate methods are based on measuring the system's decay rate when the system or structure is subjected to an impulse (a pistol shot, tone burst, or balloon pop) or an interrupted stationary excitation. The system's response amplitude when plotted versus time will have a slope proportional to $e^{(-\pi f \eta t)}$, where f is the frequency, η is the loss factor, and t is time. The energy of the system is proportional to the peak amplitude squared; therefore, the system response will decay at a rate equal to $Ge^{(-\pi f \eta t)}$, where G is related to the peak response amplitude of the system. The decay slope between two instances in time t_1 and t_2 with amplitudes G_1 and G_2, respectively, in decibels, is given by[1]

$$20 \log_{10}\left(\frac{G_1}{G_2}\right) = 27.3 f \eta (t_2 - t_1). \tag{1}$$

Solving for η and defining the decay rate, D, in terms of decibels per second gives

$$\eta = \frac{D}{27.3 f}. \tag{2}$$

The loss factor can be found for a single, well-defined modal resonance or as the average loss factor of many modal resonances over a given frequency range. In both cases, the initial decay slope is found by plotting the system's response on a logarithmic amplitude scale, with a linear time scale as shown in Fig. 1.[1]

The IRDM is similar to the initial decay rate method described above, but utilizes the impulse response of the system instead of the output response. The impulse response can be obtained from an inverse Fourier transform of a system's

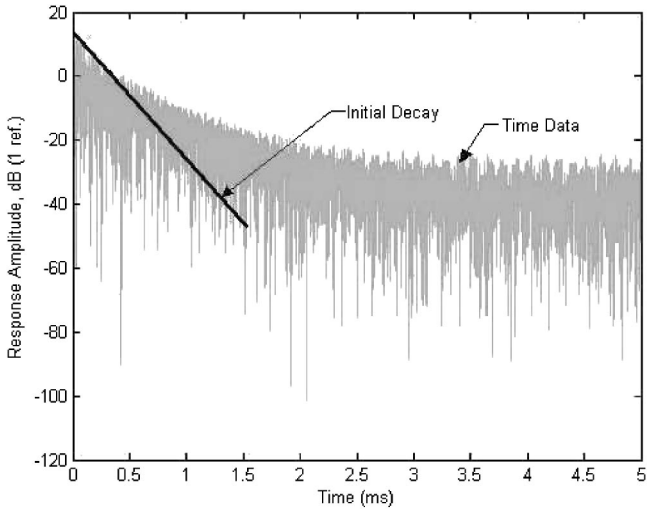


FIG. 1. Illustration of initial decay rate from a system's response.

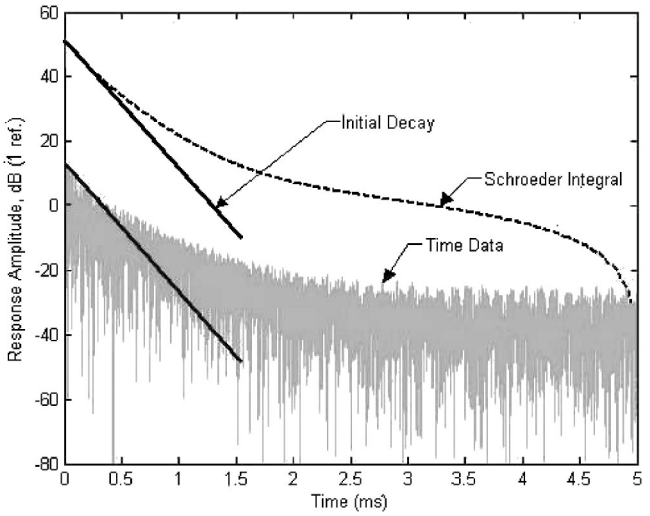


FIG. 2. Illustration of decay rate using both raw data and the Schroeder integral.

frequency response function (FRF) or using the maximum length sequence (MLS) method.[4,5] In the MLS method, the impulse response is obtained by calculating the cross correlation between the MLS input and the system output with a great amount of accuracy and repeatability. The initial decay rates can often be difficult to identify due to irregularities or noise caused by excitation signal randomness and other factors. The slope may not be straight or more than one slope will appear in the time domain. The Schroeder integral method[6] is used to smooth random fluctuations in the decay curves, which also helps alleviate identification problems. In the conventional method of measuring decay, an interrupted continuous noise signal is generally used. Schroeder[6] has shown that the expected response $x(t)$ to an excitation switch off at time $t=0$ is related to the impulse response $h(t)$ by the following equation:

$$E(x^2(t)) = G\int_t^{\infty} h(\tau)d\tau, \tag{3}$$

where G is a constant related to the excitation level.

The main advantage of using Schroeder's method is that the decay curve obtained from a single measurement is identical to the average of many decay curves that would be obtained from the system using the interrupted bandpass-filtered signal. The method is particularly useful for measuring the reverberation time in architectural acoustics. Here, since it is assumed that $x^2(t)$ is exponential, integrating an exponential function is also an exponential function with the same exponential factor. As can be seen in Fig. 2, after implementing the Schroeder integral the same initial slope of the data is maintained, while the decay curve is much smoother and easier to interpret. It should be noted that the decay represented in Fig. 2 is ideal (i.e., not experimental) data, so it is relatively easy to pick the initial decay rate in this instance without the smoothing. A comparison of reverberation time measurements using Schroeder's impulse response and decay-curve averaging methods is given in Ref. 7. The accuracy of the decay rate estimations can further be improved using a nonlinear regression approach proposed by Xiang[8] instead of the traditional linear models.

The upper limit of the integration in Eq. (3) is displayed as ∞, but in practical cases, the upper limit of the integration is limited by the maximum time of the recorded decay curve. As the lower limit of integration progresses towards the upper limit of integration, few measurements can be summed—this causes the Schroeder integral curve to slope off very sharply at the end. This is usually not a problem since only the initial decay rate is of interest.

B. Power injection method (PIM)

The PIM is based on a comparison of the dissipated energy of a system to its maximum strain energy under steady-state vibration. The PIM, as described in Refs. 9–11, approximates the loss factor of a system by the ratio of the energy dissipated within a system per radian of motion to its total strain energy. The loss factor is defined as

$$\eta(\omega) = \frac{\Delta E}{E_{\mathrm{SE}}}, \tag{4}$$

where E_{SE} is the strain energy, ΔE is the dissipated energy, and η is the loss factor within the frequency range ω considered. If the input force is stationary at a fixed location, then the dissipated energy, ΔE, can be replaced by the input energy, E_{in}. This is acceptable since input energy is equal to dissipated energy under steady-state conditions. In Eq. (4), neither the input energy nor the total strain energy can be measured directly. However, input energy acting on a system can be calculated by simultaneously measuring the input force and the input velocity at the excitation location. The numerator of Eq. (4) can then be replaced by

$$E_{\mathrm{in}} = \frac{1}{2\omega}\mathrm{Re}[h_{fv}(\omega)]G_{ff}(\omega), \tag{5}$$

where h_{fv} is the driving point mobility (velocity/force) function and G_{fv} is the power spectral density function of the input force. To find an estimate of total strain energy, a few assumptions must be made. First, the total strain energy is replaced with total kinetic energy—this can be represented as

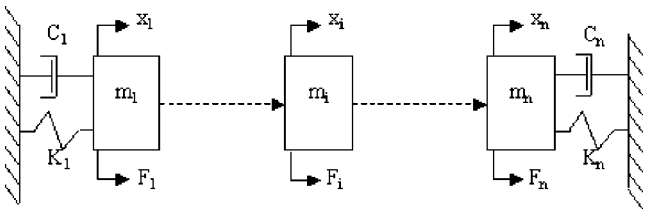


FIG. 3. Lumped mass–spring–damper n-degree-of-freedom system.

$$E_{\mathrm{KE}}=\frac{1}{2}\int_{v}\rho G_{vv}(\omega)dv, \tag{6}$$

where E_{KE} is the total kinetic energy, ρ is the system density, and G_{vv} is the power spectral density (PSD) function of the velocity vector. Replacing strain energy with kinetic energy is a valid assumption at natural frequencies or when averaged across many modal resonances within a frequency band. The second assumption needed to obtain loss factors via the PIM is replacing the volume integral in Eq. (6) with a summation. With N number of response measurement locations distributed evenly over the entire system, kinetic energy can be estimated by

$$E_{\mathrm{KE}}(\omega)=\frac{1}{2}\sum_{i=1}^{N} m_i G_{ii}(\omega), \tag{7}$$

where m_i is the mass of the ith portion of the discretized system and G_{ii} is the power spectral density function of the velocity response at measurement location i. Finally, the transfer mobility function is written as

$$|h_{if}(\omega)|^2=\frac{G_{ii}(\omega)}{G_{ff}(\omega)}, \tag{8}$$

where h_{if} is the transfer mobility function, defined as the ratio of PSD of velocity response at measurement location i to PSD of input force at the chosen input location. Equation (8) is written with the assumption that the measurements are taken with minimal noise content. By utilizing such assumptions, having equally spaced measurement locations, and equal mass per portion (such as in a uniform plate), Eq. (4) can be approximated by combining Eqs. (5), (7), and (8) as

$$\eta(\omega)=\frac{\mathrm{Re}[h_{fv}(\omega)]}{\omega m\Sigma_{i=1}^{N}|h_{if}|^2}, \tag{9}$$

where m is now a constant mass for each portion. All of the variables in Eq. (9) can be determined experimentally.

II. ANALYTICAL COMPARISON

A. Ideal systems for loss factor comparison

An ideal analytical model of n degrees-of-freedom system was used to compare the PIM and the IRDM estimated loss factors. This system takes the general form of a lumped spring, mass, and damper system as shown in Fig. 3.

Given the force input, F_i, into the system at location i, from dynamics, the potential energy can be expressed as

$$E_{\mathrm{SE},i}=\tfrac{1}{2}F_i x_i, \tag{10}$$

where x_i is the displacement at location i. For a multi-degree-of-freedom system, the total potential energy can be expressed as a summation of all individual potential energies

$$E_{\mathrm{SE}}=\frac{1}{2}\sum_{i=1}^{n} F_i x_i. \tag{11}$$

The force can be expressed in terms of the spring constants, K_{ji}, and the displacements as

$$F_i=\sum_{j=1}^{n} K_{ji}x_i. \tag{12}$$

This allows the total potential energy to be written as

$$E_{\mathrm{SE}}=\frac{1}{2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} K_{ij}x_j\right)x_i=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_{ij}x_j x_i. \tag{13}$$

In matrix form, Eq. (13) can be written as

$$E_{\mathrm{SE}}=\tfrac{1}{2}\{x\}^t[K]\{x\}, \tag{14}$$

where $\{x\}$ is the displacement vector and $[K]$ is the stiffness matrix of the complete system. Given that $[m]$ is the mass matrix, $[C]$ is the damping matrix, and $\{x\}$ is the velocity vector of the system, similar derivations to the one outlined above allow the total kinetic energy and dissipated energies to be written, respectively, as[12]

$$E_{\mathrm{KE}}=\tfrac{1}{2}\{\dot{x}\}^t[m]\{\dot{x}\} \tag{15}$$

and

$$\Delta E=\tfrac{1}{2}\{x\}^t[C]\{\dot{x}\}. \tag{16}$$

Using displacement and velocity vectors found from modal analysis and user-defined matrices for mass, stiffness, and damping, the standard equation of motion for the system can be written in matrix form as

$$\{F\}=[m]\{\ddot{x}\}+[C]\{\dot{x}\}+[K]\{x\}. \tag{17}$$

A MATLAB program was written to find the simulated loss factor using Eq. (1) for assumed values of M, K, and C matrices. The simulated values were compared with the loss factor estimated from the PIM and the IRDM. Even though the simulated method and the PIM are both capable of finding loss factors for narrow frequency bands, the comparison between the methods was performed using frequency band-averaged loss factors since the IRDM can only estimate loss factors for well-defined modal resonances or when averaged across many modal resonances.

B. Averaging techniques in the IRDM

To determine loss factors with the IRDM for an entire system, it is necessary to spatially average loss factors of individual measurement points across the complete system. For real systems, this can be done by a number of averaging schemes. For the ideal analytical case, it is only possible to average all the columns from the FRF matrix. In effect this amounts to averaging all modes over the complete system. The averaging can be done prior to filtering the responses into frequency bands; or, it is possible to average the decay rates obtained from each individual column after each decay

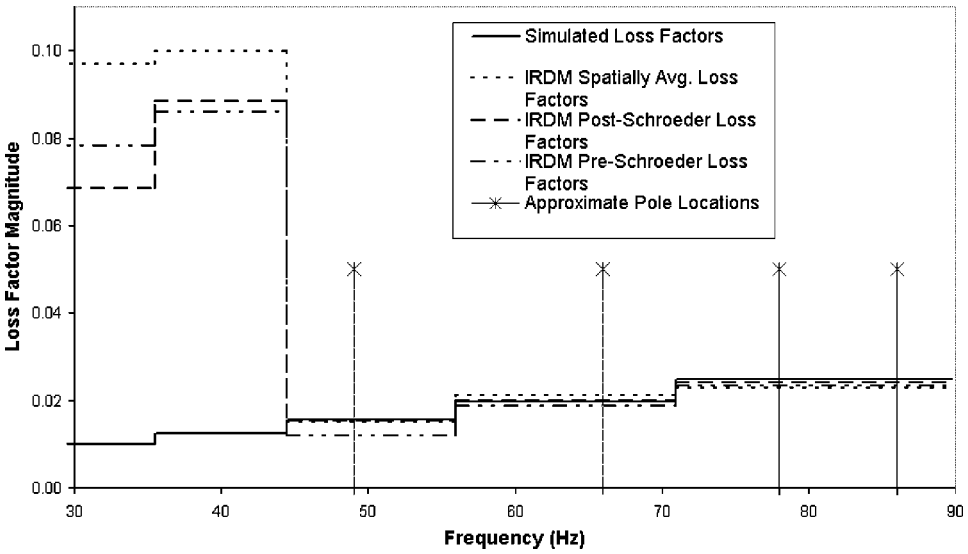


FIG. 4. Comparison of simulated and IRDM loss factors with different averaging techniques.

slope is identified for a given frequency band. Also, when averaging the individual decay slopes, it is possible to average before or after the Schroeder integral has been implemented.

A comparison of simulated loss factors and loss factors obtained from the IRDM utilizing the above-mentioned averaging schemes can be seen in Fig. 4. It should be noted that IRDM loss factors found within frequency bands where no resonances exist were significantly greater than the simulated loss factors, regardless of averaging technique. Also, spatial averaging of the modes and averaging the individual decay slopes vary greater from the simulated loss factor than loss factors found by averaging after implementing the Schroeder integral.

The post-Schroeder averaging scheme compares the simulated loss factors better than the pre-Schroeder averaging because the data utilized for the averages are smoothed—i.e., reduced fluctuations in the decay curves. Averaging the decays instead of the frequency response curves also alleviates the possibility of modes being averaged out of the FRFs. The post-Schroeder integral averaging scheme had the least amount of error so it was subsequently used for the duration of this study, both for the analytical and the experimental portion.

C. The effect of user input

During the different averaging techniques described above, it was necessary for the user to determine the initial slope of the decay signals manually. It became apparent that this user input could be a possible source of error for IRDM loss factors. Figure 5 shows the possible differences to find the decay slopes from the same data using the same averaging technique, but with three unique initial slopes.

Except for the lowest frequency band in Fig. 5, the three-user input trials show relatively consistent results, even for the frequency bands with no resonant frequencies present. However, it is important to have a consistent and automated scheme for finding the initial decay slope in order to reduce the variability associated with this factor. This will allow for a better comparison with the PIM loss factor estimations.

A best-fit line through the first 10 dB of decay was used to locate the initial slope of the data. A curve fit of the first 10

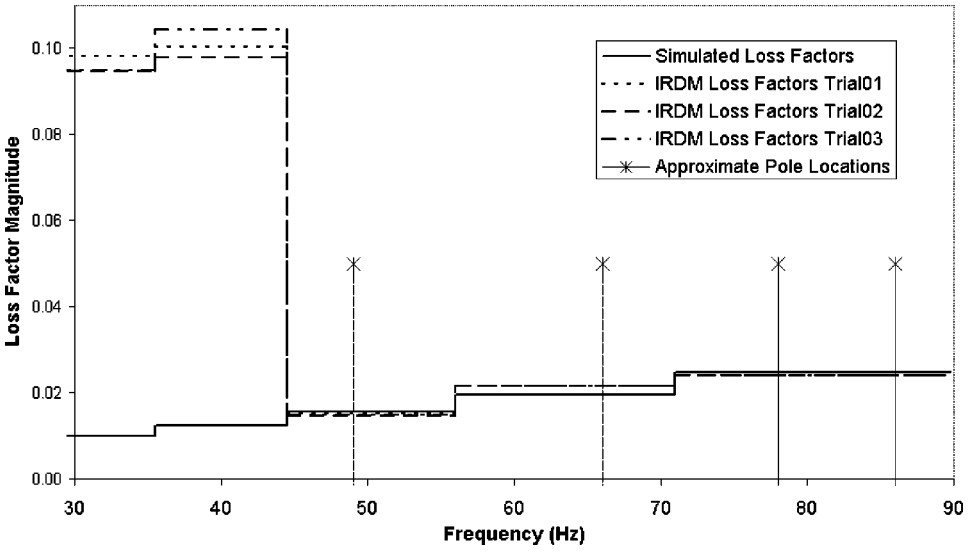


FIG. 5. Illustration of the effect of user input on IRDM loss factor estimation.

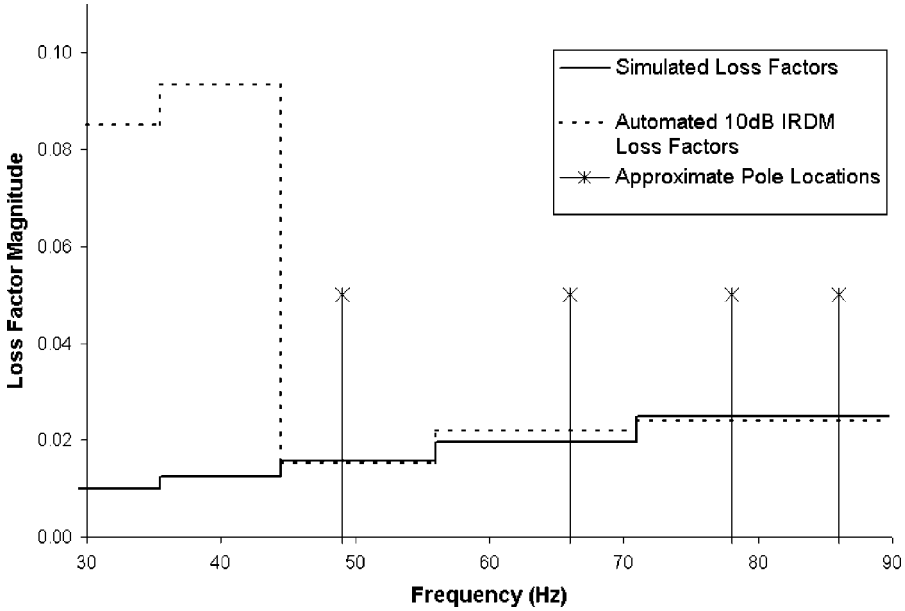


FIG. 6. Comparison of simulated loss factors and loss factors found through IRDM.

dB allows the proper damping level to be found even if the initial decay expired (i.e., changed to a different slope) rather quickly. The longest decay curve possible should be used to find the slope, but since the length of the initial decay changes from frequency band to frequency band, it would require user input at each stage—this would negate the use of an automated best-fit technique. Using less than the first 10 dB could cause erroneous damping values due to small variations in the Schroeder integrated curve. Figure 6 shows that the IRDM loss factors estimated via the 10-dB curve-fit technique compare well with the simulated loss factors except for the two bands where there are no poles.

The various averaging schemes mentioned previously were re-examined using the automated slope-picking technique to make sure that the good agreement found through the post-Schroeder integral averaging was not simply a matter of the user picking appropriate decay slopes. As was found originally, the post-Schroeder averaging has the highest level of agreement but the results were not shown here due to brevity. Since the automated 10-dB IRDM has the best level of agreement with the simulated loss factors when used with the post-Schroeder averaging, it was used for the rest of this study.

D. Comparison of loss factors from IRDM and PIM

Figure 7 shows the one-third-octave band averaged loss factors of the ideal system using the PIM, the IRDM, as well as the simulated loss factors. Similar to the IRDM loss factors, the PIM values correlated with the simulated loss factors better when poles are located within a frequency band Unlike the IRDM, the PIM does a much better job of estimating the loss factors when no modal resonances exist within a frequency band. When the width of the frequency bands increases, the divergence between the PIM loss factors and the simulated loss values increases. This is true until at least two modal resonances are present in a frequency band.

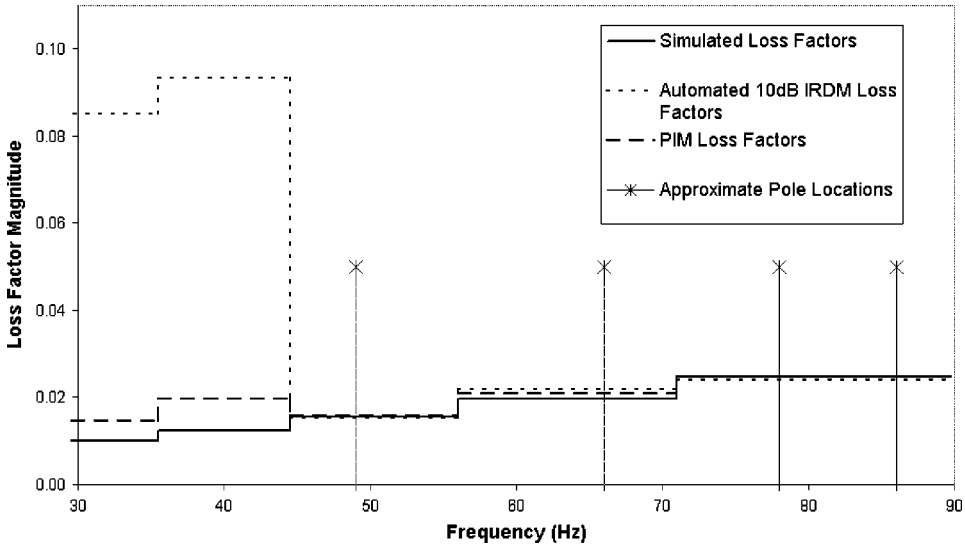


FIG. 7. Comparison of PIM, IRDM, and simulated loss factors.

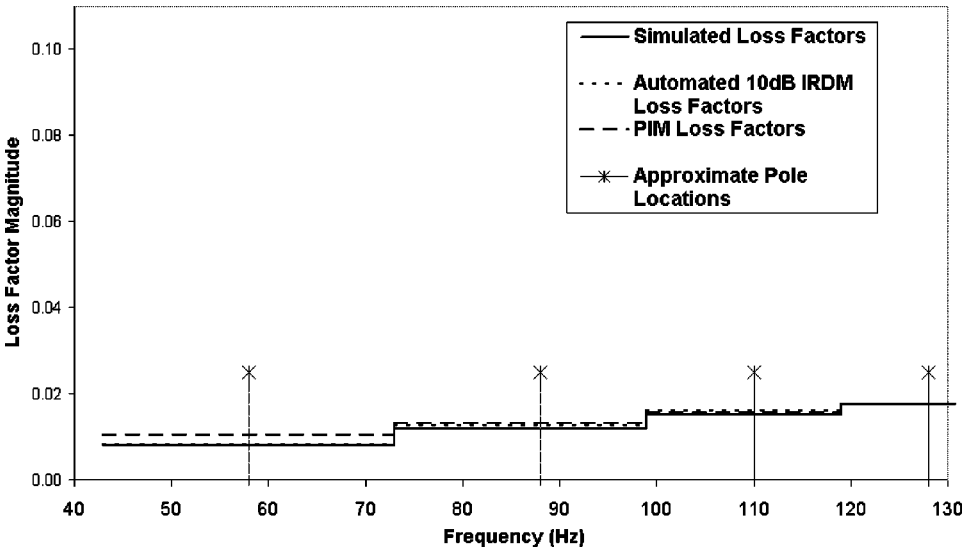


FIG. 8. The loss factor of a 4-degrees-of-freedom system.

E. Effects of frequency bandwidth

A new lumped mass–spring–damper system, with the modal resonances more equally spaced, was created to investigate whether changing the width of the frequency bands used to average the loss factors has any effect on the loss factor estimates for both methods. Figure 8 shows the simulated loss factors and the estimated loss factors of the new system from both methods. The frequency bandwidths are dictated by the distance between the modal resonances—the cutoff frequencies for each band reside halfway between the modal resonances.

Two additional lumped-mass systems were created with identical mass and damping values as the system whose loss factors are shown in Fig. 8. The only difference between the three systems was the stiffness matrices. The matrices were multiplied by a different constant for each system, which in turn changed the distance between the modal resonances and, effectively, the frequency bandwidths. Approximately 10% and 20% increases in frequency bandwidths over the original were created—this caused an indiscernible change in the loss factor estimates. The lack of change may be attributed to the pole locations within the frequency bands since they were basically in the center of the averaging bands. The difference between loss factor estimates for the systems found using the IRDM are attributed to small fluctuations in the initial 10 dB of the decay slope.

F. Effects of increased damping

Since the experimental utilization of the two methods will be performed on systems with varying damping values, it is important to investigate the error in the estimation techniques due to change in damping values. The same system used for the frequency bandwidth comparison was used here, but instead of changing the stiffness matrix, the damping matrix was manipulated. Table I shows the estimated loss factors for systems with different damping levels. The damping matrix was increased by 2, 4, and 8 times the initial values. Except for the case of damping matrix multiplication factor of 8, both methods estimate the loss factors relatively close to the simulated loss factors. At some frequency bands the correlation actually improves as the damping is increased. The IRDM maintains approximately the same amount of error for the first three levels of damping.

As with different frequency bandwidths, the IRDM estimation errors for the lowest three levels of damping could be attributed to small variations in the first 10 dB of the decay slope instead of the averaging scheme used or the increase in damping. The estimated damping values from the IRDM appear to have a ceiling limit for the level of damping that can exist in a system to obtain acceptable results. The highest level of damping used in this comparison is significantly greater than that exhibited in most linear systems. The error levels with the PIM is constant regardless of damping level—this indicates that the estimations from PIM are independent of the amount of damping in the system.

TABLE I. Comparison of different levels of damping for the PIM and the IRDM.

Simulated LF	PIM LF	IRDM LF
	Initial loss factors	
0.0040	0.0052	0.0046
0.0059	0.0066	0.0064
0.0075	0.0078	0.0081
0.0087	0.0088	0.0089
	Damping matrix multiplied by 2	
0.0080	0.0104	0.0081
0.0118	0.0132	0.0125
0.0150	0.0155	0.0160
0.0175	0.0176	0.0175
	Damping matrix multiplied by 4	
0.0161	0.0209	0.0160
0.0236	0.0263	0.0255
0.0300	0.0311	0.0294
0.0349	0.0351	0.0267
	Damping matrix multiplied by 8	
0.0322	0.0418	0.0322
0.0473	0.0528	0.0467
0.0599	0.0624	0.0455
0.0698	0.0700	0.0384

TABLE II. Percent error caused by changing the location of modes within frequency bands.

Centered modes		Shifted frequency bands 50% higher		Shifted frequency bands 50% lower	
PIM % dif. from simulated LF	IRDM % dif. from simulated LF	PIM % dif. from simulated LF	IRDM % dif. from simulated LF	PIM % dif. from simulated LF	IRDM % dif. from simulated LF
29.72	1.32	22.10	3.49	39.56	32.61
11.32	5.93	4.82	6.95	17.14	24.32
3.65	6.51	2.60	0.21	3.31	9.32
0.66	0.34	−1.00	−11.99	1.74	−6.64

G. Effects of location of modal resonance within a frequency band

In practice, the location of a modal resonance will rarely be at the center of a frequency band. It is useful to understand the amount of error associated by changing the location of the modal resonance within a frequency band. The same 4-D.O.F. system as used in earlier simulations was used, but this time, the upper and lower frequencies of the bands used to average the loss factors were varied such that the modal resonances were effectively shifted higher or lower in the frequency bands. Table II shows that the IRDM damping estimates vary greater from the simulated values when the modes are shifted either positively or negatively from the original position. For the PIM, it appears as though the error increases as the frequency bands are shifted down and decreases when the frequency band limits are shifted up. Upon closer inspection of the narrow-band loss factor estimations from the PIM, the error is reduced when the frequency is increased—at least for the system used in this comparison. This can be seen in Fig. 9, where the simulated loss factors and the narrow-band PIM loss factors are plotted against each other. Error decreases as frequency increases, but this particular estimation pattern is a property of the system alone and not indicative of the errors associated with the PIM. In other words, other analytical systems used throughout this investigation have shown distinctly different PIM narrow-band loss factor patterns.

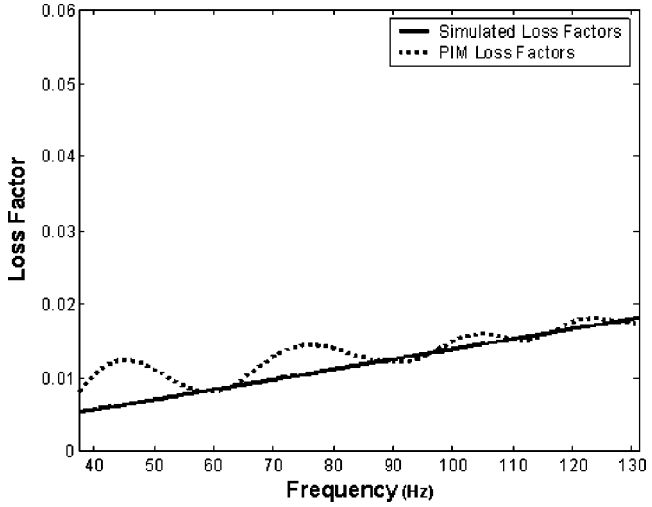


FIG. 9. Comparison between simulated loss factors and PIM loss factors in narrow-band frequencies.

H. Effect of increasing the number of modes

The ideal systems used for various comparisons shown earlier were all relatively simple systems with only a few modes. Many real-world systems are not as simplistic and the differences caused by the increase in the total number of modes need to be investigated. Ideally the effects of adding more lumped masses to the system of interest should only increase the number of modes, without having any adverse effects on the damping estimation errors. As shown in Fig. 10, both methods estimate the loss factors relatively well, even as the mode count is increased to a 9-lumped mass system. As shown previously, having no modal resonances within a frequency band causes both estimation methods to yield large errors. The PIM gives better estimates than the IRDM when no modal resonances are present in a given frequency band. The IRDM estimates loss factors for the upper two frequency bands that are much lower than the simulated values—this is thought to be due to having a modal resonance straddle the cutoff frequency exactly. The increase in modes is not thought to be the factor because Figs. 5–7 show multimode bands with very accurate loss factor estimations via the IRDM.

III. EXPERIMENTAL VALIDATION

A relatively simple system was chosen for the experimental validation of the similarities between the PIM and the IRDM. A steel plate was used as the system because of its low material damping that could be altered easily through additive damping treatments. Additionally, a plate has higher modal density when compared against other simple structures (i.e., bars, etc.). The damping of the plate, with dimensions $205\times200\times1.75$ mm $(8.1\times7.9\times0.069$ in), was used in this study with and without damping treatment glued to one side. The damping treatment and glue had a combined surface density of 5.43 kg/m^2. The plate was suspended with bungee cords to create free–free boundary conditions as shown in Fig. 11.

To eliminate the effects of mass loading due to an accelerometer, a Polytec laser, model OFV-302, in conjunction with a laser controller, model OFV-3000, was used to mea-

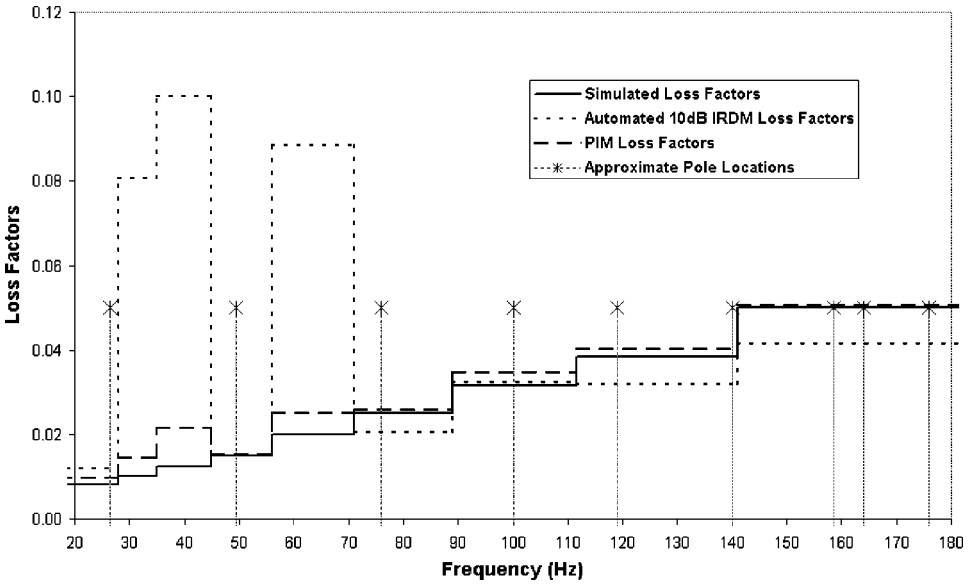


FIG. 10. Estimated loss factor of a 9-degrees-of-freedom system.

sure the response of the plate. Reflective materials were applied to the steel plate to enhance the response measurement capability of the laser. Sections of reflective tape, 3M model 510-10X, of approximately 4 mm square were used to create a 33-mm grid over the entire surface of the plate. This grid spacing was chosen to create suficient points, 36 in all, on the surface to reduce errors introduced through approximating the surface integral with a summation. As the number of points increases, the loss factor measurements asymptotically approach a constant value.

The plate was excited from 20 Hz to 8 kHz via a Labworks Inc. model ET-132-2 electrodynamic shaker and model PA-118 power amplifier. A model 288D01, PCB impedance head, connected to the shaker with a stinger and attached to the back of the plate, was used to measure the input force and acceleration. The shaker input signal was generated, and the impedance head and laser responses were measured, using an LMS SCADA-III front end. Data were collected and analyzed using LMS CADA-X 3.5C.

Figure 12 shows the FRFs of the plate with and without damping treatment. As expected, the damping treatment significantly alters the modes of the plate, especially at the higher frequencies. Characteristic of a lightly damped structure, the bare steel plate has sharper peaks with higher levels than the plate with the damping treatment applied.

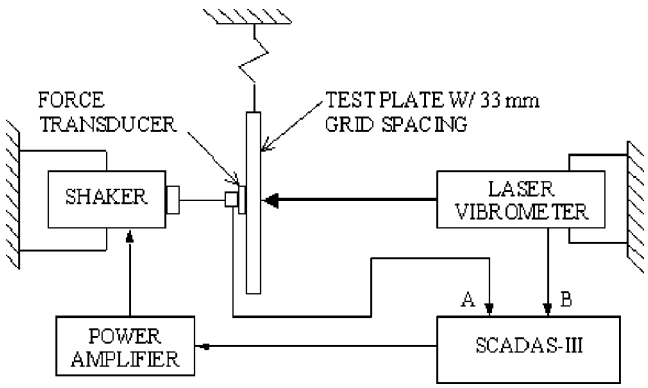


FIG. 11. Schematic of the experimental setup.

Since the loss factors were frequency-band averaged and since the analytical comparison between the PIM and the IRDM showed that at least one mode is needed in each frequency band to obtain accurate results, the FRFs were broken up into one-third octave bands. Figure 13 shows the driving point FRFs from the two plate arrangements in the range from 22 to 1130 Hz. Modes are not consistently present in the frequency bands for both arrangements until the 630-Hz one-third-octave band. The fundamental natural frequency for the plate is above 100 Hz so no loss factors can be estimated with either method below this frequency.

The PIM and the IRDM were used to estimate the loss factors for both plate arrangements. The PIM loss factors were found across narrow frequency bands (the frequency resolution of the FRFs) and then averaged over one-third-octave bands for comparison with the IRDM loss factors. The IRDM loss factors were obtained by filtering the FRFs into one-third-octave bands and then utilizing an inverse Fourier transform to obtain the impulse response decays.

Figure 14 shows the loss factors of the plane plate and

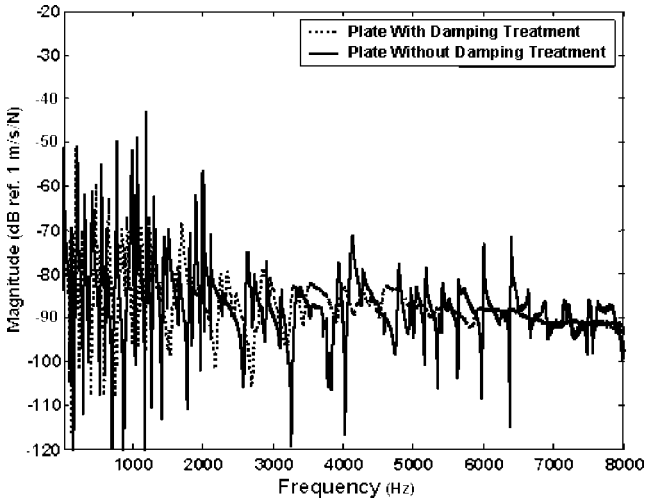


FIG. 12. Comparison of driving point FRF of the steel plate with and without damping treatment.

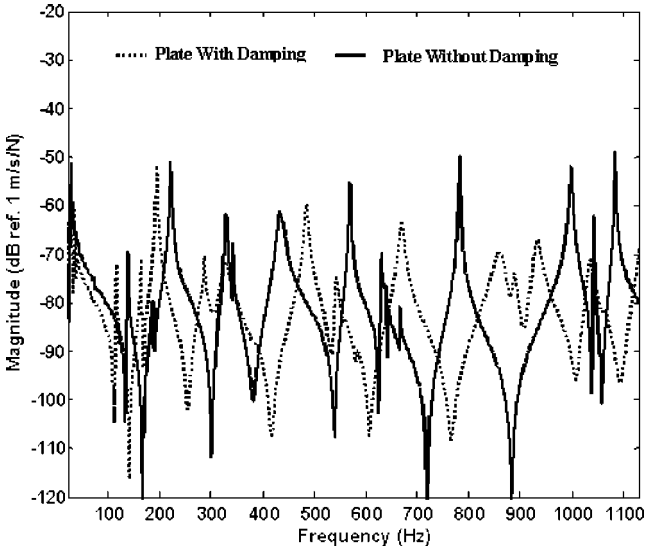


FIG. 13. One-third-octave band separation of FRFs of plate with and without damping.

the damped plate using both the IRDM and the PIM. As was shown analytically, the frequency bands with no modal resonances present and the bands below the fundamental natural frequency appear to have extremely large loss factors when compared with neighboring frequency band loss factors. Also, it appears as though the PIM may do a better job of estimating the loss factors than the IRDM for frequency bands with no modes present since the values are not as extreme. It can be recalled that the same observation was made with the results from the analytical model. For frequency bands with modes present, the loss factor values for the bare plate show less variance between the two techniques than the damped plate.

IV. CONCLUSIONS

Using analytical models of mass/spring/damper systems, the PIM and IRDM estimated loss factors were compared against simulated loss factors. The comparisons were done using frequency-band averaged values for differing frequency bandwidths; different averaging schemes for the IRDM; different user-defined and predetermined initial decay slope identification techniques; different mode spacings; and different damping levels. It was shown that as long as one modal resonance was present in each frequency band, both methods yield accurate results. When the simulated damping values were increased to very large values, the IRDM predicted loss factors were significantly lower than both the PIM and the simulated loss factors.

The results obtained from the analytical study were examined experimentally. Both the PIM and the IRDM were used to estimate the loss factors of a steel plate with and without damping treatments applied. As with the analytical study, both methods returned similar results but variance between the two existed. The PIM is based on a comparison of the dissipated energy of a system to its maximum strain energy. As long as each measurement location is properly weighted by the corresponding mass, little or no bias errors occur in this method. Some errors may be introduced through the measurement technique because this method is sensitive to power input measurement accuracy, but it is fundamentally unbiased at the natural frequencies of well-defined modes or when the loss factors are frequency-band averaged over many modes. Decay methods of finding the damping of a structure are more attractive as long as the initial decay slope of structure can be obtained once the input power is shut off. The impulse response decay method (IRDM) is used to alleviate the difficulties of finding the starting point of the initial decay. The IRDM will be looked at as a viable alternative to the PIM in simplifying damping estimation acquisitions.

The number of measurement points used with the PIM to obtain accurate, consistent loss factors is considered high. Since this study was limited to a comparison between the two methods, the same number of points was used for both techniques. The number of response locations needed to obtain accurate results from the IRDM was not considered. If one could reduce the number of measurement points needed

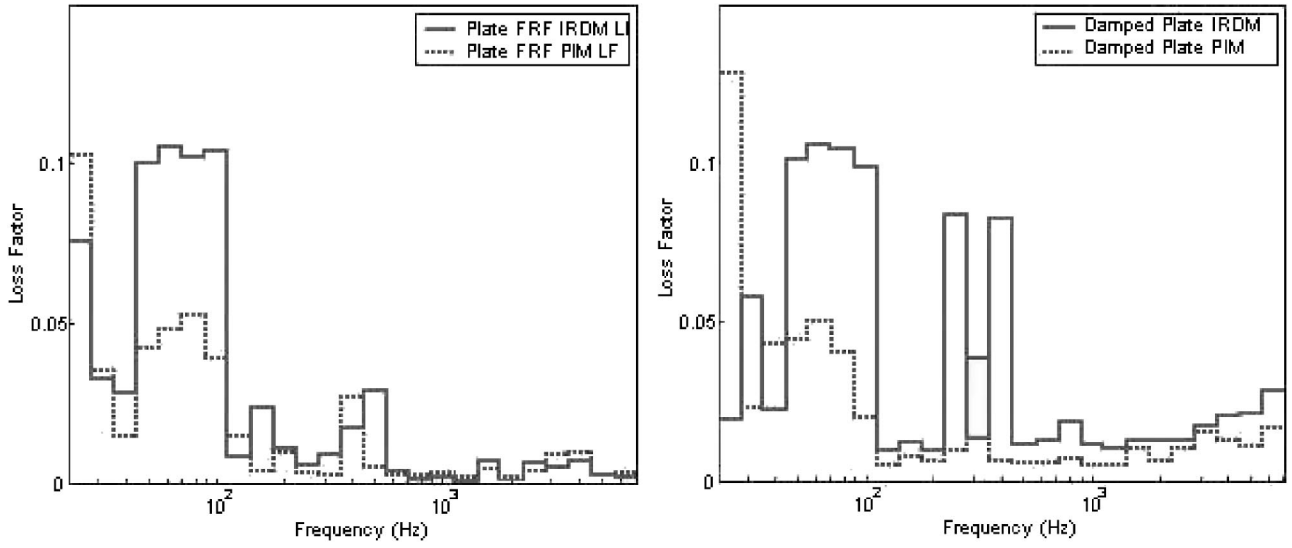


FIG. 14. Plate loss factors for damped (left) and bare (right) steel plate.

with the IRDM, it could gain a substantial advantage over the PIM in its usefulness for determining loss factors of large systems.

[1] R. H. Lyon and R. G. DeJong, *Theory and Application of Statistical Energy Analysis*, 2nd ed. (Butterworth-Heineman, Newton, MA, 1995).
[2] M. P. Norton and D. Karczub, *Fundamentals of Noise and Vibration Analysis for Engineers*, 2nd ed. (Cambridge University Press, Cambridge, 1989).
[3] F. H. Chu and B. P. Wang, ''Experimental Determination of Damping in Materials and Structures,'' Damping Application for Vibration Control, edited by P. J. Torvik, ASME Winter Annual Meeing, Chicago, pp. 113–122 (1980).
[4] D. D. Rife and J. Vanderkooy, ''Transfer–function measurement with maximum-length-sequences,'' J. Audio Eng. Soc. **37**(6), 419–443 (1989).
[5] M. Vorlander and M. Kob, ''Practical aspects of MLS measurements in building acoustics,'' Appl. Acoust. **52**(3/4), 239–258 (1997).
[6] M. R. Schroeder, ''New method of measuring reverberation time,'' J. Acoust. Soc. Am. **37**(3), 409–412 (1965).
[7] W. T. Chu, ''Comparison of reverberation measurements using Schroeder's impulse method and decay-curve averaging method,'' J. Acoust. Soc. Am. **63**(5), 1444–1450 (1978).
[8] N. Xiang, ''Evaluation of reverberation times using a nonlinear regression approach,'' J. Acoust. Soc. Am. **98**(4), 2112–2121 (1995).
[9] M. Carfagni and M. Pierini, ''Determining the loss factors by the power input method. I. Numerical investigation,'' J. Vibr. Acoust. **121**, 417–421 (1997).
[10] D. A. Bies and S. Hamid, ''*In situ* determination of loss and coupling loss factors by the power injection method,'' J. Sound Vib. **70**(2), 187–204 (1980).
[11] M. Carfagni, P. Citti, Paolo, and M. Pierini, ''Determining loss factors using the power input method with shaker excitation.'' 1998 Proceedings of the XVIth IMAC, CA, 585–590 (1998).
[12] S. S. Rao, *Mechanical Vibrations*, 3rd ed. (Addison-Wesley, Reading, MA, 1995).

Perceptually motivated time-frequency analysis[a)]

Jonathan J. O'Donovan[b)] and Dermot J. Furlong
Music and Media Technologies, Department of Electronic and Electrical Engineering, Trinity College, Dublin, Ireland

(Received 15 February 2003; revised 1 October 2004; accepted 4 October 2004)

This paper describes the design of a bilinear time-frequency distribution which is a joint model of temporal and spectral masking. The distribution is used to generate temporally evolving excitation patterns of nonstationary signals and systems and is conceived as a tool for acousticians and engineers for perceptual time-frequency analysis. Distribution time and frequency resolutions are controlled by a separable kernel consisting of a set of low-pass time and frequency smoothing windows. These windows are designed by adapting existing psychoacoustic models of auditory resolution, rather than using mathematical window functions. Cross-term interference and windowing clutter are highly suppressed for the distribution, ensuring resolution accuracy over a dynamic range sufficient to encompass that of the auditory system (in excess of 100 dB). Application to the analysis of a synthetic and two real signals are included to demonstrate the approach. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1831283]

PACS numbers: 43.60.Hj, 43.66.Ba, 43.75.Yy, 43.72.Ar, 43.71.An [JCB] Pages: 250–262

I. INTRODUCTION

The analysis of nonstationary signals is the domain of time-frequency (TF) analysis. TF analysis is used to express signal energy in terms of a three-dimensional joint function of time, frequency, and intensity. Since there are many approaches to TF analysis available, a particular TF analysis scheme is generally chosen based on the class of signal being studied. Examples of the wide ranging application that TF analysis has seen in the past are biological signal analysis, machine monitoring, seismic signal analysis, and sonar and radar analysis (Various, 1996; Boashash, 2003). TF analysis has also been successfully applied to many areas in audio and acoustics including loudspeaker design (Janse and Kaizer, 1983) and music and speech analysis (Pielemeier and Wakefield, 1995; Cooke *et al.*, 1993). In order to resolve fine signal detail and to accurately estimate the physical properties of the audio signal under investigation, audio TF analysis typically aims to achieve as high a TF resolution as possible.[1] However, such resolution is not a reflection of auditory processing or perception. In many areas of audio research, the ability to investigate the perceptual features of a particular signal or system would also be highly desirable (Gerzon, 1990), for example in the evaluation of audio equalization filters or the analysis of musical instrument timbre.

Human hearing is subject to many nonlinear effects due to its complex physiology. One of the most important features of the auditory system studied in the psychoacoustic literature, which has a marked effect on our perception of audio, is that of auditory masking (Moore, 1997). Masking is classified as either temporal or spectral; both may be modeled in terms of auditory temporal and spectral resolutions. These auditory resolutions are dependent on several factors including frequency and signal level.

This paper describes a TF signal or system analysis tool, known as the EarWig distribution (EWD), which incorporates joint models of both temporal and spectral masking. The EWD is developed using the framework of Cohen's class (CC) (Cohen, 1995) of bilinear TF distributions. Members of this class consist of a combination of the Wigner distribution (WD) (Cohen, 1995) and a smoothing kernel, which determines the properties of the distribution, including its time and frequency resolutions. In this work, the EWD has a separable kernel which is frequency dependent and is designed to reproduce the resolutions of psychoacoustic models of temporal and spectral masking. The use of a separable kernel allows independent control of EWD time and frequency smoothing, which is not possible in linear time frequency analysis (Cohen, 1995). EWD spectral smoothing is achieved by designing frequency-dependent, frequency smoothing windows based on the gammatone auditory filter function (Moore, 1997). Similarly, EWD auditory temporal smoothing is achieved by employing frequency-dependent temporal smoothing windows which are based on a frequency-dependent version of the temporal window model of temporal masking (Plack and Moore, 1990). Significantly, adopting psychoacoustic models in this manner is a departure from past approaches to kernel design for audio analysis, which have typically used mathematical window functions such as the Hamming or Hanning windows for high resolution.

Despite the advantages of bilinear TF methods over linear methods (Cohen, 1995) and their potential for audio applications (Gerzon, 1990), difficulties associated with bilinear TF analysis such as cross-term interference, limited dynamic range, and computational complexity have led some researchers to view bilinear TF methods as being unsuited to auditory modeling purposes (Agerkvist, 1996). To date there has been no successful attempt at auditory modeling using

[a)]Portions of this work were presented at the 136th Meeting of the Acoustical Society of America, Norfolk, VA, October, 1998 (O'Donovan and Furlong, 1998).

[b)]Corresponding author; electronic mail: jodonova@tcd.ie

bilinear TF methods. The significance of the EWD is that it overcomes these difficulties and therefore resolves doubts concerning the applicability of bilinear TF methods to auditory modeling. The EWD demonstrates that the flexibility of bilinear TF analysis may be exploited in the design of a smoothing kernel modeling both temporal and spectral masking.

In Sec. II, the rationale of this work is presented, and the difficulties associated with using bilinear TFA for auditory modeling are described. Cohen's class of TF analysis techniques are well documented in the literature (Boashash, 2003, Chaps. 1–5, and the references therein), and therefore Sec. III A gives only a brief summary of this class and describes the discrete smoothed pseudo-Wigner distribution (Sec. III B).

The EWD distribution algorithm is described in Sec. IV A. Sections IV B and IV C detail the EWD auditory kernel design approach, and describe briefly the psychoacoustic models upon which this kernel design is based. Finally, Sec. V presents the EWD of a complex tone and its marginals to demonstrate the effect of jointly modeling temporal and spectral masking. A comparison with the constant-Q modal distribution (Pielemeier and Wakefield, 1995) is also made for two musical signals to highlight some features of EWD analysis.

II. MASKING AND TIME-FREQUENCY ANALYSIS

Much research effort in the psychoacoustic literature has been devoted to the study of masking, and models of auditory temporal resolution and auditory spectral resolution (auditory filter models) are well established. Examples include the *roex* (rounded exponential) (Moore, 1997) and gammachirp (Irino and Patterson, 1997) auditory filter shapes and the temporal window model of temporal masking (Moore *et al.* 1988; Plack and Oxenham, 1998). Models of auditory processing incorporating both temporal and spectral masking have been developed for the investigation of the auditory system (rather than signal analysis), including the auditory image model (Patterson *et al.*, 1995) and the spectro-temporal excitation pattern (Plack, 1990). Joint models of temporal and spectral masking are also used in audio compression scheme research. In compression research, efficiency and robustness are important considerations and in general an exact match to psychoacoustic auditory models, though desirable, is not necessary.

There are few joint models of temporal and spectral masking intended specifically for perceptually based acoustic signal or system analysis purposes. Furthermore, few attempts at using bilinear TF analysis techniques to simultaneously model temporal and spectral masking have been made. Some attempts using linear time-frequency approaches such as the short-time Fourier transform (STFT)[2] (Cooke *et al.*, 1993; Jeong and Ih, 1999) or wavelet transform (Agerkvist, 1996) do exist. However, these methods can only achieve an approximation to auditory resolution because the temporal and spectral resolutions of such linear methods cannot be independently set (Cohen, 1995).

Bilinear signal analysis techniques, such as the smoothed pseudo-Wigner distribution (SPWD) (Cohen, 1995), do allow the independence of temporal and spectral smoothing required to model masking and have been successfully applied to a wide range of audio analysis tasks (Various, 1996). Despite this, these methods have not been extensively exploited for auditory modeling in the past, mainly because of the presence of undesirable cross-term interference and the computational complexities associated with this form of TF analysis (Agerkvist, 1996). Cross-term interference, discussed in Sec. III A, makes distribution interpretation difficult and also limits useful distribution dynamic range. Although much research effort has been devoted to the design of smoothing kernels optimized for cross-term suppression (Various, 1996), the amount of suppression achieved has not been sufficient to provide a dynamic range encompassing that of the auditory system (in excess of 100 dB for much of the audible spectrum). A consequence of the limited dynamic range available in most forms of bilinear TF analysis is that intensity can only be displayed in linear rather than logarithmic units.

A further difficulty with using bilinear methods for auditory modeling has been the computational complexity of implementing a frequency-dependent smoothing kernel. Such a kernel is necessary, since both the temporal and spectral resolutions of the auditory system are frequency dependent. Computational schemes for Cohen's class (CC) distributions have typically utilized the FFT algorithm for efficiency, resulting in a uniform bin spacing, rather than nonlinear frequency scales, such as the equivalent rectangular bandwidth (ERB) or Bark scale (Moore, 1997) used in psychoacoustics. Furthermore, it is also only recently that techniques for the analysis of long signals have been developed, due to the memory and computation requirements of the smoothed Wigner distribution (Pielemeier and Wakefield, 1995). The current approach resolves these various issues.

III. BILINEAR TIME-FREQUENCY ANALYSIS

A. Cohen's class

Cohen's work demonstrated that a wide variety of TF techniques proposed after Wigner were all related via linear transformation and could therefore be grouped into a single class of TF distributions. This provided a new unified framework within which the properties and relative merits of various TF techniques could be assessed. Members of Cohen's class (CC) may be expressed in the time-frequency plane as the convolution of the Wigner distribution (WD) and a convolution kernel as follows (Cohen, 1995):

$$C(t,\omega)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W(u,\xi)\Phi(t-u,\omega-\xi)du\,d\xi. \qquad (1)$$

Here, $W(t,\omega)$ is the Wigner distribution (WD) expressed in the time-frequency domain (u and ξ being convolution variables). $\Phi(t,\omega)$ is the smoothing kernel. The (unsmoothed) WD satisfies many desirable TF properties including reality, evenness, energy conservation, and time and frequency shift covariance (Claasen and Mecklenbräuker, 1980a). The effect of the smoothing kernel is to modify these properties, producing the properties of the resulting CC distribution, includ-

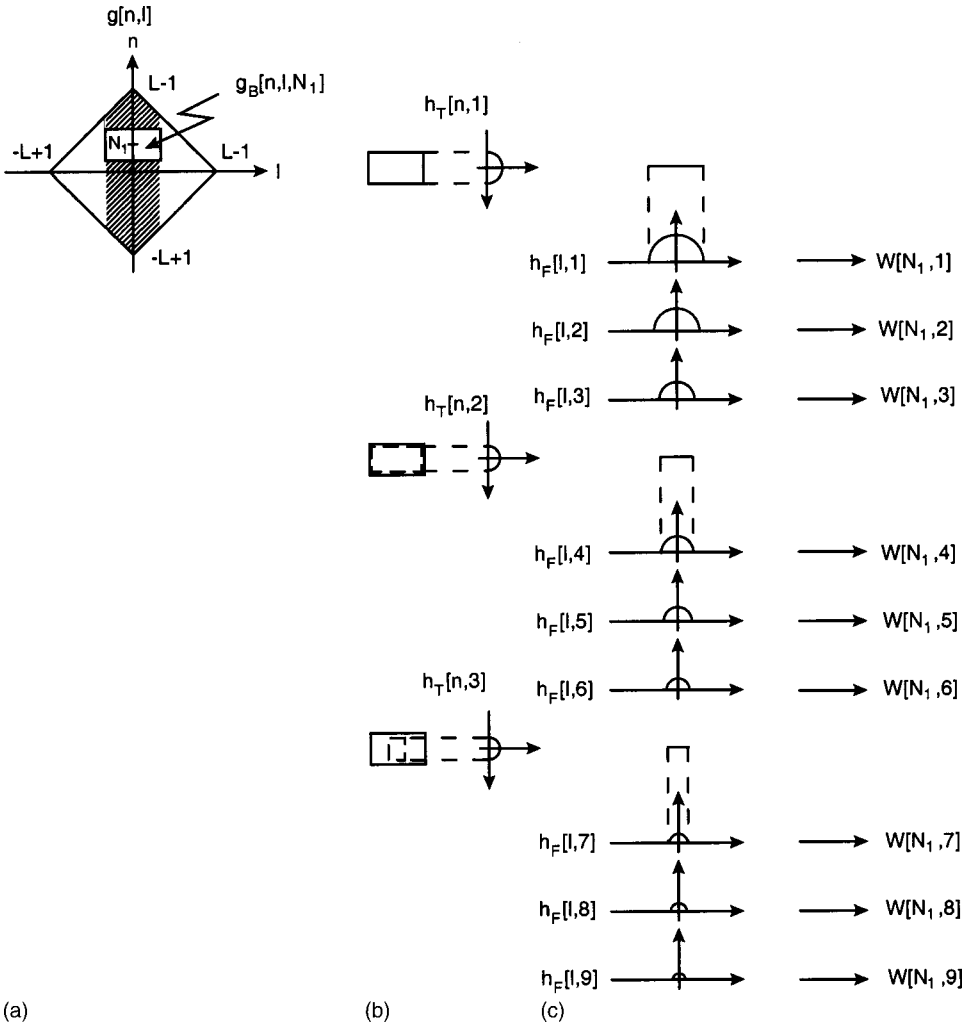


FIG. 1. Operations required to calculate an EWD frequency slice with frequency-dependent time and frequency resolution: (a) TCF block calculation, (b) TCF time smoothing, (c) TCF frequency smoothing.

ing its time and frequency resolutions. The Wigner distribution of a signal $s(t)$ is

$$W(t,\omega)=\int_{-\infty}^{\infty} s^*\left(t-\frac{\tau}{2}\right)s\left(t+\frac{\tau}{2}\right)e^{-j\tau\omega}d\tau, \tag{2}$$

where $s^*(t-\tau/2)s(t+\tau/2)$ is the *temporal correlation function* (TCF). Here, the WD is expressed in the time-relative time domain as the Fourier transform with respect to relative time τ of the TCF. The WD is classed as bilinear because the TCF depends on a product of signal $s(t)$ with itself.

The main advantage that distributions belonging to CC offer over linear TF distributions such as the STFT or wavelet transform is independent control of time and frequency resolution. This can be seen by re-expressing $\Phi(t,\omega)$ as a product kernel (Martin and Flandrin, 1985):

$$\Phi(t,\omega)=h_t(t)\times H_\omega(\omega). \tag{3}$$

Here, $h_t(t)$ controls the temporal resolution of the distribution and $H_\omega(\omega)$ controls the spectral resolution. The freedom to independently control resolution is offset however by the presence of *cross-term interference* (Hlawatsch and Flandrin, 1997) which complicates distribution interpretation. Because bilinear TF distributions are expressed in terms of signal products, bilinear TF distributions are composed of both *auto-terms* and *cross-terms*. Auto-terms are located in TF where one would expect a signal's energy to exist. Any two auto-terms which are separated in the TF plane also interfere to create a cross-term at their geometrical midpoint. Cross-terms therefore appear at locations in the TF plane where no signal energy exists. Cross-terms oscillate at a frequency proportional to the distance between the auto-terms and may be suppressed by smoothing via low-pass smoothing kernels. Therefore, for CC distributions there is a trade-off between cross-term suppression and TF resolution. Typical TF distribution research involves the design of smoothing kernels which provide a good compromise between cross-term suppression and TF resolution.

B. The discrete smoothed pseudo-Wigner distribution

A popular member of Cohen's class for audio analysis (Janse and Kaizer, 1983) is the smoothed pseudo-Wigner distribution (SPWD), which is a time and frequency smoothed WD. The SPWD is the basis of the modal distribution (Pielemeier and Wakefield, 1995) and EWD. If the smoothed pseudo-Wigner distribution (SPWD) is to be computed on a digital computer, a discrete form of the distribution is required. The derivation of the *discrete* smoothed pseudo-Wigner distribution (DSPWD) is presented in detail in the literature (Claasen and Mecklenbräuker, 1980b; Peyrin and

Prost, 1986; Richman *et al.*, 1998; O'Neill *et al.*, 1999). Specifically, for a discrete sequence $s[n]$, the discrete smoothed pseudo-WD is:

$$\mathrm{DSPWD}[n,k]=\sum_{l=-P+1}^{P-1} h_F[l] \sum_{m=-Q+1}^{Q-1} h_T[m] \times g[n-m,l]e^{-j2lk\pi/M}, \quad (4)$$

$$g[n,l]=s[n+l]s^*[n-l]. \quad (5)$$

Here, $s[n]$ is the discrete sequence representing the continuous signal $s(t)$ sampled at regular intervals nT, where T is the sampling interval. Because relative time τ in Eq. (2) is replaced by $2lT$ here, the sampling frequency, f_s, of the discrete time signal, $s(nT)$, must exceed *twice* the Nyquist rate. In other words $T=1/f_s$, where $f_s \geqslant 4f_c$, f_c being the highest signal frequency present in the continuous signal $s(t)$.[3] The discrete WD is a function of discrete time n and discrete frequency $k\pi/M$ (abbreviated to k here), and is periodic in frequency with period π. Apart from this periodicity, the properties of the discrete smoothed pseudo-WD (DSPWD) are similar to those of the continuous smoothed pseudo-WD (SPWD) (Claasen and Mecklenbräuker, 1980a; Claasen and Mecklenbräuker, 1980b).

The kernel is separable, being a combination of finite length frequency smoothing window, $h_F[l]$, of length $2P-1$, and a finite length time smoothing window, $h_T[n]$, of length $2Q-1$. The temporal correlation function (TCF) $g[n,l]$ is smoothed in the time direction by the time smoothing window $h_T[n]$ prior to FFT calculation. The multiplication in l of the frequency smoothing window $h_F[l]$ with the TCF becomes a convolution in discrete frequency k after transformation via the DFT. Assuming that the time and frequency smoothing windows are low pass, their effect is to reduce cross-term interference oscillating parallel to the time and frequency axes, respectively. Note that the temporal convolution of the smoothing window $h_T[n]$ is unaffected by the Fourier transform.

IV. THE EARWIG DISTRIBUTION

A. EWD algorithm

In Pielemeier and Wakefield (1995), a member of CC designed for high resolution audio signal analysis, called the *modal distribution*, is described. By employing a stepwise approach to TCF calculation and performing time smoothing prior to FFT calculation, the modal distribution computation time is reduced from being order CN^3 to $CN\log N$, where N is the signal length. As an extension to the modal distribution, Pielemeier and Wakefield also demonstrate that frequency-dependent frequency resolution may be obtained by varying the frequency smoothing window length in proportion to frequency. The result is the *constant-Q modal distribution*. The following section describes how the constant-Q modal distribution algorithm may be extended to also include variable temporal resolution.

Figure 1 portrays the steps necessary to compute the EWD at a particular time $n=N_1$ for an example involving nine individual frequency channels, C. The signal length is $2L-1$ and the diamond-shaped region indicates all possible nonzero TCF values. The shaded area indicates the TCF values required for EWD calculation, which is approximately proportional to the signal length times the length of the longest frequency smoothing window (the signal is typically much longer than the width of the longest frequency smoothing window). The windows $h_F[l,C]$ are the frequency smoothing window for channels 1–9. The windows $h_T[n,B]$ are the temporal windows corresponding to three frequency bands which encompass all 9 channels. In this example, temporal windows $h_T[n,1]$, $h_T[n,2]$, and $h_T[n,3]$ cover channels 1–3, 4–6, and 7–9, respectively. Values EWD$[N_1,1]$ to EWD$[N_1,9]$ represent the final computed values of the EWD for channels 1–9 at time $n=N_1$. Because auditory temporal resolution changes relatively slowly with frequency in comparison to auditory frequency resolution (Sec. IV C), the same temporal smoothing can be performed once for a range of channels. In this work, it was found that seven temporal window shapes were sufficient to cover the nominal audible frequency range 20 Hz–20 kHz. As auditory frequency resolution overlaps strongly, generally 100–150 frequency channels are sufficient to encompass the entire audible frequency range. Therefore, reducing the number of temporal smoothing stages typically results in computational savings of a factor greater than 10 for time smoothing, depending on the analysis frequency range.

The EWD algorithm proceeds as follows. A temporal correlation function (TCF) *block*, $g_B[n,l,N_1]$, of width equal to the width of the longest frequency smoothing window $h_F[l,1]$ and length equal to the length of the longest time smoothing window $h_T[n,1]$, centered around $[n,1]=[N_1,0]$ is first calculated [step (a)]. The length of both spectral and temporal windows is greatest for low frequencies. This TCF block is then smoothed for each of the three temporal windows (TCF time smoothing). For the lowest frequency temporal window (top), each column of the TCF block is first multiplied by $h_T[n,1]$ and summed in n giving a TCF strip [step (b)]. This strip is then multiplied by each of the frequency smoothing windows in the frequency band of this temporal window ($h_F[l,1]$, $h_F[l,2]$, and $h_F[l,3]$). A single frequency DFT calculation is performed after each frequency smoothing operation, yielding the value of EWD$[N_1,1]$, EWD$[N_1,2]$ and EWD$[N_1,3]$; this is [step (c)]. Using a single value DFT calculation allows the frequency channel spacing to be chosen arbitrarily, for example according to the logarithmic or equivalent rectangular bandwidth (ERB) scale.

Similar calculations are performed for the other two temporal window stages yielding EWD$[N_1,4]$ to EWD$[N_1,9]$. The overall algorithm is an extension of the constant-Q modal distribution algorithm, which employs a single temporal smoothing stage. For the following time slice, $n=N_2$, a new TCF block $g[n,l,N_2]$ is computed and the entire process is repeated. Because time smoothing is performed before DFT transformation, the distribution need only be computed in time at a rate sufficient to satisfy the sampling requirements of the (low-pass) temporal smoothing windows, rather than at the signal sampling frequency, without any loss of distribution detail. This results in very significant computational savings because signal bandwidth is

TABLE I. Values of averaged parameters T_p, T_s, w and a for frequency-dependent temporal window model.

Frequency (Hz)	T_p (ms)	T_s (ms)	w (dB)	a
300	12.7	70.7	−58.3	1.6
900	9.3	68.3	−47.0	1.8
2700	8.1	46.0	−39.0	2.1
8100	7.5	48.0	−39.7	1.9

generally far greater than temporal window bandwidth. For example, if the signal sampling rate is 44.1 kHz, and the distribution is calculated every 5 ms, which is sufficient for the temporal window, a computational saving of 220.5 to 1 is achieved. Overlap between adjacent TCF blocks may also be exploited to reduce TCF computation costs. Furthermore, since the TCF is computed incrementally, large memory capacity is not required and signal length is no longer a limiting factor.

B. Frequency smoothing window design

In O'Donovan (2000), four auditory filter shapes were evaluated for their suitability in the task of designing an EWD smoothing kernel, namely the roex (Moore and Glasberg, 1987), gammatone (Moore, 1997), gammachirp (Irino and Patterson, 1997), and all-pole gammatone filter (Slaney, 1993). The most suitable was deemed to be the gammatone filter. The gammatone filter has been used extensively as an auditory filter model (Cooke *et al.*, 1993; Jeong and Ih, 1999; Cooke, 1993). Unlike the roex, all-pole gammatone, and gammachirp filterbanks, its simple time domain response is readily demodulated to dc, which is useful for EWD smoothing kernel design, as shown in the following. Although the gammatone filter is level independent, this is not a serious limitation for the purposes of excitation pattern calculation because differences between level dependent and independent filter models only become significant for narrowband signals at high signal levels, and not for typical broadband musical signals such as speech and music (Plack, 1990). If required, level dependency may be readily incorporated in the design of the EWD smoothing kernel.

The bank of low-pass, frequency-dependent frequency smoothing windows $h_F[l,C]$ was derived from a baseband version of the gammatone auditory filter shape. The expression for the gammatone filter is given in Eq. (6). The de-

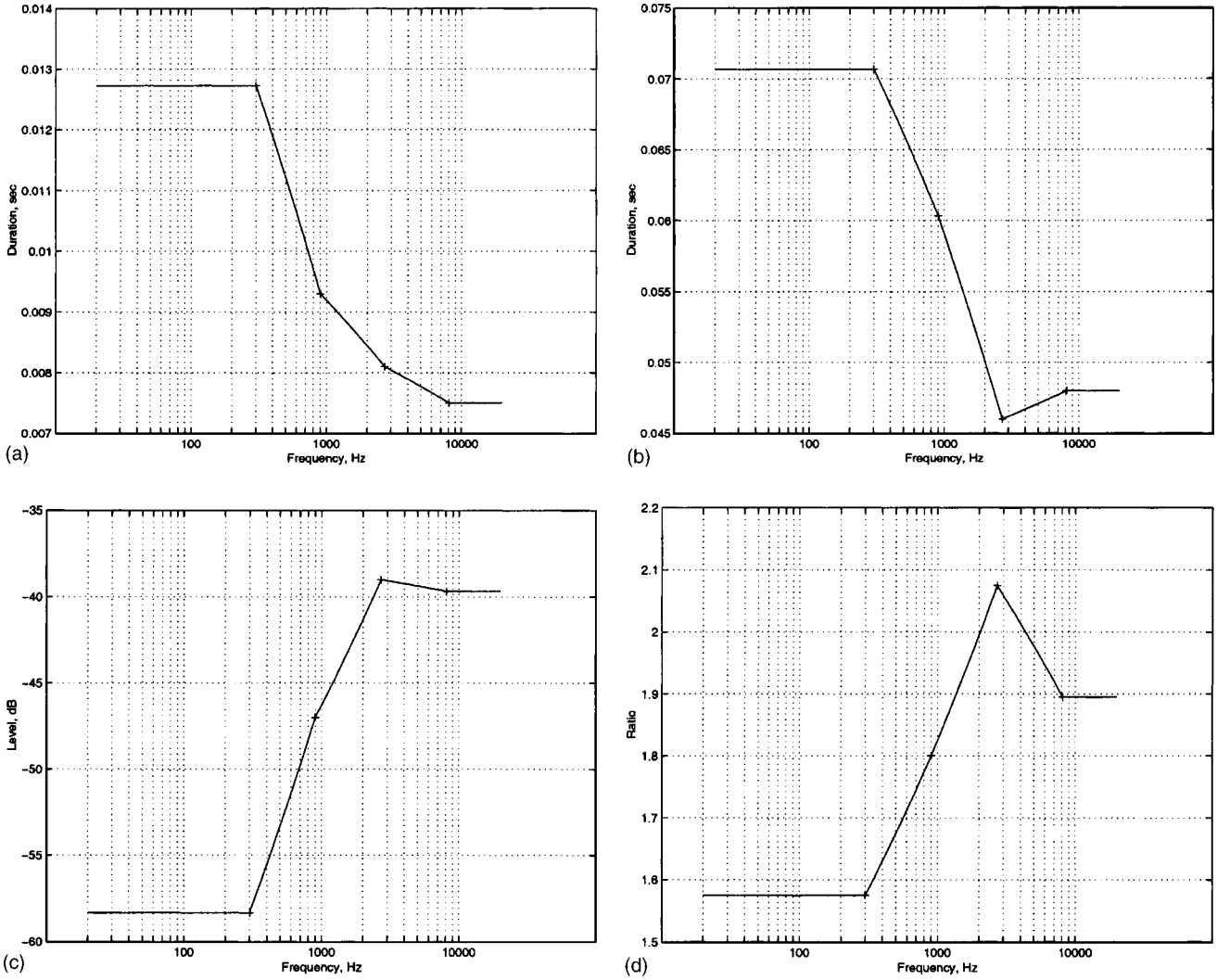


FIG. 2. Interpolated temporal window shape parameters for the design of a frequency-dependent temporal window model: (a) total peak duration, T_p, (b) total skirt duration, T_s, (c) skirt level, w, and (d) window asymmetry, a.

TABLE II. Start, center, and end frequencies for each of the seven temporal windows.

Window No.	1	2	3	4	5	6	7
Band start	0	300	533	949	1687	3000	8100
Band end	300	533	949	1687	3000	8100	20000
Window F_c	300	400	711	1265	2250	4930	8100

modulated (baseband) response, expressed in Eq. (7), is specified by omitting the ''tone'' part of the expression. The filter shape is thus centered at dc. This is referred to by Cooke as a *gamma response* or *low-pass gammatone filter* (Cooke, 1993) and by Slaney as a *baseband gammatone filter* (Slaney, 1993).

$$g(t)=at^{n-1}\exp(-2\pi bt)\cos(2\pi f_c t+\phi), \quad t>0, \tag{6}$$

$$g_W(t)=at^{n-1}\exp(-2\pi bt), \quad t>0. \tag{7}$$

Parameters b and n control the bandwidth and slope of the gamma and gammatone response. A filter order of 3–5 gives a magnitude response which is close to that of the roex filter shape (Moore, 1997). For order $n=4$, b is 1.019 ERB, where ERB is the equivalent rectangular bandwidth defined as follows:

$$\mathrm{ERB}=24.7+0.108f_c. \tag{8}$$

For the purpose of designing frequency smoothing windows a finite impulse response (FIR) was calculated by sampling and windowing the gamma response using the ''window method'' of FIR filter design. The resulting window is referred to here as a *gamma window* and is a real function. To ensure the correct EWD resolution, the gamma window must be autocorrelated with itself prior to multiplication with the time smoothed TCF block (Pielemeier and Wakefield, 1995).[4] After autocorrelation it becomes symmetric, producing a symmetrically smoothed and real EWD. The gamma window is then normalized to ensure its frequency response is 0 dB at dc by dividing by the sum of its samples.

If spectral resolution level dependency is desired, the auditory filter shapes are asymmetric (Moore and Glasberg, 1987), resulting in a complex FIR. In such cases, the frequency smoothing window's frequency response must be flipped about dc for correct EWD frequency resolution (O'Donovan, 2000). This may easily be achieved by using the complex conjugate of the FIR, which is taken either before or after autocorrelation. The overall window design process is summarized in the following, where $w[n]$ is the demodulated FIR (real for symmetric filters, complex for asymmetric), $w_c[n]$ is the response after conjugation, $w_a[n]$ is the response after conjugation and autocorrelation, and $h_F[n]$ is the final normalized kernel smoothing window where N_F is the length of $w_a[n]$,

$$w_c[n]=w^*[n],$$

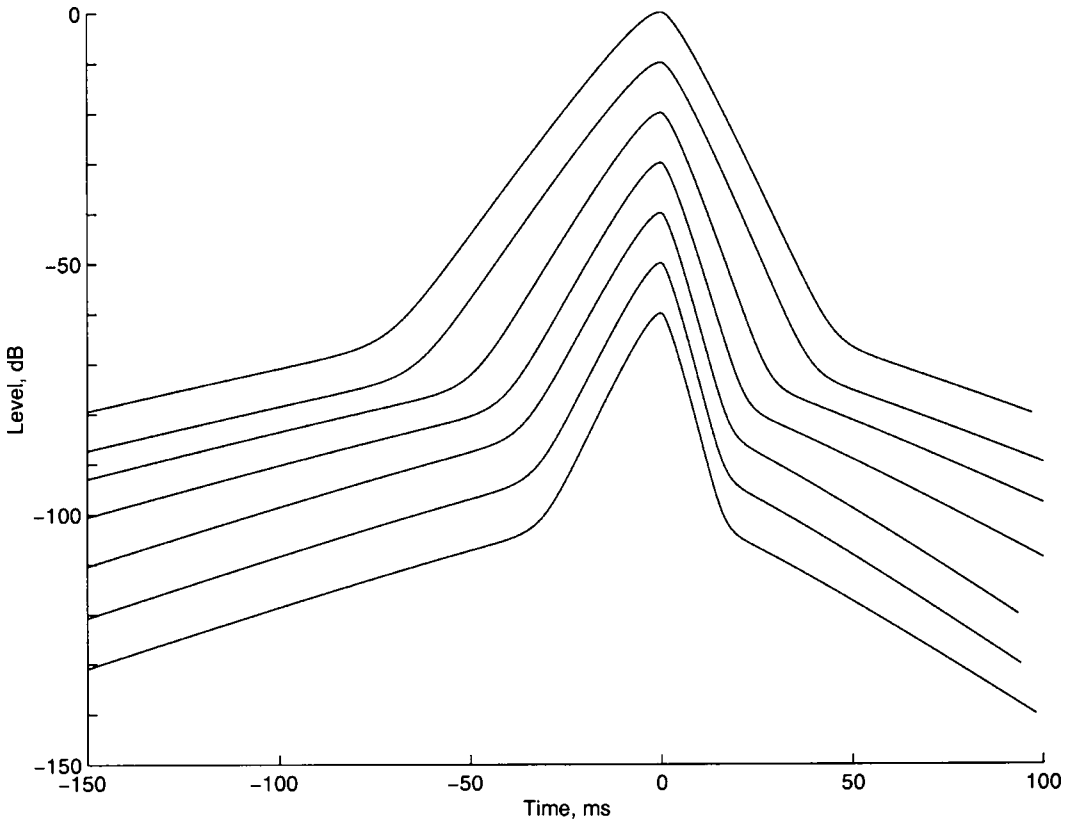


FIG. 3. Temporal windows 1–7 (top to bottom) used for the EWD smoothing kernel. The temporal windows are plotted at offsets of 10 dB for comparison purposes.

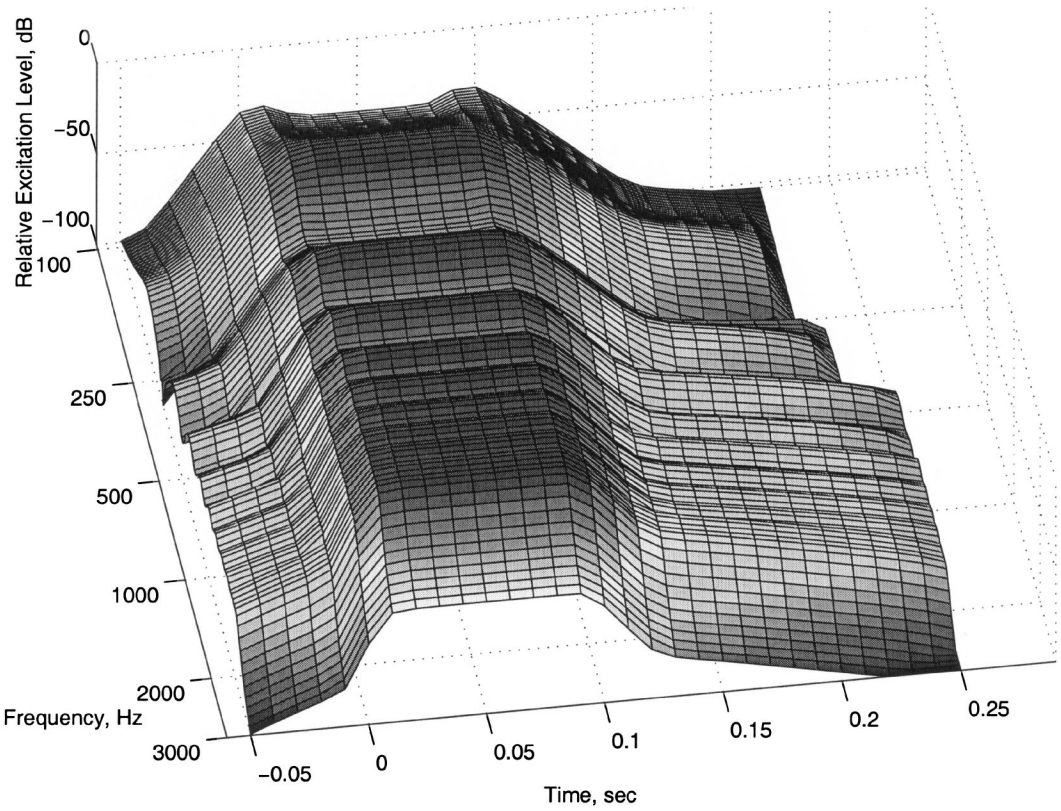


FIG. 4. EWD of a complex tone consisting of the first ten harmonics of a 200 Hz fundamental illustrating spectral and temporal masking.

$$w_a[n]=w_c[n]\otimes w_c^*[-n], \quad h_F[n]=\frac{w_a[n]}{\sum_{n=1}^{n=N_F} w_a[n]}.$$

C. Temporal smoothing window design

Auditory temporal resolution is closely related to auditory masking and refers to the extent to which the ear is accurately able to follow a given stimulus's temporal detail (Plack, 1990). In Plack and Moore (1990), an intensity weighting function, analogous to the auditory filter, known as the *temporal window* model of masking, is proposed. Here, the forward and backward spread of masking in time is modeled by an asymmetric window shape which functions as a running averager of stimulus energy. The window's asymmetry ensures that the forward spread of masking is greater than the backward spread. Each side of the window is expressed as the weighted sum of two rounded exponentials:

$$\begin{aligned} W_b(t)&=(1-w)(1+2t/T_{pb})\exp(-2t/T_{pb})\\ &\quad +w(1+2t/T_{sb})\exp(-2t/T_{sb}),\\ W_a(t)&=(1-w)(1+2t/T_{pa})\exp(-2t/T_{pa})\\ &\quad +w(1+2t/T_{sa})\exp(-2t/T_{sa}). \end{aligned} \tag{9}$$

Forward masking is modeled by $W_b(t)$, the window shape for negative time, and $W_a(t)$ gives the window shape for positive time, modeling backwards masking. Parameters T_{pb} and T_{sb} control the shape of the window peak and skirt for negative time, respectively, with T_{pa} and T_{sa} having the corresponding meaning for positive time. Parameter t is absolute time relative to the window center. The relative peak to skirt level for both sides of the window is controlled by parameter w. In total, twelve averaged temporal window shapes were derived in the study by Plack and Moore by fitting this window shape to masking data for three levels and four center frequencies. Since the temporal window shape is intended as a smoother of excitation level it is ideally suited to EWD kernel design. Furthermore, it is low pass in nature and therefore suppresses cross-term interference oscillating parallel to the time axis (O'Donovan, 2000). The window shape is expressed in units of energy, and therefore, unlike the gamma window, does not require autocorrelation and is applied directly as an EWD time smoothing window.

Plack and Moore derive a temporal window shape for three stimuli levels at four center frequencies of 300, 900, 2700, and 8100 Hz. Interpolation of their data is thus necessary to design a frequency-dependent window which is level independent. Table I presents the new data. Here, T_p is the total peak time constant ($T_p=T_{pb}+T_{pa}$) with T_s being the equivalent measure for the filter skirt ($T_s=T_{sb}+T_{sa}$). Parameter w is averaged across level and ranges between about -60 and -40 dB. Since the values of T_{pb}/T_{pa} and T_{sb}/T_{sa} agree closely at each center frequency and level investigated they are averaged across level giving parameter a, characterizing overall window asymmetry for each of the four center frequencies.

The values of T_p, T_s, w, and a were interpolated with respect to a logarithmic frequency scale in order to design a frequency-dependent window shape. Graphs of the interpolated data are presented in Figs. 2(a), (b), (c), and (d); note that the data are assumed to remain constant below 300 Hz and above 8100 Hz, respectively, since these frequencies

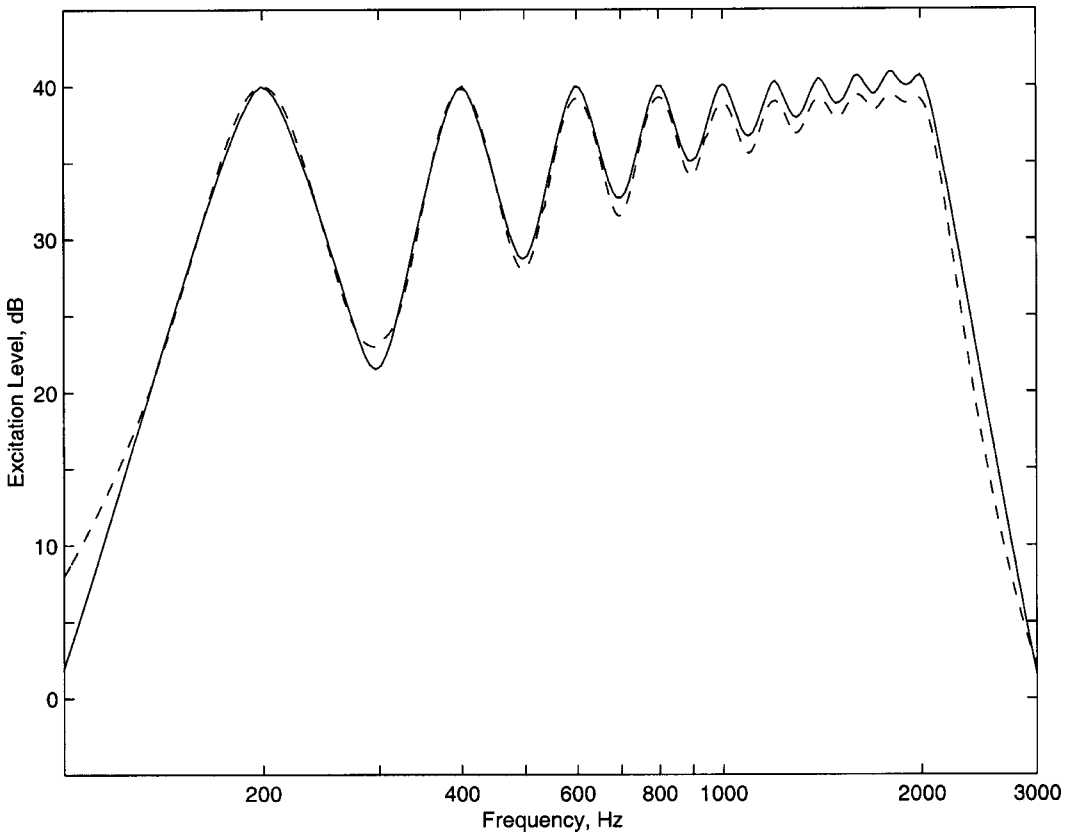


FIG. 5. Comparison of the excitation patterns generated for the symmetric roex(p) filter shape (continuous curve) and the EWD (dashed curve).

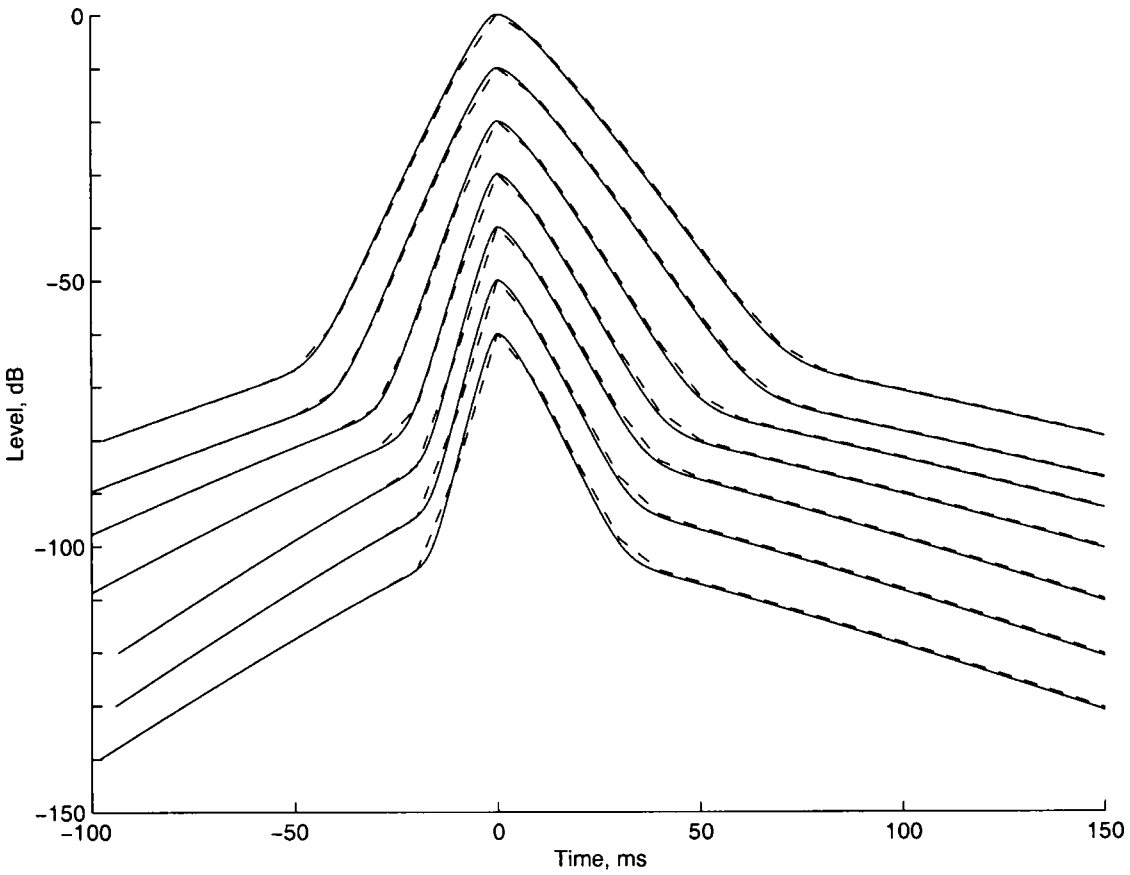


FIG. 6. A comparison of the time-reversed temporal window shape (solid curve) and EWD TEPs (dashed curve) for a short, impulsive signal. The temporal windows are displayed from 1 to 7 (top to bottom), with the graphs normalized and offset by 10 dB for clarity.

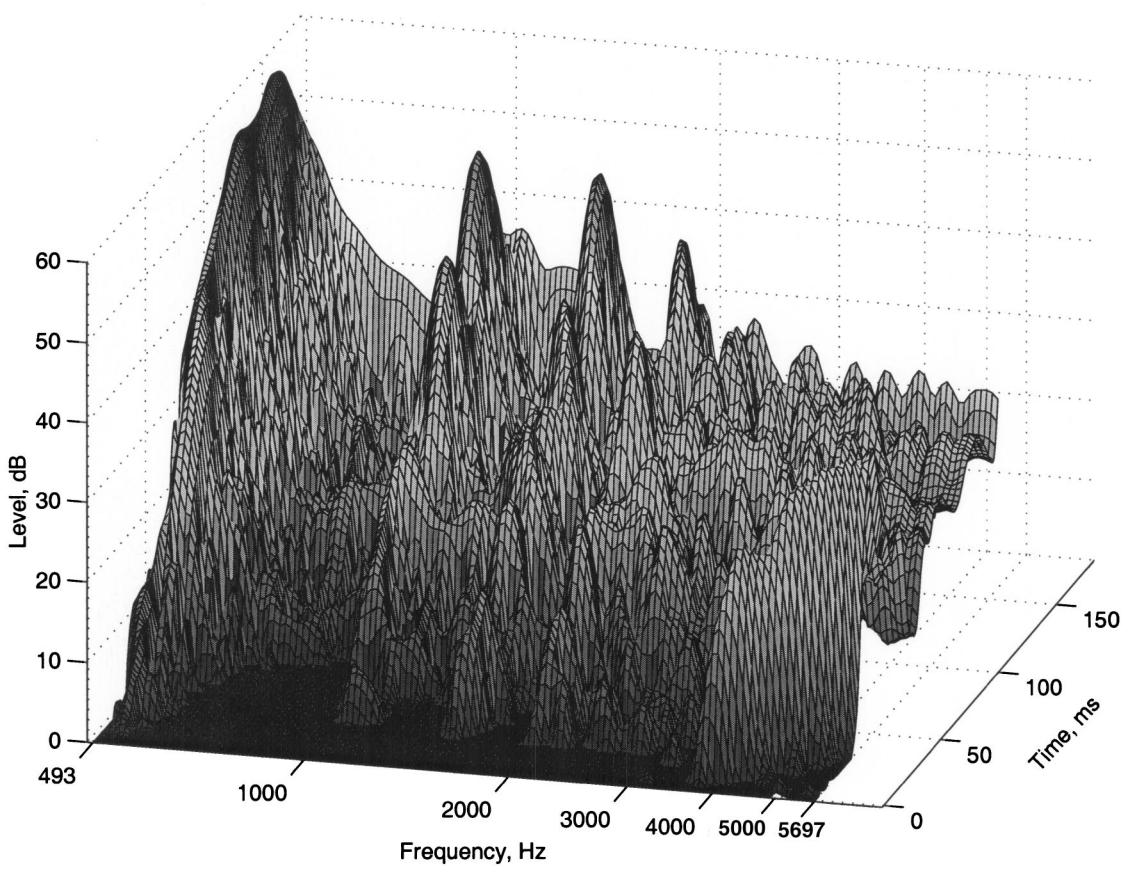


FIG. 7. Constant-Q distribution of a flutter-tongued flute attack for the note C5. The first 60 dB of the distribution is shown.

were the limits of Plack and Moore's study. Overall, the logarithmic interpolation generated a temporal window shape which changed smoothly with frequency and did not diverge significantly from the original twelve window shapes derived in Plack and Moore (1990).

For the purposes of EWD analysis, it was found that a total of seven temporal windows was sufficient to cover the nominal audible frequency range 20 Hz–20 kHz, with each temporal window covering a specific band of frequencies. Table II gives the frequency band of each of these windows. The parameter ''Window F_C'' gives the center frequency of the temporal window assigned to that particular band. For all but the first and last window, this figure is equal to the logarithmic midpoint of ''Band Start'' and ''Band End.'' These 7 temporal windows, $h_T[l,C]$, are plotted in Fig. 3(b), with the temporal window number increasing from top to bottom.

As stated earlier, the EWD's kernel is not currently level dependent. Although this is not a serious limitation for the modeling of spectral masking, temporal masking is more strongly dependent on signal and masker level than frequency masking (Plack and Oxenham, 1998). Therefore, the accuracy of a level-independent model of temporal masking may be somewhat limited, especially for signals with abrupt changes in level, where temporal masking effects become more noticeable. Level dependency has not been incorporated at this stage but may be achieved by varying the window shape with level. Since changes in EWD temporal smoothing may be made without affecting spectral smoothing and vice versa, the incorporation of new psychoacoustic models in the EWD kernel is greatly simplified.

V. EXAMPLES OF THE EWD

An evaluation of the EWD in terms of resolution accuracy, cross-term suppression, window clutter, dynamic range, and computational efficiency is presented in O'Donovan (2000). Here, three examples are included to illustrate some EWD features and also to compare it with its closest time-frequency (TF) relative, the constant-Q modal distribution. The first example presents the EWD of a complex sinusoidal signal, a signal which is commonly used to demonstrate auditory spectral masking. The signal consists of the first 10 harmonics of a 200 Hz fundamental. The total signal duration is 100 ms and the signal is ramped and damped by two ''halves'' of a 10 ms Hanning window to give an equal signal attack and decay of 5 ms. The EWD of this signal is presented in Fig. 4, with a logarithmic frequency axis. The EWD smoothing kernel causes the distribution to dilate with increasing frequency in accordance with the gammatone auditory model. The inter-ripple height of the harmonics decreases with frequency. Note that inter-partial cross-term interference is suppressed below auto-term level and the distribution is clutter free. EWD temporal smoothing is also evident, extending excitation beyond signal duration.

Because of its frequency marginal property, integration of the Wigner distribution along the time axis yields the signal spectrum (Cohen, 1995). For the case of the EWD, the

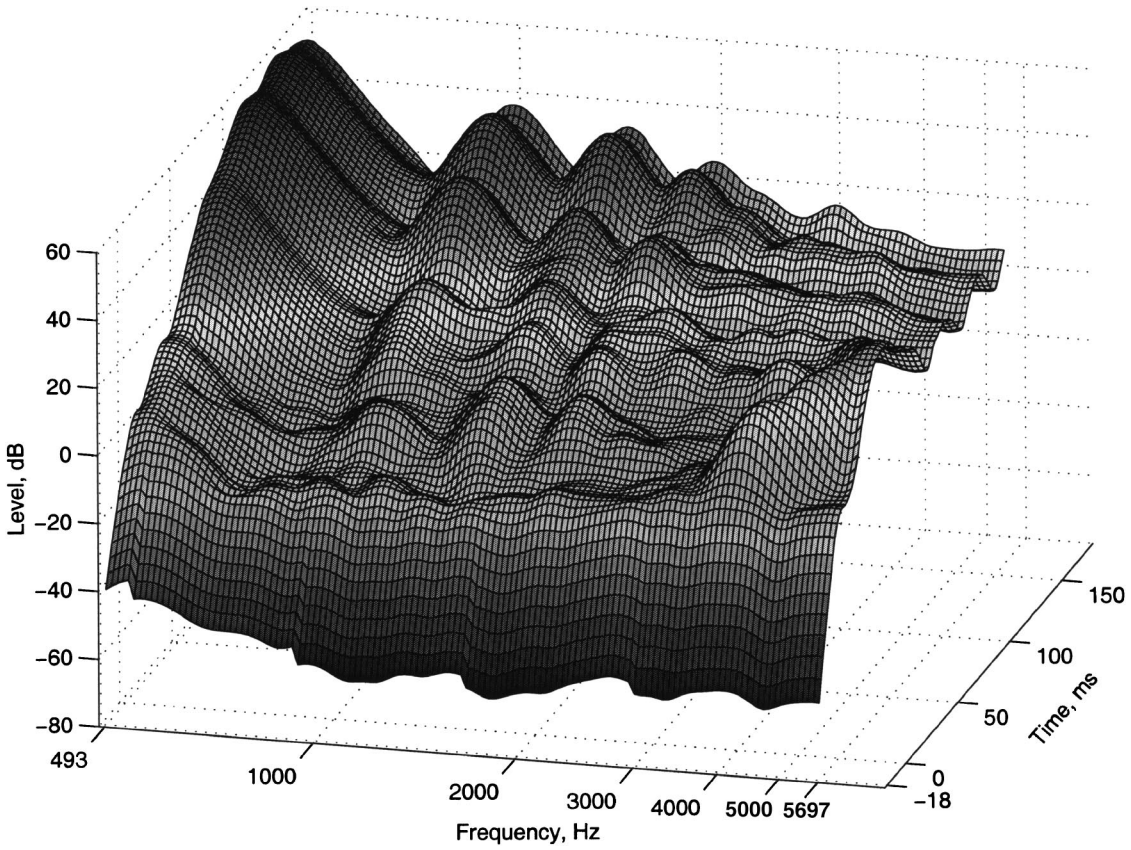


FIG. 8. EWD of a flutter-tongued flute attack for the note C5. The full dynamic range of the distribution is shown.

signal excitation pattern is obtained, as shown in Fig. 5, which presents the frequency marginal of the EWD (dashed curve) of Fig. 4. Superimposed on the same plot is the excitation pattern calculated using the method of Glasberg and Moore (solid curve), for the roex auditory filter shape, the roex and gammatone filter shapes being close for moderate signal levels (Moore, 1997). A high degree of correspondence between the two curves is evident. The EWD excitation pattern is slightly lower for the upper harmonics; this is due to the shortening of the temporal window shape with frequency. The small changes in temporal resolution at frequencies 300, 533, and 949 Hz (see Table I) are due to the switch in temporal window shape at these frequencies.

The Wigner distribution time marginal property means that integration in frequency yields the instantaneous signal power (Cohen, 1995). Correspondingly, for the EWD the temporal excitation patterns (TEPs) (Plack and Moore, 1990) of a signal's partials may be computed by integrating in frequency around each partial of interest using the power estimation techniques described in Pielemeier and Wakefield (1995). In order to verify EWD temporal resolution, Fig. 6 plots the (time-reversed) temporal window shapes of Fig. 3 (solid curves) against the TEPs calculated from the EWD of a short impulsive signal (dashed curves). A 20th-order FIR low-pass filter, with cut-off 9 kHz and duration 1 ms was used here. The EWD frequency integration regions were centered on the 7 temporal window center frequencies given in Table II. The high degree of correspondence between the TEPs and the temporal window shapes verifies EWD temporal resolution accuracy.

Figures 7 and 8 compare the constant-Q modal distribution and EWD of the first 160 ms of a flutter-tongued flute signal. Although these two forms of analyses differ in purpose, the constant-Q distribution aiming for high resolution and the EWD modeling masking, it is useful to compare them as together they highlight the effect of incorporating masking in time-frequency analysis. The signal is taken from Volume 2, Track 2, Index 13 of the McGill University Master Samples (MUMS) CD's (Opolko and Wapnick, 1989). The note value is C5 (semitone 72), giving a fundamental of $f_0 = 523.5$ Hz. The time and frequency range and smoothing used in Pielemeier and Wakefield (1995) for the constant-Q distribution were used here. The distributions are both computed and displayed from semitone 71 (493.9 Hz) to $113\frac{1}{3}$ (5696.3 Hz). The constant-Q distribution and EWD are computed and displayed at a time interval of 1 and 2 ms, respectively, and the EWD is computed before signal onset in order to highlight backward masking effects.

From Figs. 7 and 8 we see that the first four harmonics of the signal are prominent, all showing amplitude modulation characteristic of flutter tonguing. The modulation depth of the fundamental is much smaller than that of harmonics 2, 3, and 4. Also, as noted by Pielemeier and Wakefield, the analysis shows an inharmonic partial present for the first 50 ms after signal onset rising from 3729 to 5587 Hz. This partial can be heard clearly as a brief rising whistle during the note attack.

Distribution resolution is sharper for the constant-Q distribution than for the EWD. The effective dynamic range of the constant-Q distribution is about 60 dB, below which in-

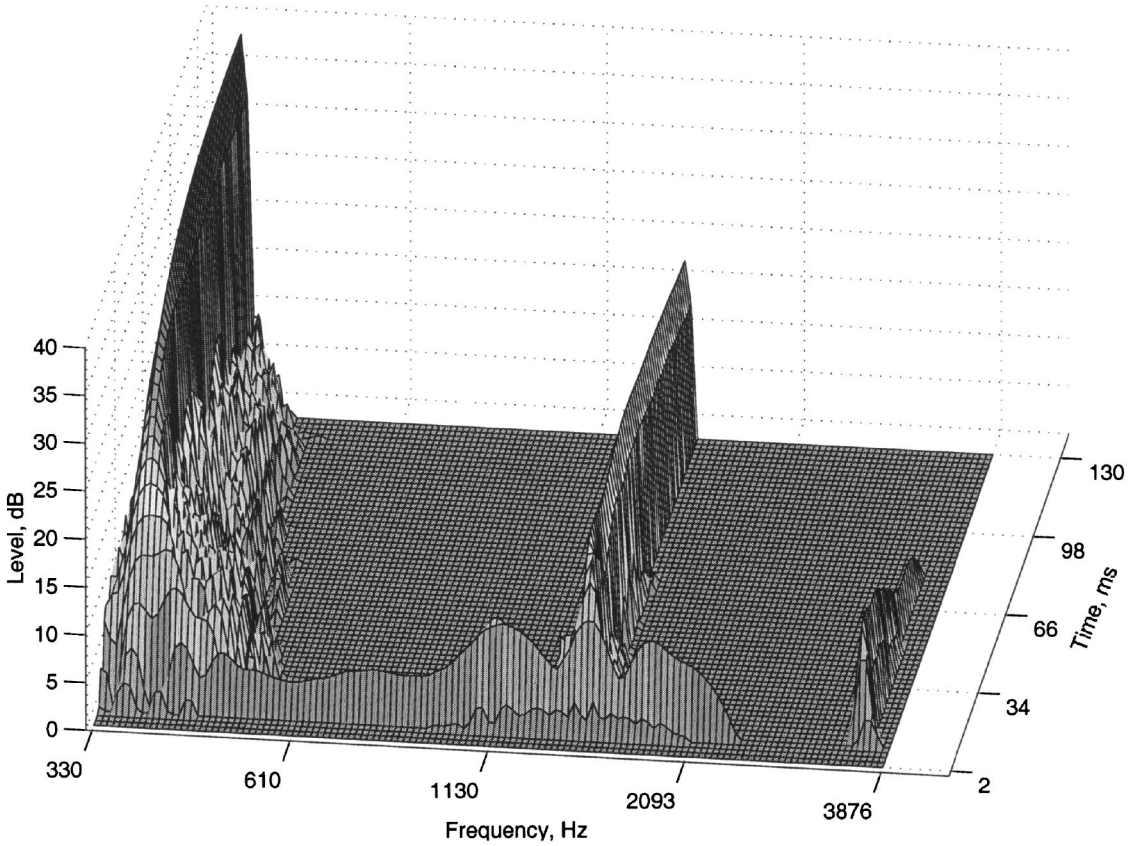


FIG. 9. Constant-Q distribution of a vibraphone note F#4 (hard mallet attack), showing the first 40 dB of the distribution.

terference dominates. In contrast, the EWD is free from interference. Fine low level signal information is resolved by the constant-Q analysis. However, it is often difficult to distinguish between signal energy and window clutter or cross-term interference below the first 30–40 dB. This is especially the case above 2 kHz, where a drop in signal energy occurs.

For the EWD, auditory masking produces the spread of excitation around the first four partials (spectral masking) and the ramp in the distribution before signal onset (backwards temporal masking). The main audible feature of the signal, namely the partial amplitude modulation of the first 4 harmonics, is clear. However, inaudible, low level temporal fluctuations which are present in the constant-Q distribution, are smoothed (masked) for the EWD. Low level signal partials above 2 kHz are also smoothed because of the increased bandwidth of the auditory filter shape. This is consistent with the finding that, in general, only lower harmonics of musical notes are audible. The underlying amplitude modulation of signal energy is still visible above 2 kHz as a low level ripple from 80 ms onwards. Note that the inharmonic partial, audible during the first 50 ms after signal onset, is still prominent and is not masked. The overall effect of incorporating auditory masking is seen to focus the analysis on audible signal information, and to remove inaudible signal features.

The constant-Q distribution and EWD of a second musical signal, a vibraphone hard mallet attack, are shown in Figs. 9 and 10, respectively. The signal sample is taken from MUMS Volume 3, Track 6, Index 14. Again, the time and frequency range and smoothing for the constant-Q distribution are as in Pielemeier and Wakefield (1995) [1/3 semitone intervals from semitone 64 (329.6 Hz) to 106 (3801.8 Hz); display interval of 1 ms]. The constant-Q and EWD distributions are computed for the first 130 and 120 ms of the signal, respectively. The EWD is computed starting at $t=-10$ ms in order to include backward masking effects and the EWD display interval is 2 ms.

Both distributions show a strong impulsive signal onset, followed by the emergence of partials at f_0 (370 Hz), $4f_0$ (1480 Hz) and an inharmonic partial at $9.5f_0$ (3515 Hz). Very brief partials at $2f_0$, $3f_0$, and $5f_0$ during signal onset are also visible; these decay rapidly after about 3–5 ms. The overall envelope and form of the impulsive onset and partial amplitudes agree closely for both distributions. Signal partials are represented by sharp fins for the constant-Q distribution. In contrast, signal partials produce excitation over a broad range of frequencies for the EWD. Due to clutter and interference, the constant-Q distribution is limited to a dynamic range of about 40 dB for this signal. This is not the case for the EWD. Backward masking produced by the impulsive signal attack is clearly visible for the EWD.

VI. CONCLUSIONS

This work describes an auditory model designed within the framework of Cohen's class (CC) of bilinear time-frequency distributions, known as the EarWig distribution (EWD). The EWD algorithm is based on an extension of the constant-Q modal distribution algorithm. In order to incorporate spectral and temporal resolution and model masking rather than achieve high TF resolution, the EWD smoothing

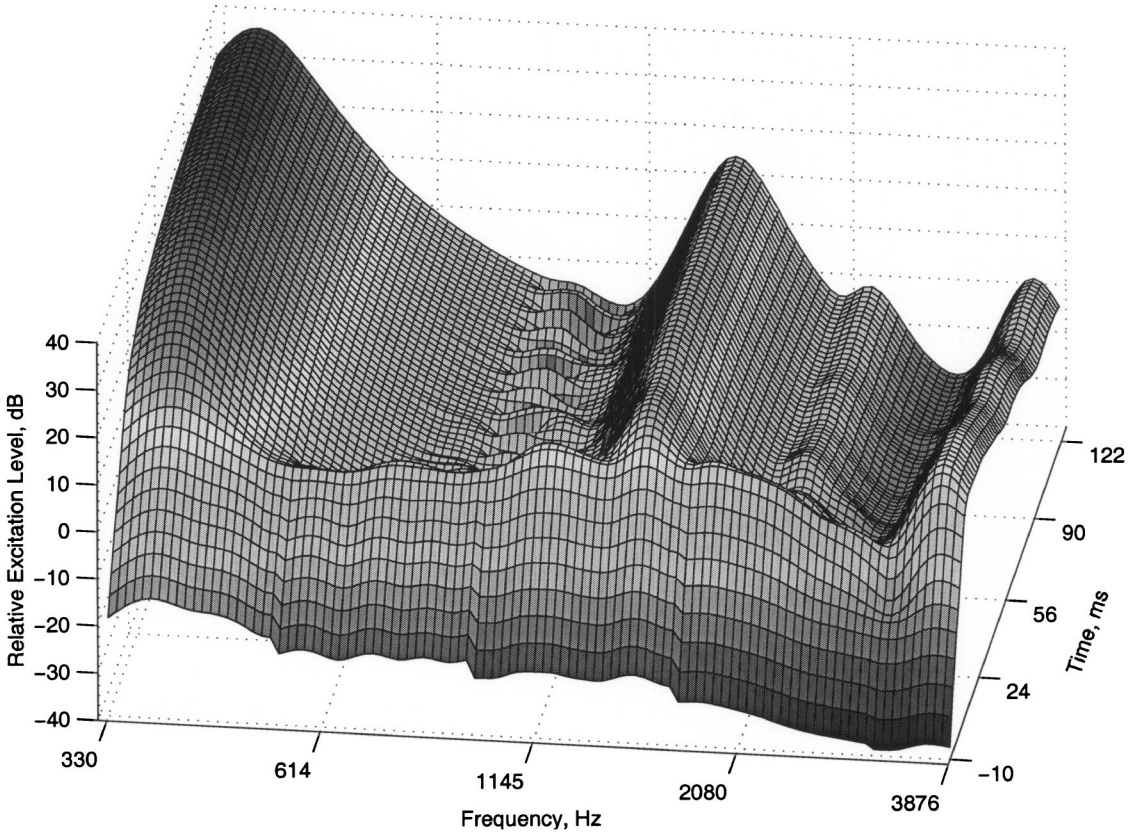


FIG. 10. EWD of a vibraphone note F#4 (hard mallet attack).

kernel is separable and frequency dependent and is based on psychoacoustic models of spectral and temporal masking, rather than standard window functions.

In order to model spectral masking, the gammatone filter model is used to design a set of frequency smoothing windows producing an appropriate degree of spectral smoothing. The temporal window model is adapted to generate a frequency-dependent model of auditory temporal resolution. This is used to provide a set of time smoothing windows which are used to model temporal masking. The combination of temporal and spectral smoothing provided by the auditory kernel ensures that EWD interference and clutter are highly suppressed. This guarantees that auditory resolution is accurate across a dynamic range exceeding 100 dB, which is sufficient to encompass that of the ear. The EWD kernel design may be enhanced to include masking level dependency, if required. This work thus demonstrates that auditory modeling is possible using bilinear TF analysis techniques.

Three examples are included to demonstrate some aspects of the EWD. The analysis of a complex tone illustrates the combined effect of incorporating both spectral and temporal masking in a single distribution. Integration of the EWD along the time axis yields the spectral excitation pattern of the signal, while integration in frequency yields the temporal excitation pattern of the signal partials. The analysis of a flutter-tongued flute signal and a vibraphone attack demonstrate that by incorporating masking, EWD analysis emphasizes perceptually important signal features while removing masked signal detail. Code implementing the EWD is available on the internet (O'Donovan and Furlong, 2002).

ACKNOWLEDGMENTS

J.J.O'D. was supported by a Studentship from Silicon and Software Systems Ltd., Ireland and a Post-Graduate Award from Trinity College, Dublin, Ireland. The authors are most grateful to Associate Editor John Burgess for helpful comments which improved the clarity and content of an earlier version of this paper.

[1]Heisenberg's uncertainty relation must always be satisfied, however, which requires an averaging over a certain area in the time-frequency plane.

[2]The spectrogram $P_{SP}(t,\omega)$, is the absolute magnitude squared of the short-time Fourier transform $\mathrm{STFT}(t,\omega)$, and may be expressed as follows: $P_{SP}(t,\omega)=|\mathrm{STFT}(t,\omega)|^2$. The spectrogram is in fact a member of Cohen's bilinear class. However, since it does not offer independence of temporal and spectral resolutions, it is generally classed as linear.

[3]If the DSPWD is evaulated from the analytic signal, rather than the real signal, the Nyquist sampling rate $f_s \geqslant 2f_c$ is sufficient to avoid aliasing. This results in significant memory and computational savings for some DSPWD applications. As detailed in O'Donovan (2000), typically over 95% of EWD computation time is taken up by TCF time smoothing, which was also the most memory intensive operation. If we assume that 4 real multiplications are required for each complex multiplication the computational advantage of using the analytic signal is unfortunately lost for this operation (although TCF memory reduction of a factor of 2 does remain). Although using the analytic signal removes cross-term interference between positive and negative frequencies for the DSPWD, this effect was not found to be significant for EWD analysis. Therefore, the real signal was used in this work.

[4]Care must be taken when designing smoothing windows for bilinear distributions. If a window exhibiting the desired frequency response is multiplied directly with the TCF, upon transformation to the TF domain the frequency resolution and out-of-band cross-term suppression of the resulting TFR is halved. This is because the TCF is quadratic, whereas the window is not. However, if the window is autocorrelated with itself prior to

multiplication with the TCF (doubling the length of the window and also necessitating a doubling of the width of the TCF), both TCF and window are quadratic, producing the desired TF resolution and cross-term suppression. Note that if the window is applied directly to the signal *before* TCF calculation, a multiplication of the window with its time-reversed complex conjugate in the lag variable results upon TCF calculation. This is equivalent to a convolution in the frequency domain of the window function with itself, which does not produce the desired TF frequency smoothing or cross-term suppression.

Agerkvist, F. T. (**1996**). ''A time-frequency auditory model using wavelet packets,'' J. Audio Eng. Soc. **44**, 37–50.

Boashash, B., editor (**2003**). *Time Frequency Signal Analysis and Processing—A Comprehensive Reference* (Elsevier, Amsterdam).

Claasen, T. A., and Mecklenbräuker, W. F. (**1980a**). ''The Wigner distribution—a tool for time-frequency signal analysis. 1. Continuous-time signals,'' Philips J. Res. **35**, 217–250.

Claasen, T. A., and Mecklenbräuker, W. F. (**1980b**). ''The Wigner distribution—a tool for time-frequency signal analysis. 2. Discrete-time signals,'' Philips J. Res. **35**, 276–300.

Cohen, L. (**1995**). *Time-Frequency Analysis: Theory and Applications* (Prentice Hall, Englewood Cliffs, NJ).

Cooke, M. (**1993**). ''Modelling Auditory Processing and Organisation,'' Ph.D. thesis, University of Sheffield.

Cooke, M., Beet, S., and Crawford, M., editors (**1993**). *Visual Representations of Speech Signals* (Wiley, New York).

Gerzon, M. (**1990**). ''Why Do Equalisers Sound Different?,'' Studio Sound and Broadcast Engineering, 58–65.

Hlawatsch, F., and Flandrin, P. (**1997**). ''The interference structure of the Wigner distribution and related time-frequency signal representations,'' in Mecklenbräuker, W. and Hlawatsch, F., editors, *The Wigner Distribution—Theory and Applications in Signal Processing* (Elsevier, Amsterdam), pp. 59–133.

Irino, T., and Patterson, R. D. (**1997**). ''A time-domain, level dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.

Janse, C. P., and Kaizer, A. J. M. (**1983**). ''Time-frequency distributions of loudspeakers: The application of the Wigner distribution,'' J. Audio Eng. Soc. **31**, 199–223.

Jeong, H., and Ih, J. (**1999**). ''Implementation of a new algorithm using the STFT with variable frequency resolution for the time-frequency auditory model,'' J. Audio Eng. Soc. **47**, 240–251.

Martin, W., and Flandrin, P. (**1985**). ''Wigner-Ville spectral analysis of nonstationary processes,'' IEEE Trans. Acoust., Speech, Signal Process. **33**, 1461–1469.

Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, New York).

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.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.

O'Donovan, J. J. (**2000**). ''Perceptually Motivated Audio Time-Frequency Analysis,'' Ph.D. thesis, Trinity College Dublin.

O'Donovan, J. J., and Furlong, D. J. (**1998**). ''A joint time-frequency model of auditory masking,'' in Proceedings of the 136th Meeting of the Acoustical Society of America.

O'Donovan, J. J., and Furlong, D. J. (**2002**). EarWig Code. http://www.mee.tcd.ie/mmt/ewd.php.

O'Neill, J., Flandrin, P., and Williams, W. (**1999**). ''On the existence of discrete Wigner distributions,'' IEEE Signal Process. Lett. **6**, 304–306.

Opolko, F., and Wapnick, J. (**1987–1989**). McGill University Master Samples (MUMS). Music Dept., McGill University, Montreal, Quebec, Canada.

Patterson, R. D., Allerhand, M. H., 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.

Peyrin, F., and Prost, R. (**1986**). ''A unified definition for the discrete-time, discrete-frequency, and discrete-time/frequency Wigner distributions,'' IEEE Trans. Acoust., Speech, Signal Process. **34**, 858–867.

Pielemeier, W. J., and Wakefield, G. H. (**1995**). ''A high-resolution time-frequency representation for musical instrument signals,'' J. Acoust. Soc. Am. **98**, 2382–2396.

Plack, C. J. (1990). ''Temporal Resolution in the Auditory System,'' Ph.D. thesis, University of Cambridge.

Plack, C. J., and Moore, B. C. J. (**1990**). ''Temporal window shape as a function of frequency and level,'' J. Acoust. Soc. Am. **87**, 2178–2187.

Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.

Richman, M. S., Parks, T. W., and Shenoy, R. G. (**1998**). ''Discrete-time, discrete-frequency, time-frequency analysis,'' IEEE Trans. Signal Process. **46**, 1517–1527.

Slaney, M. (**1993**). ''An efficient implementation of the Patterson-Holdsworth auditory filter bank,'' Apple Computer Technical Report No. 35, Apple Computer Inc.

Various (**1996**). ''Special Issue on Time-Frequency Analysis,'' Proc. IEEE **84**.

Inferring basilar-membrane motion from tone-burst otoacoustic emissions and psychoacoustic measurements

Michael Epstein[a)]
Institute of Hearing, Speech & Language, Communication Research Lab, and Communications & Digital Signal Processing Center, Electrical & Computer Engineering Department (440 DA), Northeastern University, Boston, Massachusetts 02115

Mary Florentine
Institute of Hearing, Speech & Language, Communication Research Lab, and Department of Speech-Language Pathology & Audiology (106/A FR), Northeastern University, Boston, Massachusetts 02115

(Received 30 June 2004; revised 18 October 2004; accepted 19 October 2004)

The amplitude of otoacoustic emissions, which arise on the basilar membrane, is assumed to be proportional to basilar-membrane motion. It should then be possible to assess basilar-membrane motion on the basis of otoacoustic emissions. The present study provides support for this possibility by comparing basilar-membrane motion inferred from emissions to that inferred from psychoacoustic measures. Three psychoacoustic measurements believed to be associated with basilar-membrane motion were investigated: (1) pulsation threshold; (2) loudness functions derived from temporal integration; and (3) loudness functions derived from loudness matches between pure tones and multitone complexes. Results of the psychoacoustic measurements and of the tone-burst otoacoustic emissions led to very similar estimations of basilar-membrane motion. Accordingly, emissions could serve as an excellent tool—one that is objective, noninvasive, and rapid—for estimating relative basilar-membrane motion.
[DOI: 10.1121/1.1830670]

PACS numbers: 43.64.Jb, 43.66.Cb [BLM] Pages: 263–274

I. INTRODUCTION

This study considers whether tone-burst otoacoustic emissions (TBOAEs) are suitable for inferring basilar-membrane input–output (BM I/O) functions. Withnell and Yates (1998) first suggested the use of otoacoustic emissions to examine basilar-membrane nonlinearity. They used distortion-product otoacoustic emissions (DPOAEs) in guinea pigs to assess BM I/O successfully. However, Buus *et al.* (2001) used DPOAEs in humans to infer the lower portion of the BM I/O function, and their results suggest that the stimuli necessary to obtain appropriate DPOAEs vary greatly across individuals; hence, a simple procedure applicable to all, or even most, human subjects would be unattainable. Although DPOAEs may not be suitable for inferring BM I/O functions, recent evidence (Epstein *et al.*, 2004) suggests that TBOAEs may provide more encouraging results.

Models (Zweig and Shera, 1995) and data (Shera and Guinan, 2003; Goodman *et al.*, 2004) in which OAEs are the result of place-fixed coherent reflections from impedance irregularities on the BM also support the idea that TBOAEs should be related to BM I/O. This would indicate that TBOAEs arise at the tip of the traveling wave near one of these irregularities and have a delay equivalent to a round-trip cochlear delay. One concern is that small variation in the location of stimulation due to the level dependency of traveling wave peaking could produce significant variation in cochlear reflectance. Epstein *et al.* (2004) suggest that TBOAEs are not significantly affected by these variations under the conditions used in these experiments.

More specifically, Epstein *et al.* (2004) assessed the impact of a variety of potentially confounding TBOAE measurement issues and developed a reasonable and repeatable measurement procedure specifically designed for measuring TBOAEs near 1 kHz. The slopes and kneepoints of the average TBOAE I/O functions resembled BM I/O functions seen in physiological measurements on animals (Rhode, 1971; Yates, 1990; Ruggero *et al.*, 1997). Nonetheless, functions varied widely and it was not clear to what extent the variability reflected true BM I/O or resulted from other factors. Although other studies using TBOAEs did not specifically estimate basilar-membrane mechanics, they do hint at the potential for the use of TBOAEs for studying level-dependent amplification within the auditory system and for comparing results with psychoacoustic measures (Norton and Neely, 1987; Neumann *et al.*, 1997). Psychoacoustic measurements also yield results that appear to closely correspond with BM I/O functions (Oxenham and Plack, 1997). The purpose of the present research is to compare TBOAEs with BM I/O functions obtained using three different established psychoacoustic measurements.

II. EXPERIMENTS

A. Overview of experiments

Tone-bust otoacoustic emissions were measured in experiment 1 and then three psychoacoustics experiments were performed using the same subject pool to obtain several estimates of BM I/O for comparison with the TBOAEs. The

[a)]Author to whom correspondence should be addressed. Electronic mail: mepstein@ece.neu.edu

TBOAE I/O measurements are compared with loudness functions based on temporal-integration measurements (experiment 2), loudness functions derived from loudness matches between multitone complexes and pure tones (experiment 3), and pulsation thresholds (experiment 4). These three experiments were chosen to complement one another. Specifically, the model used for generating loudness functions from temporal-integration measurements is best for assessing loudness at moderate levels, and the model used for generating loudness functions from the multitone matches is best for assessing loudness at low levels. Pulsation threshold provides a more direct assessment of BM I/O and requires no modeling to examine its correspondence with TBOAE I/O. Because of the general complexity and subjectivity of assessing the quantitative correspondence of the various methods, quantitative analysis and a discussion of why it may be misleading is provided in the Appendix.

1. Subjects

Eight subjects, three male and five female, were tested monaurally. Four subjects participated in all of the experiments. Two subjects, S5 and S6, were replaced in experiment 3 by S7 and S8. No subjects had a history of hearing difficulties, and their audiometric thresholds did not exceed 10 dB HL at octave frequencies from 250 Hz to 8 kHz (ANSI, 1989), except one subject, S6, whose threshold was 15 dB HL at 8 kHz. Their ages ranged from 20 to 27 years.

2. Apparatus

Each subject was tested in a sound-attenuating booth. A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, recorded responses, and executed the experimental procedure.

For the OAE experiment (1), a TDT DA1 converted the waveform calculated by the AP2 to an analog signal. Stimulus level was adjusted using a TDT PA4 attenuator and the stimulus was antialias filtered using a TDT FT5 whose output was sent to a TDT HB6 headphone buffer. All sounds were presented to the subject via an Etymotic Research ER-10C system, which included a microphone for recording OAEs. The recordings from the ER-10C were antialias filtered, amplified, and converted to a digital signal using a TDT AD1. The digital signal was sent back to the AP2, which averaged the responses to the stimuli and wrote the averaged signals to disk for further analysis.

For the pyschoacoustics experiments (2–4), a similar system was used, but instead of the ER-10C, a Sony MDR-V6 headphone was used to present the sounds. In addition, a second PA4 attenuator came after the FT5 antialias filter. Subjects made responses on a small terminal.

B. Experiment 1: Click-evoked/tone-burst otoacoustic emissions

1. Introduction

Experiment 1 uses the method developed in Epstein *et al.* (2004) to measure TBOAE I/O functions for the eight subjects who participated in the psychoacoustics experiments.

2. Procedure

In order to identify a frequency region near 1 kHz that produced strong otoacoustic emissions, the entire basilar membrane was first stimulated by clicks and the resulting click-evoked otoacoustic emissions (COAEs) were measured. Frequencies near 1 kHz were selected for each subject individually for use as the tonal stimuli in the measurement of TBOAEs to simplify comparisons with psychoacoustic data and because TBOAEs tend to be strongest in this frequency region.

The COAEs and TBOAEs were determined from the real, positive part of the cross spectrum of two independent averages of the time-windowed responses measured in the ear canal in order to help eliminate as much noise as possible. This method is based on the assumption that the OAE is phase locked to the stimulus and has highly repeatable components with identical phases in the two buffers in the time domain, and with zero phase in the cross spectrum. Thus, the emissions will be the real and positive part of the cross spectrum. As the phase of the measured responses becomes more variable, the phase in the cross spectrum is increasingly likely to differ from zero. Averaging and using only the real, positive part of the cross spectrum reduces the level of the recording that is not phase locked to the signal. The two averages used to get the cross spectrum were obtained by alternatingly entering the results of recordings into two different buffers. Subjects were presented with 200 clicks and responses were recorded at a sample rate of 50 kHz for 81.92 ms (4096 samples), including the time of the click itself. These recordings were windowed using a 20.48-ms (1024 samples) Hanning window covering the first portion of the recorded signal. Because the stimulus was a click, the direct stimulus was effectively eliminated by the very small value of the windowing function during the signal. The cross spectrum of the two sets of recordings was examined visually. The strongest peak around 1 kHz was chosen for further examination. To ensure that the peak was stable, this measurement was performed three times and peaks were used only if they appeared clearly in all three measurements. Once chosen, measurements of TBOAEs were made at this frequency.

Tone-burst evoked emissions were measured by presenting a series of brief tone bursts and recording the sound in the ear canal for 81.92 ms at a sample rate of 50 kHz. To allow calculation of the cross spectrum, the recordings were stored for further analysis in two interleaved sets over 360 trials. The stored responses were Hanning windowed using a 1024-sample window (20.48 ms). The window started after a delay of 15 ms. The spectrum of the windowed ear-canal response was bandpass filtered between 400 and 1400 Hz.

To reduce the effect of external noise, a response to a tone burst would be included in the average only if the noise floor was estimated to be less than 10 dB SPL. The noise floor for the rejection decision was calculated using a 20.48-ms Hanning windowing (with a window delay of 13 ms) of the difference between two samples. The noise intensity was then calculated as the sum of the squared magnitudes of the real FFT components between 32 and 819 Hz. This measure is highly conservative and guarantees that any

TABLE I. COAE peak frequencies for all subjects.

Subject	Test frequency (Hz)
S1	988
S2	990
S3	1025
S4	960
S5	1160
S6	1123
S7	1233
S8	950

data points above the noise floor are presented with a high level of confidence. Some of the measurements below the noise floor are also likely to be reliable, but the level of confidence is lower. Measurements were made for levels ranging in 5-dB steps from approximately 10 dB below the minimum level (typically 15–30 dB SPL) that produced an emission whose level exceeded the background noise to the maximum obtainable level. These measurements were repeated three times for each condition and averaged.

3. Stimuli

The clicks were 70-dB SPL 0.1-ms stimuli that contained energy between 450 Hz and 8 kHz. For each subject, a dominant peak in the click-evoked otoacoustic emissions (COAE) spectra around 1000 Hz was identified. The tone burst presented had an equivalent rectangular duration of 6/frequency. These tone bursts had a Gaussian envelope to help contain the spectral spread of the signal due to onset and offset. The duration was chosen to obtain a useful compromise between frequency specificity and the ease of separating the stimulus from the emission produced by it. Using a long-duration signal makes the OAE difficult to measure because the signal would overlap the emission. Using a short-duration signal decreases frequency specificity because, as the duration of the signal decreases, the bandwidth of the signal increases.

Table I gives the COAE peak frequency determined in experiment 1 and used for each subject as the the stimulus frequency for all other experiments.

4. Results and discussion

Figure 1 shows TBOAE levels as a function of stimulus level for eight subjects. The noise floor, open circles, is a conservative measurement indicating the level above which there is a very strong confidence regarding the reliability of the measurements. Even below the noise floor, measurements are likely to have been real and repeatable.

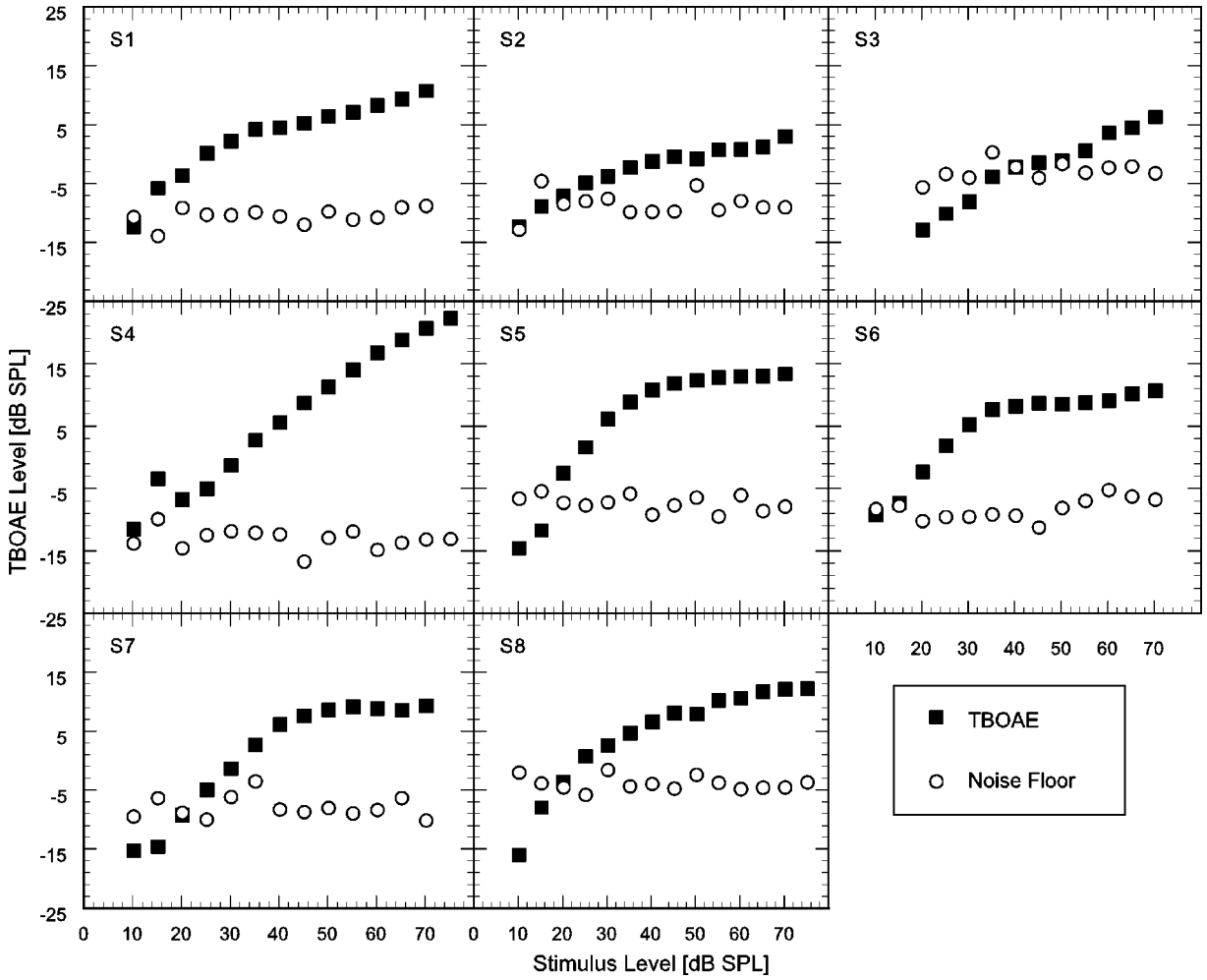


FIG. 1. The TBOAE level plotted as a function of stimulus level for each subject. The filled squares show the TBOAE level and the open circles show the noise floor.

TABLE II. Slopes and kneepoints for TBOAE I/O function fits for all subjects.

	S1	S2	S3	S4	S5	S6	S7	S8	Mean
Slope above kneepoint	0.19	0.13	0.29	0.51	0.08	0.08	0.04	0.22	0.19
Slope below kneepoint	0.41	0.26	0.61	0.78	0.72	0.68	0.74	0.43	0.58
Kneepoint	32.36	37.17	35.00	37.99	37.80	33.27	42.46	35.00	36.38

Table II gives the slope of the lower part and the compressive upper part of the TBOAE I/O function and the kneepoint where the two meet for all subjects. Kneepoints were determined by varying the slope and intersection of two line segments in a square-error minimization fit with the TBOAE data between 25 and 70 dB SPL, which is above the noise floor for all subjects at all points except the lowest levels for S3. Most subjects' functions show a more linear growth (estimated slope mean = 0.58) at low levels and a much more compressive region (estimated slope mean = 0.19) at moderate levels. Some functions also start to show an increase in slope at the highest levels measured. The TBOAEs for S4 exhibit only a slight change in slope across the entire range, with the estimated slope dropping from 0.78 at low levels to 0.51 at moderate levels.

If TBOAE I/O is closely related to BM I/O then the compressiveness of the functions should be consistent with physiological measures in mammals (approximately 0.2) (Ruggero *et al.*, 1997). The mean results are consistent with physiological measurements, but the individual measures varied over a larger range (0.04 to 0.51) than expected. The kneepoints for all subjects were generally close to the previously reported value of 35 dB SPL (Plack and Oxenham, 1998; Buus *et al.*, 2001). These results will be compared with the psychoacoustic experiments (experiments 2–4) in order to see if the individual variations seen in TBOAE I/O are correlated with BM I/O and are not the clear result of some other mechanism. Additional discussion of the fitting method and slopes obtained in this experiment is given in the Appendix.

C. Experiment 2—Temporal integration of loudness

1. Introduction

Loudness functions derived from psychoacoustic measures have been found to resemble BM I/O functions measured in animals (Buus *et al.*, 1997; Schlauch *et al.*, 1998; Buus and Florentine, 2001); both functions show midlevel compression. In the present experiment, loudness functions were determined from loudness matches between long and short tones and compared with TBOAE I/O functions. Previous work has shown that loudness functions, based on measurements of temporal integration—the increase in loudness with an increase in signal duration up to about 200 ms—closely correspond to BM I/O functions (Buus *et al.*, 1997; Buus and Florentine, 2001).

2. Procedure

a. Absolute threshold. Thresholds were measured to provide the basis for setting the sensation levels (SLs) used in loudness matching. Individual thresholds were measured for 5-ms and 200-ms tones at the peak COAE frequency selected in experiment 1. Measurements were performed using an adaptive, two-interval, two-alternative forced-choice (2I, 2AFC) paradigm with feedback. On each trial, two observation intervals, marked visually, were presented with an interstimulus interval of 500 ms. The tone was presented in either the first or second observation interval with equal *a priori* probability. The subject's task was to press one of two buttons corresponding to the interval containing the stimulus. After a 100-ms delay, a 200-ms feedback light indicated the correct answer. Following the feedback, the next trial began after a 500-ms delay.

For each subject and duration, three threshold measurements were made. A single measurement consisted of three interleaved tracks, each of which ended after five reversals. Reversals occurred when the signal level changed from increasing to decreasing or vice versa. On each trial, the track was selected at random among the tracks that had not yet ended. For each track, the level of the signal was initially set approximately 15 dB above the subject's expected threshold. It decreased following three correct responses and increased following one incorrect response, such that the signal converged on the level yielding 79.4% correct responses (Levitt, 1971). The step size was 5 dB until the second reversal, after which it decreased to 2 dB.

The threshold for each track was calculated as the average signal level of the fourth and fifth reversals; the mean of all three tracks was considered a single threshold estimation. Three such estimations (for a total of nine tracks) were obtained for each subject and duration. The mean of the three estimations was used as the reference to set the SL for each subject and stimulus in the remaining portions of the experiment.

b. Temporal integration. Previous studies (Buus *et al.*, 1997; Buus, 1999) have shown that loudness functions may be derived from measurements of the amount of temporal integration (TI), defined as the SPL difference between equally loud short and long tones. In the present study, loudness matches were made between 5-ms and 200-ms tones using an adaptive, roving-level, 2I, 2AFC procedure. Ten concurrent loudness matches were obtained by randomly interleaving ten adaptive tracks. Five of these tracks varied the short tone and five varied the long tone. The fixed stimulus for each of the tracks was set to SLs between 5 and 100 dB for the long tone or 110 dB for the short tone in equally spaced steps. This procedure ensured that subjects could not identify the stimulus being varied and forced them to use only the two stimuli presented in the current trial to make the loudness judgment (Buus *et al.*, 1997; Buus, 1999).

On each trial, the subject heard two tones separated by a 600-ms interstimulus interval. The fixed-level tone followed the variable tone or the reverse with equal *a priori* probability. The subject indicated which sound was louder by press-

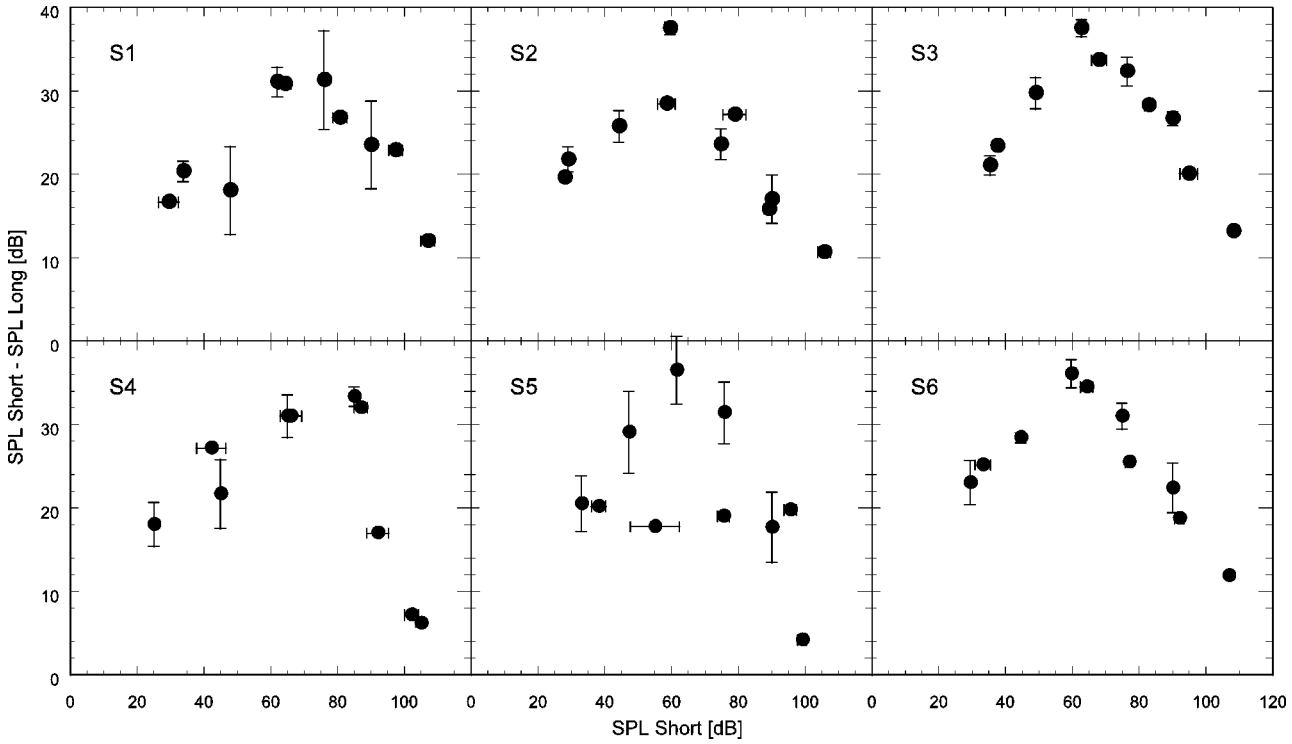


FIG. 2. Temporal integration (difference in level between equally loud short and long tones) as a function of the level of the short tone. Error bars show the standard error for three measurements at each level.

ing one of two keys on a response terminal. The next trial began after a 1-s delay. The level of the variable tone was adjusted according to a simple up–down method. If the subject indicated that the variable tone was louder than the fixed tone, its level was reduced; otherwise it was increased. The step size was 5 dB until the second reversal, after which it was 2 dB.

For each track, the variable stimulus was initially set approximately 15 dB below the expected equal-loudness level, but no lower than threshold. This starting level ensured that the subject would initially hear some trials where the short tone was definitely louder and some trials where the long tone was definitely louder. On each trial, the track was chosen at random among those that had not yet ended, which they did after nine reversals. This method made the variable tone converge toward a level at which it was judged louder than the fixed tone in 50% of the trials (Levitt, 1971). The average level of the last four reversals of each track was used as an estimate of the level at which the loudness of the variable tone was equal to that of the fixed-level tone.

3. Stimuli

Loudness matches were made between short and long tones with equivalent rectangular durations of 5 and 200 ms. The short tone consisted of a 6.67-ms raised-cosine rise followed by a 6.67 ms raised-cosine fall. The long tone had the same rise and fall with a 195-ms steady-state segment in between. This ensured that most of the energy was near the center frequency. (Table I gives the frequencies used for each subject.)

4. Results and discussion

Figure 2 shows the difference in level for equally loud short and long tones as a function of the level of the short tone. All subjects showed a nonmonotonic pattern of TI with the largest amount at moderate levels. This is the form of the function seen in previous measurements of TI (for review, see Florentine *et al.*, 1996).

In order to infer loudness functions from these results, it is assumed that loudness functions for short and long tones are parallel, which is likely to be true, except perhaps at low levels, according to the equal-loudness-ratio hypothesis (Florentine *et al.*, 1996). The hypothesis, which has been supported by data (Florentine *et al.*, 2001), states that the ratio of the loudness of tones of different durations differs by a constant scale factor and is unaffected by SPL. The loudness matches were used to fit loudness functions to the data using a modified version of Zwislocki's loudness model (Zwislocki, 1965; Buus and Florentine, 2001). This model adjusts Zwislocki's modified power function model to allow the low-level slope to vary. The loudness (N) for a single tone using this model is

$$N(\mathrm{SL}) = k[(1 + (\mathrm{SNR}_{th} \cdot 10^{\mathrm{SL}/10})^{S_{-\infty}})^{S_{\infty}/S_{-\infty}} - 1], \qquad (1)$$

where k is a scale factor that does not affect predictions, SNR_{th} is the signal-to-noise ratio at threshold, which is estimated within a critical band centered on the frequency of the tone, SL is the sensation level of the tone in dB, $S_{-\infty}$ is the asymptotic local exponent of the loudness function at low levels, and S_{∞} is the asymptotic local exponent of the loudness function at high levels.

Loudness functions derived from Eq. (1) are compared with TBOAEs in Fig. 3. The TBOAE functions were arbi-

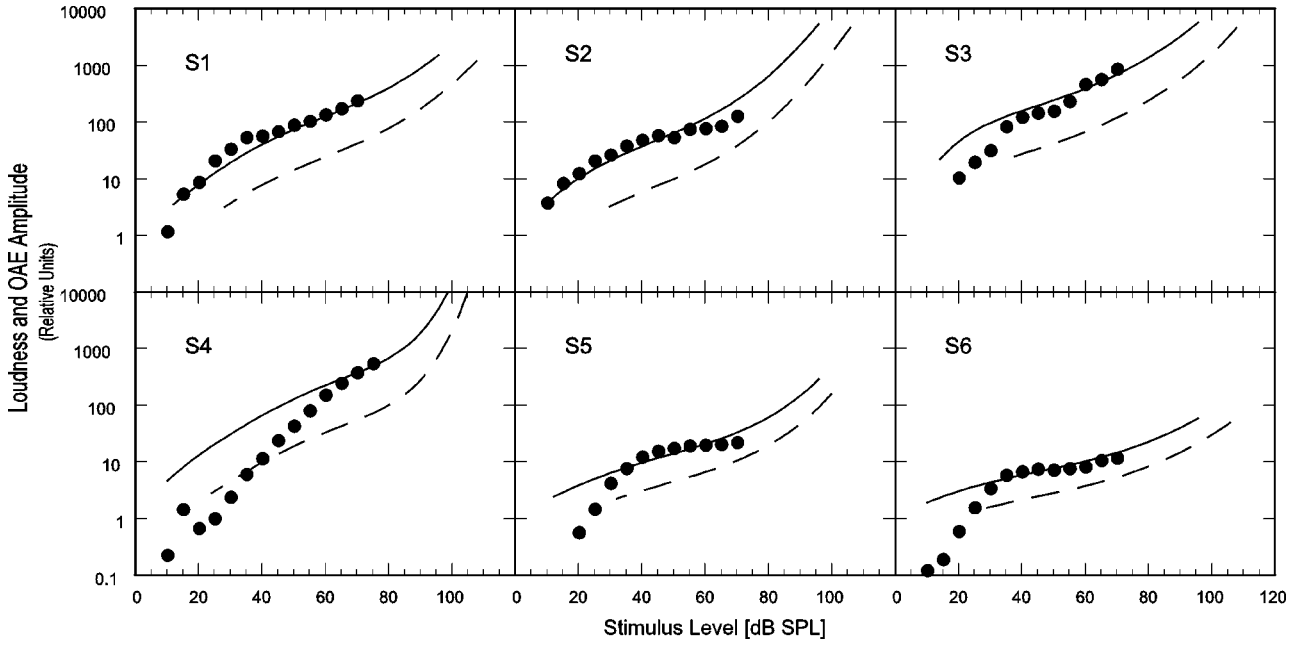


FIG. 3. Loudness functions for short (dashed lines) and long (solid lines) tones generated using the Buus and Florentine (2001) loudness model are shown with TBOAE I/O data (filled circles). The TBOAEs are plotted on a fixed-size 50-dB scale with an arbitrarily shifted range for all subjects. For the exact values of the range, see Fig. 1.

trarily shifted vertically; maintaining the same number of unit divisions, in order to obtain a good visual fit with the loudness functions.

The comparisons between TBOAE I/O functions and the model-generated loudness functions should show close correspondence at moderate levels if the TBOAE I/O functions and loudness functions are related to BM I/O. Five of the six subjects show very good correspondence between 35 dB SPL, the TBOAE I/O function kneepoint, and 60 dB SPL. The remaining subject, S4, shows a very steep TBOAE I/O function that does not correspond well with the midlevel loudness function. The slope for S4 is more similar to the high-level loudness function.

The similarity of the two functions indicates that they are likely to be showing the same aspect of auditory processing, at least at moderate levels. The loudness model used assumes that the loudness functions for the short and long tones are parallel. Since this is untrue for low levels (Florentine *et al.*, 2001), it is not surprising that there is discrepancy between the low-level TBOAEs and low-level loudness functions. In this case, loudness functions for short tones would be more comparable because TBOAEs are evoked with brief stimuli and the model is primarily based on the form of the loudness function for long tones.

Furthermore, it is likely that S4 (the first author) exhibits an outlying rapid growth of loudness that cannot be accounted for by the model. In fact, if TBOAEs are related to loudness growth, this is expected because S4 shows a steep growth of TBOAE I/O. In addition, other measurements of loudness for this individual have consistently shown an abnormally rapid growth.[1]

D. Experiment 3—Multitone loudness summation

1. Introduction

In addition to the loudness functions derived from loudness matches between tones of different durations, it is also possible to derive functions from loudness matches between pure tones and multitone complexes. If the loudness of each component of the tone complex is equal and their loudnesses sum linearly, then the loudness of the tone complex must be equal to the loudness of one component times the number of components (Fletcher and Munson, 1933). This is true if all components are at the same SL, loudness grows at the same rate for at all frequencies (within the range studied), and there is no mutual masking between components.

2. Procedure

a. Threshold. Threshold was measured the same way as in experiment 2 except that the tone duration was 500 ms. This change was made in order to replicate the conditions in Buus *et al.* (1998).

A Békésy tracking procedure was used to examine local threshold variations at frequencies used for the multitone-complex loudness matches. Tones were presented monaurally sweeping across from -0.7 to $+0.5$ Barks from the frequency being examined at a rate of 0.005 Barks per presentation (800 ms). Subjects were asked to hold down a button as long as the sound was audible, and to release the button as soon as the sound became inaudible. The level either increased or decreased by 1 dB at each presentation. This procedure allowed frequencies that were in areas of very steep threshold change to be avoided. It has been hypothesized that frequencies within steep threshold changes may not yield the same loudness growth near threshold as those outside of these steep regions (Kemp, 1979; Cohen, 1982). Frequencies that were found to be within these steep areas were shifted to outside the region. This occurred for all subjects only at their test frequency (Table I).

b. Loudness matching. The present study used a procedure identical to that used by Buus *et al.* (1998). Comparisons were made using an adaptive procedure with a 2I,

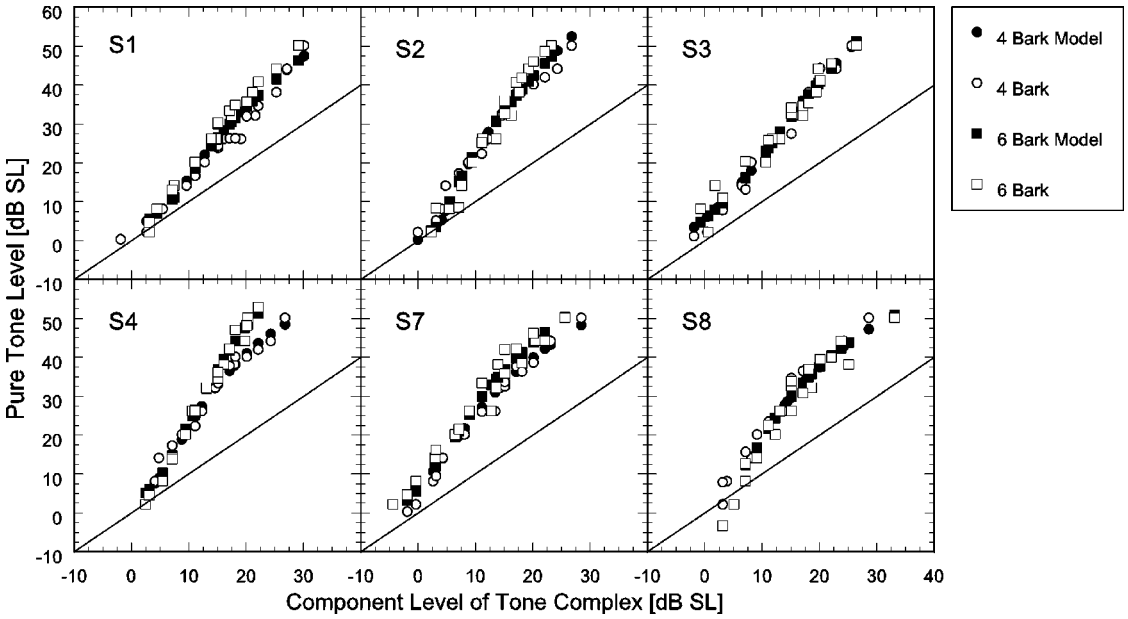


FIG. 4. Loudness matches between pure tones and multitone complexes with overall spacings of 4 (filled circles) or 6 (filled squares) Barks compared with the loudness model fits (open symbols).

2AFC paradigm. On each trial, the subject heard a pure tone and a tone complex presented in random order with equal *a priori* probability and a 600-ms interstimulus interval. The subject indicated which interval contained the louder stimulus. The next trial began 400 ms after the response. The level of the variable sound initially was set 10 dB above the expected equal-loudness level and was adjusted according to a simple up–down procedure (Jesteadt, 1980). The step size was 5 dB until the second reversal, after which it was reduced to 2 dB. This procedure converges on the 50% point on the psychometric function (Levitt, 1971). Ten interleaved tracks were used to obtain concurrent loudness matches at five levels with the pure tone varied and five levels with the complex varied. On each trial, the track was selected at random from all tracks that had not been completed, which occurred after nine reversals. The equal-loudness level for one track was calculated as the average of the levels at the last four reversals. Three such matches were obtained for each subject and condition. For further discussion of roving-level loudness matching procedures, see Buus *et al.* (1997). The matches were divided into a low and a high range, each consisting of five levels for the condition in which the pure tone was fixed (4, 6, 10, 14, and 18 dB SL for the low range; 18, 22, 27, 33, and 40 dB SL for the high range) and five levels in which the complex was fixed (-1, 1, 4, 7, and 10 dB SL for the low range; 10, 12, 14, 17, and 20 dB SL for the high range). The highest level in the low range and the lowest level in the high range were identical to check if the range of levels presented affected the loudness matches.

3. Stimuli

The multitone complexes were centered at 1600 Hz and the pure tones were the same frequency as the stimuli in the TBOAE measurements (Table I). Both signals consisted of a 20-ms raised-cosine rise and a 20-ms raised-cosine fall with a 480-ms steady-state segment, such that the duration between the half-amplitude points of the rise and fall was 500 ms. The complexes consisted of four tones with an overall spacing of either four or six critical bands in barks (Scharf, 1970; Zwicker and Fastl, 1990). Component frequencies were chosen far enough apart to reduce the likelihood that components would interact in any way. If a particular component, or the pure-tone comparison stimulus, was at a frequency where a sharp peak or trough was seen in the Békésy tracking, it was shifted downward by one-quarter critical band to place it in a more stable area. This helped reduce the possibility of an abnormal relationship between SL and loudness that has been observed at microthreshold irregularities (Kemp, 1979; Cohen, 1982). At least at low levels, equal-SL, rather than equal-SPL, stimuli are equally loud (Hellman and Zwislocki, 1961).

4. Results and discussion

Figure 4 shows the levels of equally loud pure tones and tone complexes for complexes with 4- and 6-bark separations as well as loudness-model fits to those data. Except for S4, the 4- and 6-bark functions vary the same way, indicating that mutual masking of complex components did not occur. Loudness matches at the highest levels measured showed a 5- to 10-dB smaller difference in level between equally loud tones and tone complexes than Buus *et al.* (1998), but those matches showed a greater difference in level than prior work (Zwicker *et al.*, 1957). It is not clear why this discrepancy exists.

Loudness functions were generated from these data on the assumption that the total loudness of the complex is the sum of the loudnesses of the components and that all components at equal SLs are equally loud. These functions were based on the same formula used in Buus *et al.* (1998) and the calculations are explained in more depth there. The loudness (N) of a single tone was calculated as

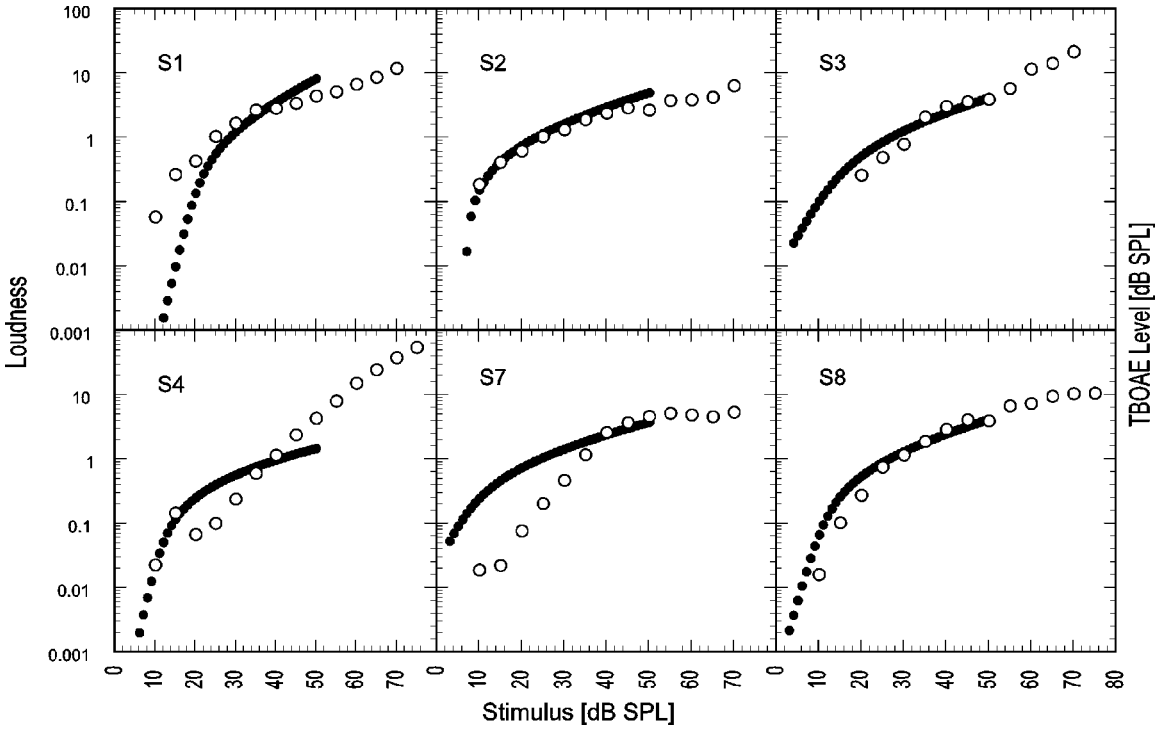


FIG. 5. Loudness functions generated from the loudness-matching model (filled symbols) compared with TBOAE I/O (open symbols). TBOAE functions are on a sliding vertical scale with a fixed number of total units. One major unit on the TBOAE scale is 10 dB.

$$N(\mathrm{SL}) = k([1 + (C * 10^{\mathrm{SL}/10})^{D}]^{B/D} - 1), \quad (2)$$

where k is a scale factor, C is the signal-to-noise ratio at threshold, B and D are fitting parameters calculated for each subject: B is the asymptotic slope of the function at high levels, and D is the asymptotic slope of the function at low levels. The fits were performed to the 4-Bark and 6-Bark complexes simultaneously using a least-squares fit with the same parameters. Figure 4 shows that the loudness model provides excellent predictions for all subjects.

Figure 5 shows loudness derived from the loudness matches plotted on a log scale compared with TBOAE measurements in dB SPL. The TBOAE functions were shifted vertically to obtain a good visual fit. There is excellent agreement for subjects S2, S3, and S8, and good agreement at moderate levels for S1. Two subjects, S4 and S7, showed poor overall agreement between the two measures. One possible explanation for this discrepancy is that the loudness model uses a restrictive formula and forces the loudness function to be a particular shape. Individual variations from that shape may not be well represented by this function and may disguise the similarities with BM I/O. The steeper slope that S4 showed in the TBOAE I/O functions may also explain why mutual masking was seen for this individual in Fig. 4. If the physiological response is greater for that individual than for other subjects, it is likely that the spread of excitation is also wider, making it more likely for component interactions to occur.

In order to obtain an overall picture of the value of this comparison between loudness derived from the tone-to-complex matches and TBOAEs, the average of all six subjects for each measure is shown in Fig. 6. It is clear from Fig. 6 that, on average, there is excellent agreement between the two measures. If loudness functions obtained using tone-to-complex matching are a good representation of BM I/O functions, then it is likely that TBOAE I/O is also closely related to BM I/O, at least for the range of levels examined here.

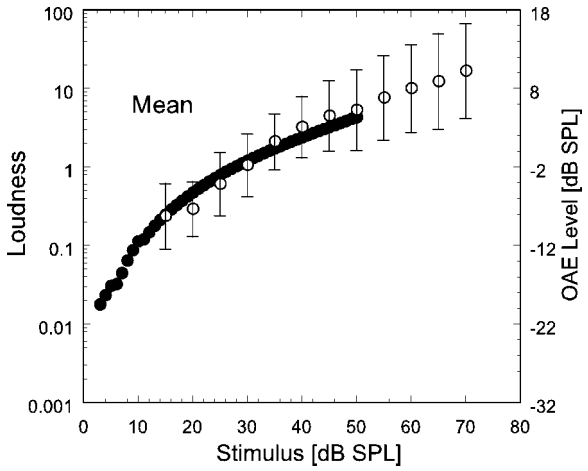


FIG. 6. The mean loudness function generated from the loudness-matching model for six subjects (filled symbols) compared with mean TBOAE I/O (open symbols). The TBOAE function was on a sliding vertical scale with a fixed number of total units. The error bars on the TBOAE I/O function show the standard deviations.

E. Experiment 4—Pulsation threshold

1. Introduction

Pulsation threshold is the lowest level of a masker for which an alternating on-frequency (signal)/off-frequency (masker) tone pair produces a continuous sounding signal, rather than a pulsed one (Houtgast, 1977). It is determined by adjusting an off-frequency tone until the perception changes to pulsed rather than continuous.

Measurements of the BM I/O function (Ruggero *et al.*, 1997) indicate that it is reasonable to assume that the response at a certain place on the BM to an off-frequency tone is linear and the response to an on-frequency tone is nonlinear (Murugasu and Russell, 1995; Ruggero *et al.*, 1997). As the on-frequency tone level increases, the off-frequency tone level necessary to mask it will increase at a slower rate. The shape of the pulsation-threshold function, at least at moderate-to-high levels, should provide an estimate of the BM response. Specifically, because the pulsation threshold is thought to provide a measure of the activity caused by the on-frequency tone, the increase in level of the masker relative to the increase in level of the signal reveals compression (Oxenham and Plack, 1997). For example, if an off-frequency masker is at a level such that a signal is at threshold and the slope of the compressed region is 0.2 (a compression ratio of 5:1), then an off-frequency masker will only need an increase of 1 dB for the signal to remain at threshold despite a 5-dB signal level increase. This is a result of compressive response only occurring at or near the best frequency response point on the BM.

2. Procedure

Pulsation threshold was measured with a single-interval procedure similar to that of Plack and Oxenham (2000). The masker and the signal were alternated three times each per trial. The subject was asked to state whether the signal sounded pulsed or continuous. The masker level was varied while the signal level was held constant at a series of selected levels. Because of the difficult nature of the task, all subjects were provided with a demonstration program at the beginning of the experiment and at any other time that they requested it or expressed confusion. The demonstration program allowed the subject to select and hear either clearly pulsed or clearly continuous sounds. The sensations were achieved by having the masking tone at a very high or very low level for a relatively low-level signal. The subject listened to these demos until accustomed to differentiating the two percepts. Each run consisted of two interleaved tracks that followed a 2-down, 1-up or 2-up, 1-down rule. One track determined the point at which 70.7% ''continuous'' responses were given and the other the point at which 70.7% ''pulsed'' responses were given (Levitt, 1971). The masker-level step size was 5 dB for the first three reversals, and 2 dB thereafter. Each track ended after seven reversals and the mean of the last four reversals was defined as the pulsation threshold. The results from the two tracks were averaged to estimate the point at which pulsed responses and continuous responses were equally likely. If the two tracks differed by more than 10 dB, the run was rerun. This occurred very infrequently. If the adaptive procedure called for a masker level higher than 100 dB SPL then the level was set to and maintained at 100 dB SPL. If the maximum level was presented more than five consecutive times in a particular track, the track was terminated. Each pulsation threshold reported is the mean of three threshold estimates. Whenever possible, masker threshold measurements were made for signals ranging from 20 to 85 dB SPL in 5-dB steps. It was often difficult to obtain measures at the lowest and highest levels due to lack of consistency or inability to mask within the allowed-level range. In these cases, after several attempts were made, the levels were excluded for that subject.

3. Stimuli

The pulsation thresholds were measured using stimuli that began and ended with a 540-ms signal. The initial signal was followed by an alternating sequence of three 240-ms maskers and two 240-ms signals. All tones had 20-ms raised-cosine onsets and offsets. There was a 30-ms overlap at all signal transitions so that no audible gap was present. The signal was at the same frequency as in the TBOAE measurements (see Table I). The masker frequency was 0.4 times the signal frequency.

4. Results and discussion

Figure 7 shows a comparison between the pulsation-threshold measurements and TBOAE data. Several subjects' data showed large variability, which is probably a result of the difficulty of the task.

The pulsation threshold measurements fit very well with the TBOAE I/O data in five of the six subjects. One subject, S3, shows reasonable correspondence, but exhibits more variability than any of the other subjects. The pulsation-threshold data points with the smallest standard errors tend to correspond best with the TBOAE data. Pulsation-threshold measurements could not be performed over the same range of input levels as the TBOAE measurements, so low-level correspondence could not be tested. Nevertheless, strong midlevel agreement is the important metric because pulsation-threshold midlevel measures are an assessment of the compressive portion of the BM I/O function.

III. CONCLUSIONS

Excellent correspondence was seen between TBOAE I/O and BM I/O estimated from loudness functions, derived from temporal integration and tone-to-complex loudness matches, and pulsation thresholds. This indicates that TBOAEs may provide a good individual measure of BM I/O in an objective and rapid manner. In addition, these results show evidence of a wide range of variability in the compressiveness of the BM I/O function. These determinations provide a framework for examining the relationship between subjective psychoacoustic phenomena and objectively measured TBOAEs. Such comparisons may prove invaluable for gaining a greater understanding of the lower levels of the auditory system and yield a framework in which some perceptual phenomena can be examined through a noninvasive physiological means.

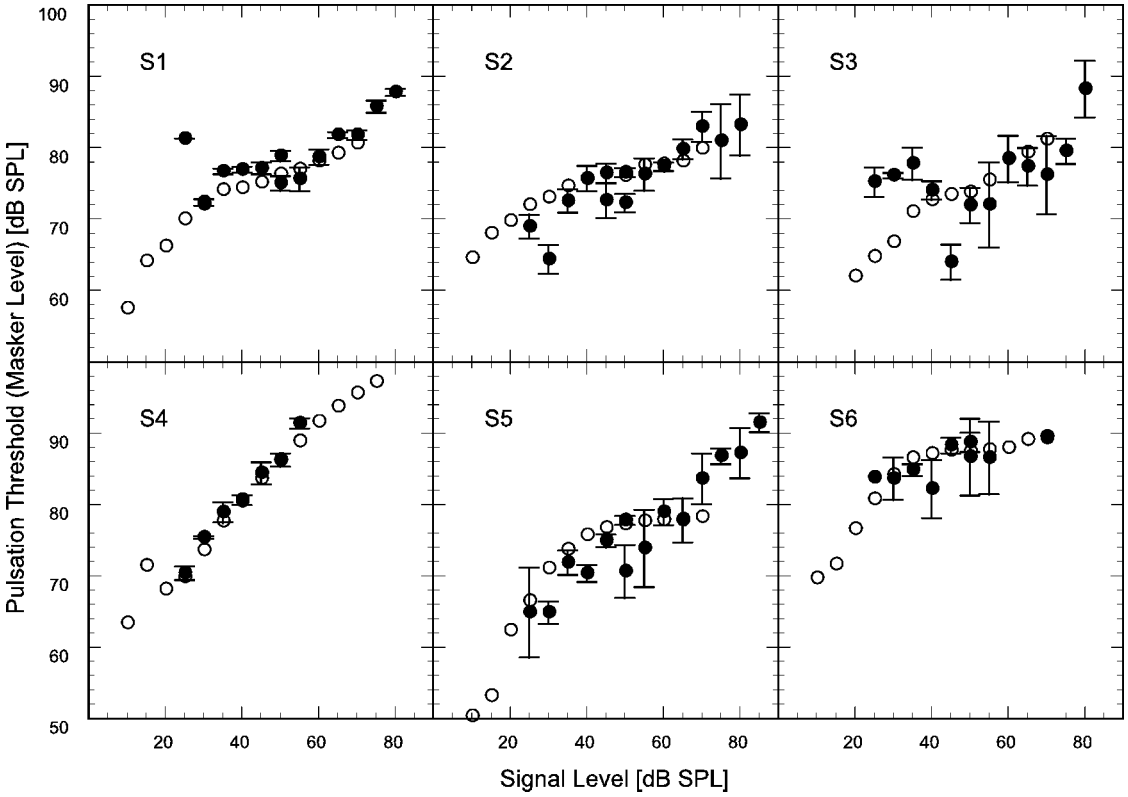


FIG. 7. Pulsation threshold (filled circles) and TBOAE data (open circles) are shown with respect to signal level. Standard errors for the pulsation-threshold measurements are also shown.

ACKNOWLEDGMENTS

Søren Buus contributed substantially to this project and passed away prior to completion of the work. The authors wish to thank Bertram Scharf, Eva Wagner, reviewer Chris Plack, and an anonymous reviewer for helpful comments. This research was supported by NIH/NIDCD Grants Nos. R01DC02241 and R01DC00187.

APPENDIX: QUANTITATIVE ANALYSIS OF CORRESPONDENCE BETWEEN TBOAES AND PSYCHOACOUSTIC MEASURES

Quantitative and statistical analyses of the data provided in this paper in an objective manner are extremely difficult and the results of such analyses vary significantly with small changes to the parameters used for modeling, local minima in linear optimizations, and the range of levels used in the parameter estimation. It would be at least somewhat misleading to give an absolute value for slope or an absolute fitting equation for any of the data sets. In particular, each of the individual experiments has different strengths and weaknesses for use as a metric of BM I/O.

The TBOAEs are highly irregular at levels below the noise floor and are at least somewhat difficult to use for estimation of the low-level BM I/O function; i.e., below the kneepoint. The estimates of slope made in Table II, both above and below the kneepoint, use simply two lines. It is unlikely that this represents the detailed complexity of TBOAE I/O, but in this experiment, measurements were spaced by 5 dB and were therefore not necessarily suitable for determining the exact subtle changes in the slope across the entire range. It would be possible to try to generate slopes using a model of BM I/O, but this would bias the way the data were fit and force the results toward looking like predetermined BM I/O functions. Therefore, data were not fit with a specific model to ensure the objectivity of the comparisons, and the estimates of slope should be considered just that. They are a valuable reference for determining whether the approximate mean slope is in the realm of what would be expected for BM I/O and roughly what a range of slopes for a set of subjects might look like.

The loudness functions derived from temporal integration are derived indirectly based on the equal-loudness-ratio hypothesis. This is a limitation because the hypothesis does not appear to hold at low levels. Owing to this limitation, BM I/O estimates from these loudness functions will be useful primarily for the compressive portion, and comparisons should be limited to this range.

The loudness functions derived from multitone loudness matches are limited because of mutual masking that occurs in the complexes when tone level gets high. This limitation prevents an assessment of the shape of the loudness function much above the kneepoint, thus making it difficult to quantitatively compare with the highly subjective determination of the low-level slope estimate of the TBOAEs.

Despite the difficulty with quantitative comparisons, some measures are provided here as a reference to indicate roughly that the value of these comparisons is not purely subjective. Table III shows the minimized mean-square error differences between TBOAEs and the three psychoacoustic measures for all subjects. The measures are all on different

TABLE III. Mean-square error differences between TBOAEs and the three psychoacoustic measures for all subjects for values normalized to 1.

	Temporal integration (Fig. 3)		Multitone matches (Fig. 5)	Pulsation threshold (Fig. 7)	Mean
	15 to 70 dB	40 to 70 dB	15 to 50 dB		
S1	0.000 98	0.000 07	0.024 27	0.004 10	0.007 35
S2	0.000 89	0.000 87	0.000 31	0.003 04	0.001 27
S3	0.001 97	0.000 47	0.000 96	0.015 76	0.004 79
S4	0.008 23	0.002 70	0.005 06	0.000 26	0.004 06
S5	0.015 50	0.000 44	N/A	0.002 17	0.006 04
S6	0.007 15	0.000 14	N/A	0.002 19	0.003 16
S7	N/A	N/A	0.010 66	N/A	0.010 66
S8	N/A	N/A	0.002 44	N/A	0.002 44
Mean	0.005 79	0.000 78	0.007 28	0.004 59	0.004 97

scales, so the ranges have all been normalized to unity while maintaining the same number of log units. Only one free parameter, vertical offset, was used for minimization. For the experiment and condition listed in each column, a separate minimization was performed. It is expected that the conditions with fewer data points will have a smaller mean because of the free parameter. Table III provides a relative charting of the correspondence between each psychoacoustic measure and TBOAEs for each subject.

Table IV shows slope estimates and their correlations with TBOAE I/O derived from the measures of the slope of the compressive region for the two appropriate psychoacoustical measures and the correlation of the slopes with TBOAE slopes. The loudness functions derived from multitone summation were not appropriate for estimating the slope far above the kneepoint; therefore, this comparison has been omitted.

Both measures show a correlation with TBOAE I/O slope and relatively close estimates of the slope itself. The pulsation threshold, which is a more direct measure, overestimates the slope slightly (0.32), but is highly correlated with TBOAE I/O (0.87).

These quantitative comparisons of the different measures provide corroboratory evidence that the conclusions determined from the visual comparison of the different measures were valid.

TABLE IV. Slope estimates and their correlations with TBOAE I/O derived from the measures of the slope of the compressive region.

	TBOAE (Fig. 1)	Temporal integration (Fig. 3)	Pulsation threshold (Fig. 7)
S1	0.19	0.25	0.16
S2	0.13	0.24	0.19
S3	0.29	0.19	0.34
S4	0.51	0.27	0.66
S5	0.08	0.16	0.28
S6	0.08	0.12	0.26
Mean	0.21	0.21	0.32
Correlation	⋯	0.65	0.87

[1]Only S4, out of nine subjects in Florentine *et al.* (2001), exhibited loudness growth for tones of different durations that was not consistent with the equal-loudness-ratio hypothesis. Because the loudness functions shown here are based on this assumption, it is not surprising that the least correspondence is seen between this subject's TBOAEs and temporal-integration-derived loudness functions.

ANSI (**1989**). ANSI-S3-6-1989, "Specifications for Audiometers," (American National Standards Institute, New York).

Buus, S. (**1999**). "Loudness functions derived from measurements of temporal and spectral integration of loudness," in *Auditory Models and Nonlinear Hearing Instruments*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen (GN ReSound, Taastrup, Denmark).

Buus, S., and Florentine, M. (**2001**). "Modifications to the power function for loudness," in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin).

Buus, S., Florentine, M., and Poulsen, T. (**1997**). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am. **101**, 669–680.

Buus, S., Müsch, H., and Florentine, M. (**1998**). "On loudness at threshold," J. Acoust. Soc. Am. **104**, 399–410.

Buus, S., Obeling, L., and Florentine, M. (**2001**). "Can basilar-membrane compression characteristics be determined from distortion-product otoacoustic-emission input–output functions in humans?," 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, Netherlands).

Cohen, M. F. (**1982**). "Detection threshold microstructure and its effects on temporal integration data," J. Acoust. Soc. Am. **71**, 405–409.

Epstein, M., Buus, S., and Florentine, M. (**2004**). "The effects of window delay, delinearization, and frequency on tone-burst otoacoustic emission input/output measurements," J. Acoust. Soc. Am. **116**, 1160–1167.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.

Florentine, M., Epstein, M., and Buus, S. (**2001**). "Loudness functions for long and short tones," in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin).

Goodman, S. S., Withnell, R. H., De Boer, E., Lilly, D. J., and Nuttall, A. L. (**2004**). "Cochlear delays measured with amplitude-modulated tone-burst-evoked OAEs," Hear. Res. **188**, 57–69.

Hellman, R. P., and Zwislocki, J. J. (**1961**). "Some factors affecting the estimation of loudness," J. Acoust. Soc. Am. **33**, 687–694.

Houtgast, T. (**1977**). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker," J. Acoust. Soc. Am. **62**, 409–415.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

Kemp, D. T. (**1979**). "The evoked cochlear mechanical response and the auditory microstructure—evidence for a new element in cochlear mechanics," in *Models of the Auditory System and Related Signal Processing Techniques*, edited by M. Hoke, and E. de Boer.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
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.
Neumann, J., Uppenkamp, S., and Kollmeier, B. (**1997**). ''Relations between notched-noise suppressed TEOAE and the psychoacoustical critical bandwidth,'' J. Acoust. Soc. Am. **101**, 2778–2788.
Norton, S. J., and Neely, S. T. (**1987**). ''Tone-burst-evoked otoacoustic emissions from normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1860–1872.
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. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.
Plack, C. J., and Oxenham, A. J. (**2000**). ''Basilar-membrane nonlinearity estimated by pulsation threshold,'' J. Acoust. Soc. Am. **107**, 501–507.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. I.
Schlauch, R. S., DiGiovanni, J. J., and Ries, D. T. (**1998**). ''Basilar membrane nonlinearity and loudness,'' J. Acoust. Soc. Am. **103**, 2010–2020.
Shera, C. A., and Guinan, Jr., J. J. (**2003**). ''Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning,'' J. Acoust. Soc. Am. **113**, 2762–2772.
Withnell, R. H., and Yates, G. K. (**1998**). ''Onset of basilar membrane nonlinearity reflected in cubic distortion tone input–output functions,'' Hear. Res. **123**, 87–96.
Yates, G. K. (**1990**). ''The basilar membrane input–output function,'' in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. M. Matthews, M. A. Ruggero, and C. R. Steele (Springer, New York).
Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics—Facts and Models* (Springer, New York).
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.
Zwislocki, J. J. (**1965**). ''Analysis of some auditory characteristics,'' in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York).

Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch

Michael J. Hautus[a] and Blake W. Johnson
Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand

(Received 29 June 2004; revised 25 September 2004; accepted 7 October 2004)

The cortical mechanisms of perceptual segregation of concurrent sound sources were examined, based on binaural detection of interaural timing differences. Auditory event-related potentials were measured from 11 healthy subjects. Binaural stimuli were created by introducing a dichotic delay of 500-ms duration to a narrow frequency region within a broadband noise, and resulted in a perception of a centrally located noise and a right-lateralized pitch (dichotic pitch). In separate listening conditions, subjects actively discriminated and responded to randomly interleaved binaural and control stimuli, or ignored random stimuli while watching silent cartoons. In a third listening condition subjects ignored stimuli presented in homogenous blocks. For all listening conditions, the dichotic pitch stimulus elicited an object-related negativity (ORN) at a latency of about 150–250 ms after stimulus onset. When subjects were required to actively respond to stimuli, the ORN was followed by a P400 wave with a latency of about 320–420 ms. These results support and extend a two-stage model of auditory scene analysis in which acoustic streams are automatically parsed into component sound sources based on source-relevant cues, followed by a controlled process involving identification and generation of a behavioral response.
[DOI: 10.1121/1.1828499]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.66.Lj [WPS] Pages: 275–280

I. INTRODUCTION

Dichotic pitch is a binaural psychoacoustic phenomenon that involves the perception of pitches from stimuli that contain no monaural cues to pitch (Cramer and Huggins, 1958; Bilsen, 1976). Dichotic pitch can be produced by presenting listeners with two broadband noises with interaurally identical amplitude spectra but with a specific interaural lag over a narrow frequency band (Dougherty *et al.*, 1998). The interaurally shifted frequency band becomes perceptually segregated from the noise, and the resulting pitch has a tonal quality associated with the center frequency of the dichotically delayed portion of the spectrum. Because the stimuli are discriminable solely by the interaural lag but are otherwise acoustically identical, the perception of dichotic pitch must ultimately depend upon the binaural fusion of interaural time differences (ITDs) within the central auditory system. The phenomenon of dichotic pitch demonstrates that the human auditory system applies its exquisite sensitivity for the fine-grained temporal structure of sounds to the perceptual segregation, localization, and identification of concurrently presented sound sources. The dichotic pitch illusion has a parallel in the visual system, which in a closely analogous manner can utilize retinal disparities in random dot stereograms to achieve stereopsis (Julesz, 1971).

While it is known that ITDs are first extracted by nuclei in the auditory brainstem (Goldberg and Brown 1969; Yin and Chan, 1990), very little is known about the higher level mechanisms that use such cues to achieve perceptual segregation of sound sources. In a recent investigation, Johnson *et al.* (2003) examined cortical processes associated with perception of a dichotic pitch by measuring event-related brain potentials (ERPs). They found that dichotic pitch stimuli elicited a late negative potential, termed the LN, in a latency region of about 150–300 ms after stimulus onset.

Interestingly, the LN response closely resembles a response associated with perceptual segregation based upon monaural cues—the inharmonicity of one component of a complex sound composed of multiple harmonics. Alain *et al.* (2002) measured ERPs from subjects presented with sounds containing tuned or mistuned harmonics. In two different listening conditions the subjects either actively attended to the sounds and indicated their perceptions (a single sound versus two sounds) with a button press, or ignored the acoustic stimuli while watching a silent movie. The perception of a mistuned harmonic as a separate sound was associated with a negative wave that has a peak latency of about 160 ms, which these authors termed the ''object-related negativity'' (ORN). The ORN was elicited in both active and passive listening conditions, while a later P400 wave was elicited by the mistuned harmonic stimuli only when subjects actively attended to these stimuli. The authors concluded that the two sequential components support a two-stage model of auditory scene analysis. In this model, complex sounds are first parsed based on an automatic process, reflected by the ORN and involving a detection of a mismatch between the harmonic template extracted from the incoming stimulus and the harmonic frequency expected based upon the fundamental of the complex sound. The second stage, indexed by the P400, is invoked only when subjects are actively attending to the stimuli and seems to reflect controlled processes responsible for identification of stimuli and selection of behavioral responses.

The present study was designed to investigate the possi-

[a]Electronic mail: m.hautus@auckland.ac.nz

bility that the mechanisms described in this two-stage model are also deployed for the perception of dichotic pitch. Although the mistuned harmonic stimuli and the dichotic pitch stimuli have radically different compositions (pure tones versus white noise) and critical acoustic cues (monaural versus binaural), the overall perceptual task—segregating two concurrently presented acoustic objects—is the same. If the ERP components elicited by dichotic pitch are functionally comparable to those described by Alain *et al.* (2002), then these are likely to reflect fairly general mechanisms of auditory stream segregation that can broadly utilize a range of cues that can help to separate simultaneous acoustic events. We evaluated two main properties of the ERPs elicited by dichotic pitch: (1) the extent to which they reflect automatic (preattentive) versus controlled (attention-demanding) processes, and (2) the extent to which they depend on the perceptual context of the stimuli, determined by the frequency of occurrence of the dichotic pitch stimuli within the context of comparison stimuli.

II. METHOD

A. Subjects

Eleven volunteers (eight right handed; four female), aged between 22 and 48 years (mean=29 years), participated. All subjects reported that they were free of audiological or neurological illness. Any subject who had not participated in previous dichotic pitch research was screened to ensure that they could detect dichotic pitch. The University of Auckland Human Participants Ethics Committee approved all procedures.

B. Stimuli

The acoustic stimuli were comparable to those employed by Johnson *et al.* (2003). In a procedure similar to that described by Dougherty *et al.* (1998), two independent broadband Gaussian noise processes of 500-ms duration were constructed digitally using LabView software at a sampling rate of 44.1 kHz. One of the noise processes was bandpass filtered using a fourth-order Butterworth filter with a center frequency of 600 Hz, a 3-dB bandwidth of 50 Hz, and corner frequencies of 575 and 625 Hz. The other noise process was notch filtered, employing the same corner frequencies as for the bandpass filter. The two filters were designed to be complementary so that the sum of the filter functions for the notch and bandpass filters was equal to one for all frequencies. Consequently, when the two noise processes were added subsequent to filtering, the resulting noise process had a flat spectrum.

To produce the dichotic pitch stimulus, the bandpass filtered noise process was reproduced and one copy was temporally delayed by 0.45 ms. Two spectrally identical noise processes were obtained by adding the notch-filtered noise process to each copy of the bandpass-filtered noise process. These were then low-pass filtered with an upper cutoff frequency of 1200 Hz before being windowed with 4-ms rise and fall times. The stimuli were generated on two channels of a 16-bit converter (National Instruments, PCI 6052E). The level of the stimulus was adjusted using programmable attenuators to yield 70 dB SPL at the eardrum. Insert earphones (Etymotic Research, Model ER2) delivered the stimuli to the listeners. The interstimulus interval was drawn from a rectangular distribution of durations between 1500 and 3500 ms.

Johnson *et al.* (2003) found a psychoacoustic right-ear bias in dichotic pitch detection during their initial screening procedure. Consequently, in the present study, the noise process that contained the temporally delayed band of noise was presented to the left ear of the subject and the other process to their right ear. We refer to this stimulus configuration as the *pitch* stimulus, which gives rise to two distinct percepts: a ''ball'' of noise perceived in the center of the head and a low-frequency pitch perceived to the right of auditory space.

The control stimulus was produced in the same manner as the pitch stimulus, except that no delay was introduced for the narrow-band component. Consequently, the stimulus at each ear was identical and the resulting percept consisted only of a centrally located noise and did not have any localized pitchlike characteristics.

C. Screening procedure

The purpose of this task was to establish the ability of subjects to detect dichotic pitch. The pitch and control stimuli were presented in a two-alternative forced-choice task. By pressing one of two keys on a computer keyboard, listeners stated whether the pitch stimulus was presented in the first or second observation interval. Feedback in the form of a flashing LED indicated whether their response was correct or incorrect. To participate further in the project, subjects were required to get more than 71% correct in this task.

D. EEG

The EEG component of this experiment comprised three listening conditions: (*i*) an ''*active-randomized*'' condition (AR), which required attentive listening and responding to randomly interleaved control and dichotic pitch stimuli; (*ii*) a ''*passive randomized*'' condition (PR), in which subjects watched silent video presentations of cartoons and were instructed to ignore randomly interleaved control and dichotic pitch stimuli; and (*iii*) a ''*passive blocked*'' (PB) condition, in which subjects watched cartoons and were instructed to ignore the acoustic stimuli, which were presented as long uniform blocks of control stimuli and dichotic pitch stimuli.

In the AR condition subjects listened to 512 stimuli, presented as four blocks of 128, in a randomized order with equal frequencies of each type (pitch and control). After each stimulus presentation subjects were required to indicate on a keyboard whether a dichotic pitch or a control stimulus had been presented. This is equivalent to a single-interval (yes/no) task without feedback. The ability of the subject to detect the presence of the pitch stimulus was subsequently evaluated by calculating the percentage of correct responses over the four blocks of trials. Because the task differed from the screening procedure, all subjects participated in some practice trials before data collection began to ensure their familiarity with the stimuli and with the nature of the task.

The PR condition consisted of the same number of trial presentations and block structure as the AR condition. For this condition subjects were instructed to attend to a series of silent cartoons displayed on the screen in front of them and not to the auditory stimuli.

In the PB condition all stimuli within a block of presentations were of the same type, that is, either all control or all dichotic pitch stimuli. Each stimulus type was presented as two blocks of 128 stimuli, and the block types were randomly interleaved. For this condition subjects were also required to attend to silent cartoons, and not to the auditory stimuli.

All subjects were required to sit for the duration of the recording sessions for all four conditions. All blocks within a condition were presented consecutively. The order of the conditions for each subject was randomized.

EEG recordings were carried out in an electrically shielded room using Electrical Geodesics Inc. 128-channed Ag/AgCl electrode nets (Tucker, 1993). EEG was recorded continuously (250-Hz sampling rate; 0.1–100-Hz analogue bandpass), with Electrical Geodesics Inc. amplifiers (200 MΩ input impedance). Electrode impedances were kept below 40 Ω, an acceptable level for this system (Ferree *et al.*, 2001). EEG was acquired using a common vertex (Cz) reference.

E. Analysis

Following completion of data collection, EEG files were segmented into 600-ms epochs, including a 100-ms prestimulus baseline. Ocular artifacts were removed from individual trial epochs using procedures from Jervis *et al.* (1985). ERPs were rereferenced to the mean reference (recovering the Cz electrode) and digitally filtered with a bidirectional three-pole Butterworth filter (Alarcon *et al.*, 2000) with lower and upper corner frequencies of 0.1 and 30 Hz, respectively. ERPs for individual listeners were combined to produce grand-averaged ERPs for each of the six (2 stimulus types ×3 listening conditions) experimental conditions. ERP waveforms were analyzed by computing mean voltages over three time intervals: N1 (88–136 ms), ORN (150–250 ms) and P400 (320–420 ms).

Behavioral responses during the AR condition were quantified by percent correct for each stimulus type and analyzed with a repeated measures analysis of variance (ANOVA). Statistical analysis of ERP amplitude data were performed with separate repeated measures ANOVAs (with stimulus type and listening condition as factors) on each of the three ERP components. Where the assumption of sphericity was violated the degrees of freedom were adjusted with the Greenhouse-Geisser epsilon. For each ERP component, data were collapsed across the ten electrode positions (Oostenveld and Praamstra, 2001) corresponding to that component's maximum amplitude distribution. Thus, for the N1 component, data were collapsed across ten fronto-central electrodes (FFC2h,Fcz,FCC1h,FFC1h,FC1,CCP1h,CCP2h, FCC2h,FC2,Cz); for the ORN component, data were collapsed across ten frontal-left electrodes (FCC1h,FC1, FFC3h,FC3,C1,CCP1h,FCC5h,C3,CP1,CP3); and for the P400 component, data were collapsed across ten frontal-right electrodes (FCC2h,FC2,FFC4h,FC4,C2,CCP2H,FCC6h,C4, CP2,CP4) homologous to the frontal-left electrodes. Statistical analyses of ERP spatial distributions were performed using a subset of 68 electrode positions chosen for their correspondence to standard International 10-10 positions (AF8,AF4,F2,Fcz,Fz,FP2,Fpz,Afz,F1,FC1,FP1,AF3,F3,AF7, F5,FC3,C1,F7,FC5,C3,CP1,F9,FT7,C5,CP3,Ft9, T7, Cp5, Tp7,P5, P3,CpZ,TP9,P9,P7,PO3,P1,Pz,Po7, Poz,O1,O2,P2, PO8,PO4,P4,CP2,P8,P6,CP4,P10,TP8,CP6,TP10,C6,C4,C2, T8,FC4,Ft10,FT8,FC6,FC2,F10,F8,F6,F4,Cz). For each spatial analysis, data were first normalized (McCarthy and Wood, 1985) to eliminate amplitude differences between conditions and thus to focus the ANOVA on amplitude differences across electrode locations.

III. RESULTS

A. Behavioral data

Subjects performed the discrimination task in the AR condition with an overall mean accuracy of 90.3% (range 73%–98%). There was no significant difference in accuracy for responses to the control and pitch stimuli (mean 90.6% and 93.7%, respectively).

B. Electrophysiological data

Figure 1 shows the grand averaged ERPs elicited by noise and dichotic pitch stimuli in each of the three listening conditions. For the AR condition the ERPs include both correct and incorrect trials as high accuracy rates make it impractical to consider these cases separately. Stimuli in all experimental conditions elicited a P-N-P complex of waves peaking at latencies of about 60 (P1), 110 (N1), and 200 (P2) ms, respectively. An additional negative peak at a latency of 300 ms (N2) was prominent in the active listening condition for both noise and pitch stimuli. This N2 peak was much reduced in amplitude in the passive listening conditions.

The dichotic pitch stimulus elicited two ERP components with distinct time courses and topographic distributions. For all three listening conditions, the pitch ERPs showed a negative displacement (ORN) that is most evident at the P2 peak. The time course of the ORN component is best visualized in the difference waveforms at the bottom of Fig. 1 obtained by subtracting the pitch ERPs from the noise ERPs. The ORN amplitude was maximal during a time window of about 150–350 ms and over the left fronto-central electrodes (Figs. 1 and 3).

A second dichotic pitch component (P400) was apparent only in the active listening condition. The P400 was manifest as a positive displacement that was maximal in amplitude during a time window of about 320–420 ms and showed a lateralized distribution over the right central electrodes (Figs. 1 and 3).

Figure 2 summarizes the experimental effects on the N1, ORN, and P400 components. The N1 component showed a distinct increase in amplitude for the active listening conditions in comparison to both of the passive listening conditions, and ANOVA confirmed a significant main effect of listening condition [$F(2,20)=4.89$, $p=0.019$, $\varepsilon=0.91$]. Planned contrasts of the different listening conditions con-

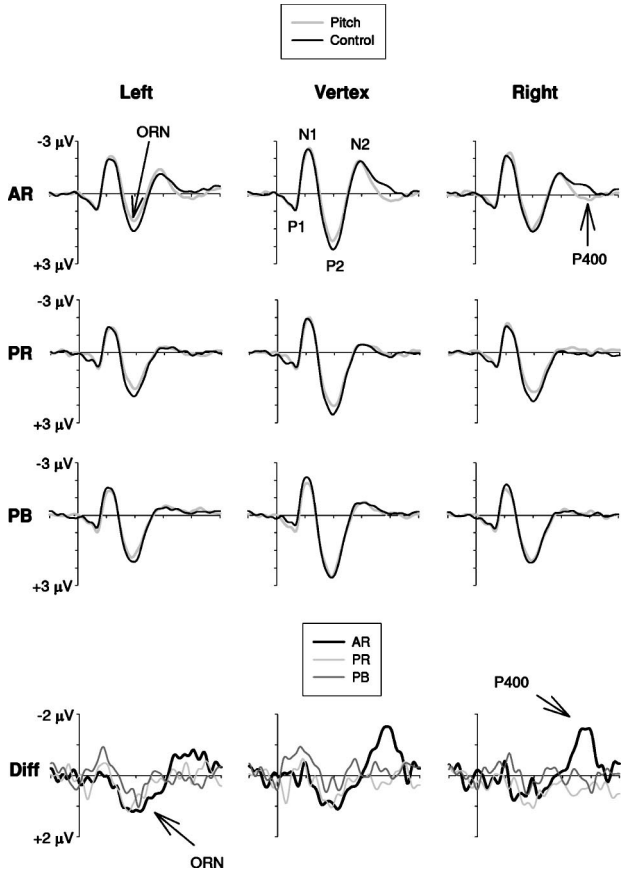


FIG. 1. Top three rows: grand-averaged ERPs evoked by control and pitch stimuli in each of the listening conditions. Bottom row: difference waveforms computed by subtracting the control from the pitch ERPs for each of the listening conditions. Left electrode=FC3; right electrode=FC4; vertex electrode=Cz. Vertical scale indicates 1 μV/division (negative up). Horizontal scales shows 100 ms/division. Vertical scalebar intersects the horizontal axis at time of stimulus onset.

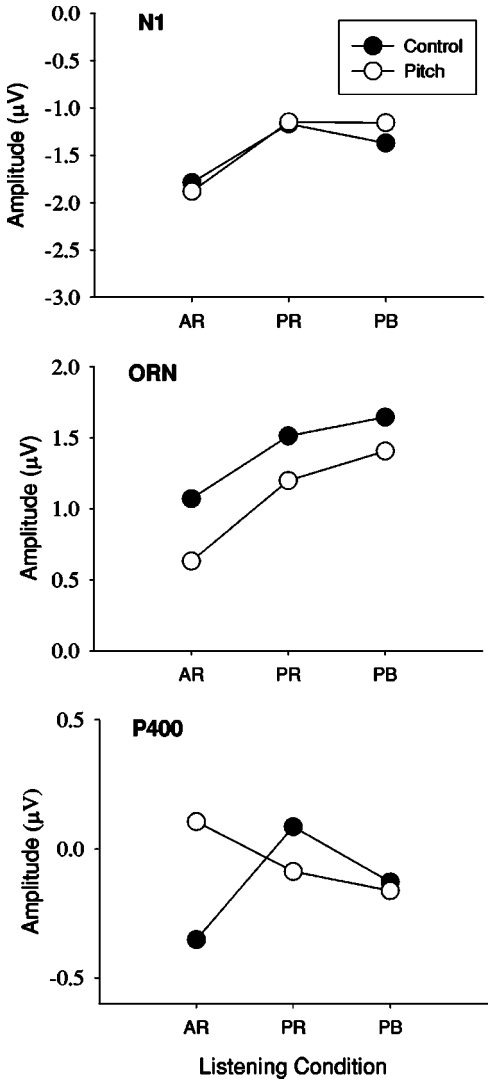


FIG. 2. Effect of listening condition and stimulus type on three different ERP components N1, ORN and P400.

firmed that the active listening N1 amplitude was significantly greater than both of the passive listening conditions [$F(1,10)=6.82$, $p=0.03$; and $F(1,10)=5.65$, $p=0.04$ for random and blocked passive listening conditions respectively], but the passive conditions were not significantly different from one another [$F(1,10)<1$].

During the analysis window of the ORN, P2 amplitude was affected by listening condition [for listening condition, $F(2,20)=9.11$, $p=0.002$, $\varepsilon=0.802$]. Planned contrasts of the different listening conditions confirmed that the AR amplitude was significantly different from both the PR and PB amplitudes [$F(1,10)=6.59$, $p=0.03$, and $F(1,10)=10.03$, $p=0.002$, respectively], but that the passive conditions were not significantly different from each other. However, the ORN elicited by the pitch stimuli was not affected by listening condition, as evidenced by the constant amplitude displacement between the control and pitch ERPs across listening conditions. ANOVA confirmed a significant main effect of stimulus [$F(1,10)=9.58$, $p=0.011$] but no significant stimulus$\times$listening condition interaction [$F(2,20)<1$].

Figure 2(c) shows that only the active listening condition elicited a P400 effect for the dichotic pitch stimulus; for both of the passive listening conditions the noise-pitch amplitude difference was much smaller and in the opposite direction. ANOVA confirmed a highly significant stimulus $\times$listening condition interaction [$F(2,20)=7.13$, $p=0.005$, $\varepsilon=0.681$].

C. Topographic analyses

The voltage isocontour maps of Fig. 3 compare the topographic distributions of the N1, ORN, and P400 components for the control and pitch conditions. The N1 component showed an amplitude distribution over fronto-central electrodes, and this distribution showed relatively little difference between stimulus conditions. The ORN component showed a comparable fronto-central distribution, however the difference map shows that the increased negativity in the pitch condition was slightly lateralized to the left hemisphere. Statistical analysis confirmed that the control and pitch topographies were significantly different [$F(67,670)=2.33$, $p=0.02$, $\varepsilon=0.137$]. For the P400, the difference maps show that the pitch condition was more positive over right central locations. Statistical analysis confirmed that the control and pitch topographies were significantly different [$F(67,670)=3.33$, $p=0.003$, $\varepsilon=0.123$].

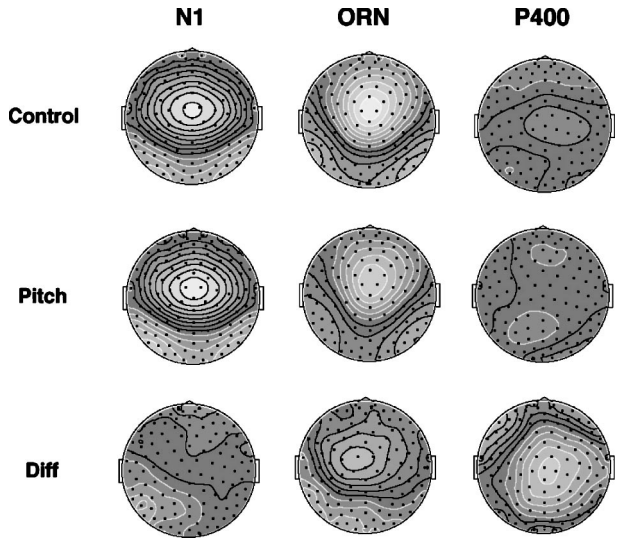


FIG. 3. Surface topographies of the N1, ORN and P400 components elicited by control (top row) and pitch (middle row) stimuli in the AR listening condition. Bottom row shows topography of difference ERPs computed by subtracting the pitch ERPs from the control ERPs. Black dots indicate electrode positions. Lighter colors indicate larger amplitudes. Black contours indicate negative polarity, white contours indicate positive polarity. Contour spacing is 0.2 μV.

IV. DISCUSSION

Because the noise sequences employed in the present experiment contained no monaural cues to pitch, the perception of dichotic pitch must be based on the extraction of binaural cues by the auditory system. In the mammalian auditory system, the first auditory structure to receive binaural inputs is the superior olivary complex in the brainstem (Goldberg and Brown, 1969; Yin and Chan, 1990). Evidence from studies of neurons in the superior olive of the cat indicate that these cells are tuned to ITDs for specific frequency regions and behave as if they are computing interaural cross-correlations for a given frequency (Yin and Chan, 1990). This is a plausible physiological system for the extraction of frequency-specific ITDs that are presumably utilized by higher levels of the auditory system for the perceptual decomposition and spatial localization of the broad spectrum noise sequences used in the present study. In real-life situations this arrangement would be useful for the detection of sounds in noisy environments and, more generally, for the perceptual segregation of concurrently presented sound sources.

The present results show that the perception of dichotic pitch was associated with two distinct ERP components indexing activity in relatively high cortical levels of the auditory system. The first component, the ORN, occurred at a latency of about 200 ms after stimulus onset and was elicited both when subjects were actively attending and responding to the stimuli, and when they directed attention instead to the video cartoons. This indicates that the ORN component reflects relatively automatic processes that are independent of the allocation of attention. The latency of the ORN overlaps with those of several other classes of frontally distributed late negative potentials elicited by acoustic stimuli, notably the mismatch negativity (MMN) and the ORN elicited by mistuned harmonic stimuli (Alain *et al.*, 2001, 2002). The present results argue against the possibility that our ORN component is a MMN. While the MMN is also a largely preattentive component, its most salient functional property is that it is elicited only by rare deviant stimuli in the context of a sequence of identical standard stimuli. In contrast, the ORN of the present study was elicited independently of listening condition, including when stimuli were presented in entirely homogeneous blocks.

On the other hand, the functional properties of the ORN described here closely resemble those of the ORN elicited by complex harmonic stimuli containing a mistuned harmonic (Alain *et al.*, 2001, 2002). The ORN to mistuned harmonic stimuli has a peak latency of about 160 ms and a frontocentral scalp distribution. Like the dichotic pitch ORN, the ORN to mistuned harmonic stimuli is elicited independently of whether subjects are actively attending to the stimuli or are attending to a silent video (Alain *et al.*, 2002), and further, the ORN can be elicited by mistuned harmonics regardless of whether they have a frequent or infrequent occurrence in the sequence of presented stimuli (Alain *et al.*, 2001).

These results provide good support for the conclusion that the negative waves elicited by mistuned harmonic stimuli and dichotic pitch stimuli index a comparable phase of auditory processing. However, two points of dissimilarity should be noted. First, the peak latency of the dichotic pitch response (about 200 ms) is somewhat later than the latency reported for the mistuned harmonic response [about 160 ms (Alain *et al.*, 2002)]. Early electrophysiological studies using abrupt interaural time changes in continuous acoustic streams showed that binaurally evoked potentials are delayed in latency on the order of 30–40 ms in comparison to those evoked by monaurally detectable stimulus onsets, likely because the restricted acoustic cues require longer processing times (Halliday and Callaway, 1978; McEvoy *et al.*, 1990). A binaural processing delay of this order neatly accounts for the longer latency of the dichotic pitch response.

A second point of difference is that both our present results and those of our previous study show that the dichotic pitch ORN was lateralized over the left frontal region, whereas the mistuned harmonic ORN has a central midline distribution. This topographic difference can likely be accounted for by the fact that we have to date used only dichotic pitch stimuli with a right-ear-leading interaural lag, resulting in a perception of a pitch that is localized to the right of auditory space. It is likely that this accounts for the left hemisphere lateralization of the response since the human auditory cortex analyses sound azimuth (based on both binaural and monaural localization cues) mainly in the hemisphere contralateral to the sound (Fujiki *et al.*, 2002).

In another striking point of similarity, both the mistuned harmonic stimuli and the dichotic pitch stimuli elicit a second component, the P400. As with our P400, the P400 described by Alain *et al.* (2002) was lateralized to the right hemisphere, and was elicited only in an active listening condition, when subjects were required to actively discriminate between tuned and mistuned stimuli.

V. CONCLUSION

Our results show that the perception of dichotic pitch is associated with two cortical processing stages—indexed by the ORN and the P400—that are functionally comparable to those elicited by the mistuned harmonic stimuli used by Alain *et al.* (2001, 2002). Since the physical compositions and critical cues of these two classes of sounds are radically dissimilar, it is reasonable to conclude that these processing events are more related to the overall perceptual problem posed by both sounds: that of partitioning a complex acoustic input stream into two distinct perceptual objects. Alain *et al.* (2002) proposed a two-stage model in which the ORN component indexes a transient automatic mismatch process between the harmonic template extracted from the incoming stimulus and the harmonic frequency expected based upon the fundamental of the complex sound, while the P400 indexes controlled processes involved in the identification of the stimuli and generation of the appropriate response. In the case of dichotic pitch, however, a neural mismatch response must be generated on the basis of location, since the noise processes by definition have no harmonic structure to match to. This suggests that the ORN reflects the activity of fairly general mechanisms of auditory stream segregation that can broadly utilize a range of cues to parse simultaneous acoustic events.

Consistent with what one would expect of mechanisms that play such a basic role in auditory perception, the present results and those of previous researchers (Alain *et al.*, 2002; Alain and Izenberg, 2003) suggest that the ORN indexes relatively automatic processes that function independently of attention. Such automaticity points to relatively low levels of the auditory system, and indeed physical modeling of the generators of the auditory P1-N1-P2 complex indicates that these waves are best modeled by current sources in or near the primary auditory cortices (Picton *et al.*, 1999; Scherg *et al.*, 1989). More evidence for the involvement of primary auditory cortex in stream segregation comes from a recent study by Dyson and Alain (2004), who reported that the first reliable effect of mistuning harmonic stimuli was a modulation of the amplitude of the Pa peak of the middle latency auditory evoked response. There is good evidence from convergent sources that the Pa wave is generated in primary auditory cortex (Scherg and von Cramon, 1986; Liegois-Chauvel *et al.*, 1991; McGee *et al.*, 1992; Pantev *et al.*, 1995), and the early latency of this response (about 32 ms) is consistent with the preattentive processing of acoustic cues to auditory stream segregation.

ACKNOWLEDGMENTS

This work was supported by University of Auckland Research Committee Grant No. 3603782/9353 to BWJ and MJH. The authors thank Helen Gaeta for useful suggestions during the design of this experiment, Leigh Martelli for contributions to an early draft of the method section and data collection, and Wes Clapp for assistance with the collection of data.

Alain, C., and Izenberg, A. (**2003**). ''Effects of attentional load on auditory scene analysis,'' J. Cogn Neurosci. **15**, 1063–1073.

Alain, C., Arnott, S. R., and Picton, T. W. (**2001**). ''Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials,'' J. Exp. Psychol. Human. **27**(5), 1072–1089.

Alain, C., Schuler, B. M., and McDonald, K. L. (**2002**). ''Neural activity associated with distinguishing concurrent auditory objects,'' J. Acoust. Soc. Am. **111**, 990–995.

Alarcon, G., Guy, C. N., and Binnie, C. D. (**2000**). ''A simple algorithm for digital three-pole Butterworth filter of arbitrary cut-off frequency: application to digital electroencephalography,'' J. Neurosci. Methods **104**, 35–44.

Bilsen, F. A. (**1976**). ''A pronounced binaural pitch phenomenon,'' J. Acoust. Soc. Am. **59**, 467–468.

Cramer, E. M., and Huggins, W. H. (**1958**). ''Creation of pitch through binaural interaction,'' J. Acoust. Soc. Am. **30**, 858–866.

Dougherty, R. F., Cynader, M. S., Bjornson, B. H., Edgell, D., and Giaschi, D. E. (**1998**). ''Dichotic pitch: a new stimulus distinguishes normal and dyslexic auditory function,'' NeuroReport **9**(13), 3001–3005.

Dyson, B. J., and Alain, C. (**2004**). ''Representation of concurrent acoustic objects in primary auditory cortex,'' J. Acoust. Soc. Am. **115**, 280–288.

Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (**2001**). ''Scalp electrode impedance, infection risk, and EEG data quality,'' Clin. Neurophysiol. **112**(3), 536–544.

Fujiki, N., Riederer, K. A., Jousmaki, V., Makela, J. P., and Hari, R. (**2002**). ''Human cortical representation of virtual auditory space: differences between sound azimuth and elevation,'' Eur. J. Neurosci. **16**(11), 2207–2213.

Goldberg, J. M., and Brown, P. B. (**1969**). ''Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization,'' J. Neurophysiol. **32**(4), 613–636.

Halliday, R., and Callaway, E. (**1978**). ''Time shift evoked potentials (TSEPs): method and basic results,'' Electroencephalogr. Clin. Neurophysiol. **45**(1), 118–121.

Jervis, B. W., Nichols, M. J., Allen, E. M., Hudson, N. R., and Johnson, T. E. (**1985**). ''The assessment of two methods for removing eye movement artefact from the EEG,'' Electroencephalogr. Clin. Neurophysiol. **61**(5), 444–452.

Johnson, B. W., Hautus, M. J., and Clapp, W. C. (**2003**). ''Neural activity associated with binaural processes for the perceptual segregation of pitch,'' Clin. Neurophysiol. **114**(12), 2245–2250.

Julesz, B. (**1971**). *Foundations of Cyclopean Perception* (Univ. of Chicago, Chicago).

Liegois-Chauvel, C., Musolino, A., and Chauvel, P. (**1991**). ''Localization of the primary auditory areas in man,'' Brain **114**, 139–153.

McCarthy, G., and Wood, C. C. (**1985**). ''Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models,'' Electroencephalogr. Clin. Neurophysiol. **62**(3), 203–208.

McEvoy, L. K., Picton, T. W., Champagne, S. C., Kellett, A. J., and Kelly, J. B. (**1990**). ''Human evoked potentials to shifts in the lateralization of a noise,'' Audiology **29**(3), 163–180.

McGee, T., Kraus, N., Littman, T., and Nicol, T. (**1992**). ''Contribution of the medial geniculate body subdivision to the middle latency response,'' Hear. Res. **61**, 147–152.

Oostenveld, R., and Praamstra, P. (**2001**). ''The five percent electrode system for high-resolution EEG and ERP measurements,'' Clin. Neurophysiol. **112**(4), 713–719.

Pantev, C., Bertand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., and Elbert, T. (**1995**). ''Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings,'' Electroencephalogr. Clin. Neurophysiol. **94**, 26–40.

Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., and Trujillo, N. J. (**1999**). ''Intracerebral sources of human auditory-evoked potentials,'' Audiol. Neuro-Otol. **4**, 64–79.

Scherg, M., and von Cramon, D. (**1986**). ''Evoked dipole source potentials of the human auditory cortex,'' Electroencephalogr. Clin. Neurophysiol. **65**, 344–360.

Scherg, M., Vajsar, J., and Picton, T. W. (**1989**). ''A source analysis of the late human auditory evoked potentials,'' J. Cogn Neurosci. **1**, 326–355.

Tucker, D. M. (**1993**). ''Spatial sampling of head electrical fields: the geodesic sensor net,'' Electroencephalogr. Clin. Neurophysiol. **87**(3), 154–163.

Yin, T. C., and Chan, J. C. (**1990**). ''Interaural time sensitivity in medial superior olive of cat,'' J. Neurophysiol. **64**(2), 465–488.

Delay dependence for the origin of the nonlinear derived transient evoked otoacoustic emission

Robert H. Withnell[a] and Sarah McKinley
Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405

(Received 19 April 2004; revised 1 August 2004; accepted 4 August 2004)

In the guinea pig it has been shown that the nonlinear derived transient evoked otoacoustic emission ($TEOAE_{nl}$) is comprised of significant amounts of intermodulation distortion energy. It is expected that intermodulation distortion arising from a nonlinear distortion mechanism will contribute to the overall TEOAE in a stimulus-level-dependent manner, being greatest when basilar-membrane vibration in response to a click stimulus is greatest; with decay of vibration of the basilar membrane subsequent to stimulation by a click, nonlinear interaction along the cochlear partition should reduce and so provide for a linear mechanism to dominate $TEOAE_{nl}$ generation, i.e., the contributions of each of these mechanisms should be delay dependent. To examine this delay dependence, $TEOAE_{nl}$ evoked by acoustic clicks of varying bandwidth were time-domain windowed using a recursive exponential filter in an attempt to separate two components with amplitude and phase properties consistent with different mechanisms of OAE generation. It was found that the part of the $TEOAE_{nl}$ occurring first in time can have a relatively constant amplitude and shallow phase slope, consistent with a nonlinear distortion mechanism. The latter part of the $TEOAE_{nl}$ has an amplitude microstructure and a phase response more consistent with a place-fixed mechanism. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1798352]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM] Pages: 281–291

I. INTRODUCTION

Studies of mammalian otoacoustic emissions (OAEs) over the past two decades have revealed notable differences between human and rodent OAEs. That is, rodents have (i) larger distortion product OAEs for the same level of stimulus, consistent with broader cochlear tuning (Shera *et al.*, 2002); (ii) a much lower prevalence of spontaneous OAEs, in contrast to humans where such OAEs are commonly found (Strickland *et al.*, 1985); and (iii) transient evoked OAEs (TEOAEs) with amplitude microstructure that is qualitatively similar to that found in humans. Studies in humans suggest a one-to-one correspondence between stimulus frequencies and TEOAE frequencies, i.e., each frequency contained within the TEOAE is evoked by only that frequency within the click stimulus (Kemp, 1986; Prieve *et al.*, 1996), while in rodents (or in the guinea pig, at least) the TEOAE appears to be comprised of significant energy generated by intermodulation distortion (Yates and Withnell, 1999). These differences could be explained by describing the rodent cochlea as having broader cochlear filters and a less irregular cochlea and/or better impedance match between the middle and inner ears than that of humans.

TEOAEs in humans are thought to arise from an independent channel-generating mechanism such as a place-fixed, linear reflection mechanism (Zweig and Shera, 1995). But, OAEs in general appear to arise from a composite of two mechanisms, a nonlinear distortion mechanism and a place-fixed mechanism (Talmadge *et al.*, 1998; Shera and Guinan, 1999; Goodman *et al.*, 2003). The contribution of each of these mechanisms is thought to be stimulus-level dependent and so with increasing stimulus level, intermodulation distortion energy arising from a nonlinear distortion mechanism should contribute more significantly to the TEOAE. In the guinea pig, it has been shown that the nonlinear derived TEOAE is comprised of significant amounts of intermodulation distortion energy, but the relative contribution of the two mechanisms has not been explicitly examined. This study examines the origin of the nonlinear derived TEOAE in guinea pig based on the hypothesis that the relative contributions of each of these mechanisms should be stimulus-level and delay dependent, some of the earliest part of the response of the basilar membrane to a high-level click stimulus involving the greatest excursion of the membrane from its resting position (Recio and Rhode, 2000) and so involving the largest nonlinear interaction. With decay of vibration of the basilar membrane, nonlinear interaction along the cochlear partition should reduce and so provide for a linear mechanism to dominate $TEOAE_{nl}$ generation.

A. Origin of the TEOAE

OAEs are thought to have a complex generation involving two discrete mechanisms: place-fixed and wave-fixed (Kemp, 1986).

(i) Place-fixed: Zweig and Shera (1995) provided a theoretical framework for the generation of the place-fixed mechanism, suggesting that place-fixed OAEs are an analog of Bragg scattering: the tall, broad peak of the traveling wave emphasizes a localized region from which energy is reflected from a distributed, place-fixed irregularity, i.e., the emission is thought to arise from the tip region of the traveling wave enve-

[a]Electronic mail: rwithnel@indiana.edu

lope. The defining signature of such emissions is a phase accumulation that will increase with frequency.

(ii) Wave-fixed (nonlinear distortion): Basilar-membrane impedance is not a linear function of stimulus level and the nonlinear change in impedance acts as a perturbation that may reflect the incident energy to become an otoacoustic emission (Talmadge *et al.*, 2000). This impedance perturbation is associated with the traveling wave envelope. The OAE arising from a nonlinear wave-related interaction has a phase accumulation that is constant, irrespective of stimulus frequency (Shera and Guinan, 1999).

The contribution of each of the above mechanisms appears to vary with stimulus level (Goodman *et al.*, 2003), stimulus type, and species. For the TEOAE, examination of the origin is further complicated by the fact that the TEOAE can be extracted in more than one way, the differing methods not necessarily being equivalent in terms of the TEOAE obtained. Ear-canal sound pressure (P_{ec}) recorded in response to an acoustic stimulus is comprised of the stimulus, the OAE, and noise. Noise is reduced by synchronous averaging of the ear-canal sound-pressure recording. Extraction of the OAE from P_{ec} is dependent on the frequency and onset of the OAE relative to the acoustic stimulus. For OAEs with a frequency that differs from the stimulus frequency, the OAE can be distinguished from the stimulus using Fourier analysis or narrow-band windowing. For OAEs with a frequency that is the same as the stimulus, i.e., stimulus frequency OAEs, or in response to a short-duration stimulus with a spectrum that significantly overlaps that of the stimulus, i.e., TEOAEs, extraction of the OAE from P_{ec} is more problematic. The TEOAE can be extracted from P_{ec} in one of three ways, of which two are considered here:

(i) Time-domain windowing,
(ii) Nonlinear derived extraction.

and the third, noise suppression, is discussed in Sec. IV.

1. Time-domain windowing

TEOAEs result from using a very short-duration stimulus to evoke an OAE[1]; ideally, there would be no overlap in time between the stimulus and the OAE. However, ringing of the loudspeaker or an acoustic stimulus that persists longer than the electrical stimulus that evoked it is typically due to resonance of the loudspeaker. It is, as a result, typical that the acoustic transient stimulus waveform will overlap in time with the TEOAE. As a result, it is not possible to isolate the *total* TEOAE from the stimulus using time-domain windowing. The degree of overlap between the stimulus and the TEOAE in time dictates how much of the total TEOAE can be extracted. Figure 1 provides an example of P_{ec} recorded from one animal. A time-domain window $[F(t)]$ is also shown. Extraction of TEOAE_{win}, that part of P_{ec} that does not include the stimulus, using this example would be

$$\mathrm{TEOAE}_{win} = P_{ec} - F(t) \cdot P_{ec}$$

If there were no overlap in time between stimulus and TEOAE, TEOAE_{win} would include all of the TEOAE. If there were overlap between the stimulus and the TEOAE, TEOAE_{win} would only contain that part of the TEOAE that did not overlap with the stimulus and was not removed by time-domain windowing.

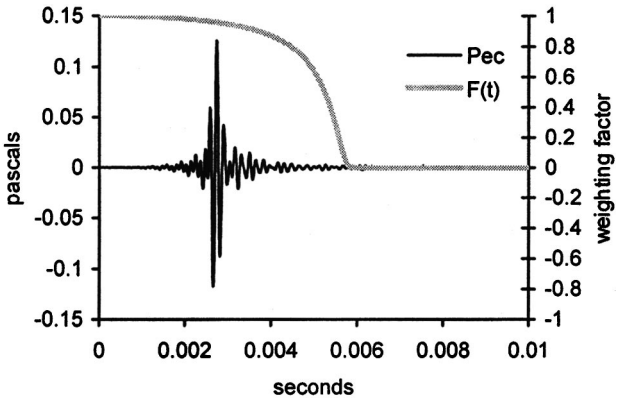


FIG. 1. Time-domain windowing: The ear-canal sound pressure (P_{ec}) when multiplied by a time-domain window $F(t)$ can be used to try and isolate the stimulus from the OAE, i.e., $\mathrm{TEOAE}_{win} = P_{ec} - F(t) \cdot P_{ec}$.

In humans, cochlear delay times are sufficiently long that some considerable part of the TEOAE is present beyond the duration of the stimulus (Kemp, 1978). Time-domain windowing of P_{ec} thus provides a means to extract a significant part of the TEOAE, although how much of the total TEOAE is extracted is not known, with stimulus ringing obscuring the onset and earliest part of the TEOAE. Rodents have much shorter cochlear delay times (Shera *et al.*, 2002), and so the TEOAE appears to be considerably obscured by the stimulus, the result being that *time-domain windowing cannot isolate some part of the TEOAE from the stimulus*. Figure 2 provides an example of the P_{ec} recorded in a live animal and postmortem [panels (b) and (c)] and the TEOAE_{nl} obtained from the live animal [panel (a)]. The *y* axis in panels (b) and (c) has been truncated to emphasize the amplitude of P_{ec} relative to TEOAE_{nl}, i.e., much of the early part of the stimulus has an amplitude that exceeds the upper bounds of the figures. It is evident that TEOAE_{nl} in panel (a) enters the noise floor before the stimulus ceases in panel (b). Panel (c), where the stimulus level in the ear canal postmortem, was matched to the stimulus used with the live animal (stimulus levels differ by 0.5 dB), provides an estimate of cessation of the stimulus: comparison of panel (b) with panel (c) suggests that there is no significant TEOAE in panel (b) beyond cessation of the stimulus [determined by panel (c)]; i.e., the sound-pressure waveform in the ear canal beyond 0.004 s is noise only.[2] Because the TEOAE has an onset time that is delayed relative to the stimulus and a duration that is at least as long as the stimulus, it follows that some part of the TEOAE must persist beyond cessation of the stimulus. However, in rodents, this component has not been observed—presumably because it is obscured by noise and so, with sufficient averaging, it might be possible to identify this delayed, small part of the TEOAE. Clearly, though, it would not constitute a significant part of the overall TEOAE.

2. Nonlinear derived extraction

Synchronously averaged human-ear-canal sound-pressure recordings in the time domain reveal that the part of

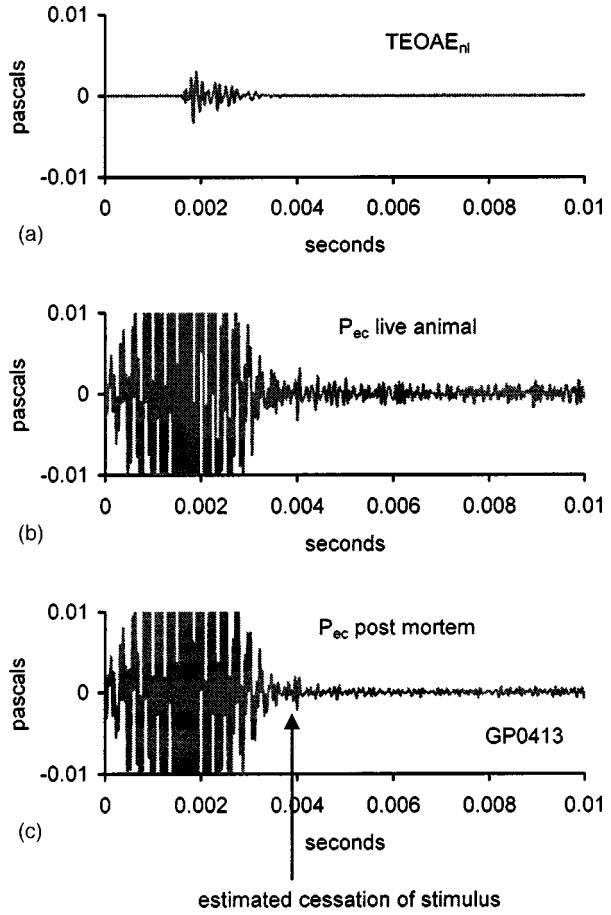


FIG. 2. An example of the ear-canal sound pressure (P_{ec}) recorded in a live animal [panel (b)] and postmortem [panel (c)] and the $TEOAE_{nl}$ obtained from the live animal [panel (a)]. The *y* axis in panels (b) and (c) has been truncated to emphasize the amplitude of P_{ec} relative to $TEOAE_{nl}$, i.e., much of the early part of the stimulus has an amplitude that exceeds the upper bounds of the figures. It is evident that $TEOAE_{nl}$ in panel (a) enters the noise floor before the stimulus ceases in panel (b). Panel (c), where the stimulus level in the ear canal postmortem was matched to the stimulus used with the live animal (stimulus levels differ by 0.5 dB), provides an estimate of cessation of the stimulus: comparison of panel (b) with panel (c) suggests that there is no significant TEOAE in panel (b) beyond cessation of the stimulus.

the TEOAE that persists beyond the stimulus grows compressively nonlinearly (Kemp, 1978), and as a result Kemp suggested that the TEOAE could be extracted from the ear-canal sound-pressure recording by making use of this nonlinear growth (see Kemp *et al.*, 1990), i.e., in response to a stimulus train consisting of *n* stimuli with a peak pressure *P* and one stimulus with a peak pressure $n.P$, the nonlinear derived TEOAE ($TEOAE_{nl}$) is given by

$$\mathrm{TEOAE_{nl}}=\frac{n\cdot P_{\mathrm{ecL}}-P_{\mathrm{ecnL}}}{n-1},$$

where P_{ecnL} is the ear-canal sound pressure recorded in response to the transient acoustic stimulus with a peak pressure $n.P$, P_{ecL} is the ear-canal sound pressure recorded in response to the transient acoustic stimulus with a peak pressure *P*. This is illustrated in Fig. 3 for $n=2$. Extraction of the total TEOAE is dependent on the stimulus level being sufficiently high that emission growth has saturated, otherwise the emission is underestimated.

Because the TEOAE in the guinea pig appears to be considerably obscured by the stimulus, i.e., most of the TEOAE is present within the duration of the stimulus, only $TEOAE_{nl}$ can be reported as representative of the TEOAE.

3. $TEOAE_{nl}$ *versus* $TEOAE_{win}$

In humans, studies of TEOAE origin have been reported for both $TEOAE_{nl}$ and $TEOAE_{win}$. Prieve *et al.* (1996) reported findings for $TEOAE_{nl}$[3] that are consistent with a one-to-one correspondence between the stimulus frequency and the frequency of the OAE, and so would suggest $TEOAE_{nl}$ arises predominantly from a place-fixed mechanism. A one-to-one correspondence between the stimulus frequency and the frequency of the OAE would not occur for an OAE arising from a wave-fixed, nonlinear distortion mechanism due to the generation of intermodulation distortion products. A $TEOAE_{nl}$ that arises predominantly from a linear, place-fixed reflection mechanism is not at odds with the method of extraction of the OAE, i.e., a method of extraction of an OAE that utilizes cochlear nonlinearity will extract on OAE that arises from a linear mechanism if the growth of the OAE is affected by the compressively nonlinear growth of basilar-membrane vibration (Shera and Guinan, 1999; Kalluri and Shera, 2004).

Avan *et al.* (1997) reported findings for $TEOAE_{win}$, where the first 2.5 ms following stimulus onset of P_{ec} was

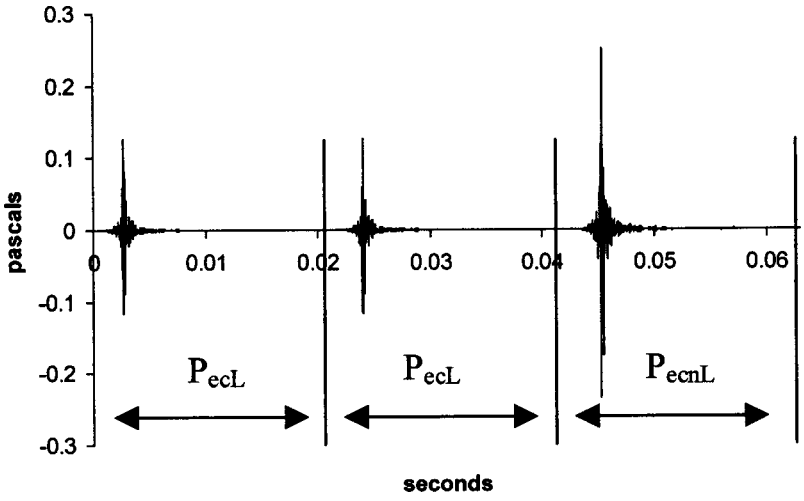


FIG. 3. An ear-canal sound-pressure measurement versus time in response to a stimulus train of three click stimuli, the first two clicks being of equal amplitude and the third twice the amplitude of the first two. The TEOAE is extracted from the ear-canal sound-pressure recordings by making use of the nonlinear growth of the TEOAE with stimulus level. In this example, the $\mathrm{TEOAE}=(2.P_{ec}-P_{ecnL})$, where the stimulus for the recording of P_{ecnL} is twice as large as the stimulus for the recording of P_{ecL}.

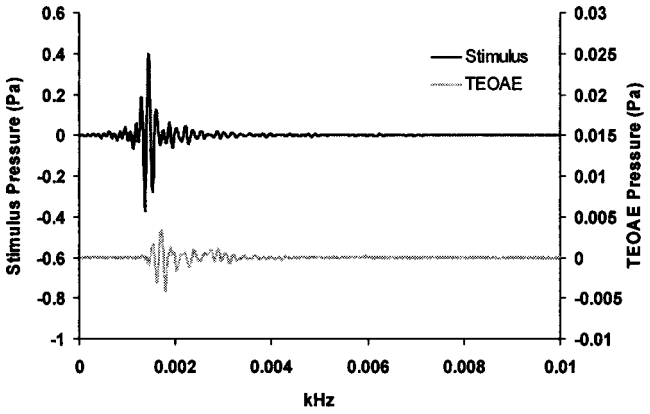


FIG. 4. An example of a P_{ec} (click stimulus, 3–9-kHz bandwidth) versus time measured in the ear canal of a guinea pig, and the corresponding $TEOAE_{nl}$.

removed by windowing. Damage to the basal turn of the cochlea was found to influence $TEOAE_{win}$ energy at frequencies corresponding tonotopically to more apical cochlear locations. Such a finding is consistent with intermodulation distortion energy contributing to this TEOAE (Yates and Withnell, 1999; Withnell *et al.*, 2000). Of course, it is also possible that damage to the basal turn introduces an additional source of reflection from which OAE energy could arise (Avan *et al.*, 1997; Kakigi *et al.*, 1998; Shera *et al.*, 2004). Withnell *et al.* (2000) reported small changes to $TEOAE_{nl}$ amplitude spectra in guinea pig with 1–5-kHz click stimuli following TTS of the basal turn, consistent with an additional reflection source, but also found that there was no change for an electrically evoked OAE (EEOAE) when the basal turn was damaged and the electrical current was injected into the third turn. No change to the EEOAE may not refute an additional reflection source associated with damage to the basal turn—modeling studies suggest such cochlear ''perturbations'' affect forward and reverse traveling waves differently with greater reflection of forward traveling waves (Shera, personal communication 2004).

Each method of extraction of the TEOAE has its limitations and neither is necessarily representative of the total TEOAE. $TEOAE_{win}$ is contaminated by stimulus artifact and/or some of the TEOAE having been removed by windowing. $TEOAE_{nl}$ will underestimate the TEOAE if the growth of the TEOAE is compressively nonlinear and the stimulus level is not sufficiently high that emission growth has saturated. Commensurate with this, Ravazzani *et al.* (1996) found $TEOAE_{nl}$ in humans to be similar to $TEOAE_{win}$ greater than 6 ms postsignal onset in response to ''high-level'' stimuli.

In the guinea pig, studies of TEOAE origin have been confined to $TEOAE_{nl}$, with recent findings suggesting that $TEOAE_{nl}$ appears to be comprised of significant energy generated by intermodulation distortion (Yates and Withnell, 1999; Withnell *et al.*, 2000). This is in stark contrast to the findings for $TEOAE_{nl}$ in humans (Prieve *et al.*, 1996), and yet $TEOAE_{nl}$ amplitude microstructure in guinea pig is qualitatively similar to that found in humans, i.e., both exhibit a quasiregular cyclical variation in the amplitude spectrum of the emission.

B. Origin of $TEOAE_{nl}$ in guinea pig

Intermodulation distortion has been shown to contribute to the $TEOAE_{nl}$ in guinea pig (Yates and Withnell, 1999). Energy arising from a nonlinear distortion mechanism, it has been suggested, should have an amplitude spectrum that is, for the most part, essentially devoid of microstructure (Talmadge *et al.*, 2000). Further, energy arising from a nonlinear distortion mechanism should be present in the earliest part of the $TEOAE_{nl}$ response when the basilar-membrane response to the acoustic transient stimulus is greatest; as the response of the basilar membrane to the acoustic transient decays, nonlinear interaction will reduce and so provide for OAE arising from a place-fixed mechanism to contribute more significantly.

Amplitude microstructure for an OAE that arises solely from a linear place-fixed reflection mechanism would arise from the complex interaction of intracochlear standing waves with the initial apically reflected wave/s and variations in cochlear reflectance. This resonant behavior (intracochlear standing waves) will produce $TEOAE_{nl}$'s with a slow decay time (Talmadge *et al.*, 1998) and spontaneous OAEs (Talmadge *et al.*, 1998; Shera, 2003). While such a $TEOAE_{nl}$ will arise without nonlinear interaction, and so each frequency in the $TEOAE_{nl}$ will correspond to its tonotopic cochlear location, intracochlear reflections will disassociate a simple temporal relationship for $TEOAE_{nl}$ frequencies. In the absence of intracochlear reflections, the $TEOAE_{nl}$ would be expected to have a fast decay time and a temporal relationship whereby the higher the OAE frequency, the shorter the delay time.

The amplitude microstructure in the $TEOAE_{nl}$ in guinea pig, if it does not arise from intermodulation distortion energy, must arise from the interaction of OAE arising from a nonlinear distortion mechanism and OAE arising from a place-fixed mechanism. As such, it should be temporally distinguishable as being dominant in the latter part of the $TEOAE_{nl}$. To investigate this temporal relationship and amplitude microstructure in $TEOAE_{nl}$ in guinea pig, the $TEOAE_{nl}$ was time-domain windowed using a recursive exponential filter in an attempt to determine if two components with amplitude and phase properties consistent with different mechanisms of OAE generation could be distinguished.

II. METHOD

A. Animal surgery

Albino guinea pigs (300 to 550 grams) were anesthetized with Nembutal (35 mg/kg i.p.) and Atropine (0.06–0.09 mg i.p.), followed approximately 15 minutes later by Hypnorm (0.1–0.15 ml i.m.). Neuroleptanaethesia was maintained with supplemental doses of Nembutal and Hypnorm. Guinea pigs were tracheostomized and mechanically ventilated on Carbogen (5% CO2 in O2) with body rectal temperature maintained at approximately 38.5 deg Celsius. The head was positioned using a custom-made head holder that could be rotated for access to the ear canal. Heart rate was monitored throughout each experiment. The bulla was opened dorso-laterally and a silver wire electrode placed on

the round-window niche for the recording and monitoring of the compound action potential (CAP). In some cases, Pancuronium (0.15 ml i.m.) was administered to reduce physiological noise associated with muscle contractions. The number of animals examined for this study was 18, of which 10 had CAP thresholds that remained stable throughout the experiment and were within laboratory norms. The data presented in this paper represent a subset of the data obtained from the 10 animals with good CAP thresholds.

B. Signal generation

The method for stimulus delivery has been described previously (Withnell *et al.*, 1998; Withnell and Yates, 1998). Briefly, the acoustic stimuli were delivered by a Beyer DT48 loudspeaker placed approximately 4 cm from the entrance to the ear canal. The stimulus waveform was calculated as a sinc function ($\sin(x)/x$), appropriately time scaled according to the required low-pass frequency and windowed over 3 ms, using the equation

$$A(t)=\begin{bmatrix}\cos^2(\pi t/0.003)[\sin(2\pi f_c t)]/(2\pi f_c t) & -0.0015<t<0.0015\\ 0 & \text{otherwise}\end{bmatrix},$$

where f_c is the low-pass corner frequency.

Acoustic transient stimuli with varying bandwidths (e.g., 3–9, 7–14, 10–18 kHz) and stimulus levels were delivered to the ear without amplification, buffered by a Tucker-Davis HB6 amplifier. Stimulus spectra were relatively flat, achieved by compensating for the loudspeaker frequency response in the ear canal (see Yates and Withnell, 1999, for more details).

C. Data acquisition

Ear-canal sound-pressure recordings (P_{ec}) were made by a Sennheiser MKE 2–5 electrostatic microphone fitted with a metal probe tube (1.2 mm long, 1.3 mm i.d., 1500Ω acoustic resistor) positioned approximately 2 mm into the ear canal. The microphone and probe tube combination was calibrated against a Bruel & Kjaer 4138 1/8-in. microphone. The output from the probe tube microphone was amplified 20 dB, high-pass filtered (0.64 kHz, 4-pole Butterworth), and transmitted as a balanced input to one of the analog input channels of the computer sound card (total gain = 30 dB). It was subsequently digitized at a rate of 96 kHz.

This study examines TEOAE_{nl} obtained from the guinea pig, obtained using the nonlinear derived response technique (Kemp *et al.*, 1990) with a 6-dB stimulus level ratio, i.e., the stimulus train consisted of three acoustic transients with a 21.4-ms interstimulus interval, one of the stimuli being 6 dB higher than the other two. Each synchronously averaged recording represented a total of 21.38 s of data collection (21.4-ms epochs, 2048 points for the FFT). Data analysis was performed using Microsoft EXCEL and MATLAB.

D. Time-domain windowing of TEOAE_{nl}

Time-domain windowing to separate components with disparate amplitude structure was performed using a recursive exponential filter developed by Shera and Zweig (see Kalluri and Shera, 2001, Shera and Zweig, 1993), i.e.,

$$\text{TEOAE}_{\text{nl}}\ \text{short latency component}=T(t)\cdot F(t),$$

where $T(t)$ is TEOAE_{nl} and $F(t)=1/\Gamma_n(\tau)$

$\tau=t/\tau_{\text{cut}}$, where t is time, τ_{cut} is the length of the window

$\Gamma_n(\tau)$ is defined recursively as

$$\Gamma_{n+1}(\tau)=e^{\Gamma n(\tau)-1},\ \text{with}\ \Gamma_1(\tau)=e^{\tau^2}.$$

The value of τ_{cut} was chosen with the goal of minimizing the amplitude microstructure of the short latency component.

Filter order (n) was 14. This filter order was chosen based on the requirement of having a filter sufficiently steep that it meaningfully separates the TEOAE_{nl} in time while not being so steep that it introduces artifacts into the response. The late component of TEOAE_{nl} is given by

$$\text{TEOAE}_{\text{nl}}\ \text{late component}=T(t)-[T(t)\cdot F(t)].$$

III. RESULTS

A. Amplitude microstructure

Figure 4 shows a click stimulus (3–9-kHz bandwidth) and the corresponding TEOAE_{nl} versus time obtained from a guinea pig. Figure 5, panels (a) to (f), illustrates the effect of time-domain windowing the TEOAE_{nl} of Fig. 4 with six different values of τ_{cut} (ranging from 0.001 05 to 0.003 15 seconds[4]), each panel having both amplitude and phase of the early OAE component, late OAE component, and the amplitude of the unwindowed TEOAE [except for panel (a), where no part of the TEOAE is within the window and so only the late component is present, this component being equivalent to the unwindowed TEOAE]. It is expected that applying a time-domain window to any signal in the time domain will separate the signal into two components—an early component and a late component. This is evident in panels (b) to (f). It is notable in panels (b) and (c) that the early OAE component has a relatively flat amplitude spectrum, with none of the amplitude microstructure that is evident in the unwindowed TEOAE or the late OAE component, and a phase slope that is not as steep as that of the late OAE component [panels (b) and (c) show the phase of TEOAE_{nl}, the phase of the early OAE component in both cases being similar to the phase of TEOAE_{nl}]. In subsequent figures, the choice of value of τ_{cut} used was that which

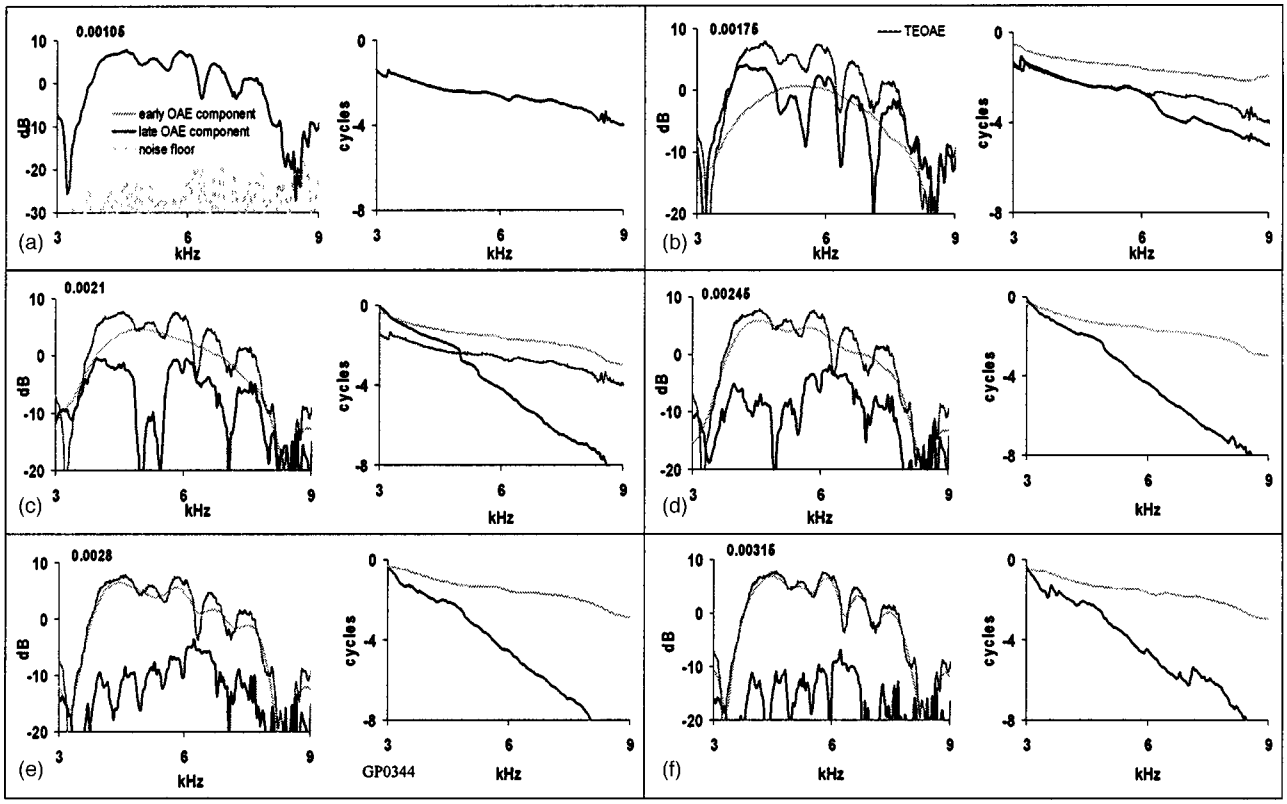


FIG. 5. Amplitude and phase spectra illustrating time-domain windowing of the $TEOAE_{nl}$ in Fig. 4 with six different cutoff values (τ_{cut}) for separating the $TEOAE_{nl}$ into two components, designated the "early component" and "late component." In panel (a), the $TEOAE_{nl}$ is not windowed at all and so the windowed component is negligible, while the remainder (the late component) is equal to the $TEOAE_{nl}$. Subsequent panels (b) to (f) show various degrees of separation of the $TEOAE_{nl}$ into two components. Panel (c) shows the $TEOAE_{nl}$ to be separated into two components, distinguishable based on amplitude and phase. See the text for further details.

(i) provided an early OAE component essentially devoid of amplitude microstructure and a late OAE component with amplitude microstructure;
(ii) had a phase for the dominant or largest component with a similar slope to the phase of $TEOAE_{nl}$; and
(iii) produced two components with distinguishable phase slopes, i.e., the slope of the early OAE component was less than that of the late OAE component.

Based on these criteria, a τ_{cut} of 0.0021 s, panel (c), would be chosen. Panel (b) with a τ_{cut} of 0.001 75 s provides for a late component with larger amplitude than the early component for a significant part of the frequency range of the OAE, but the phase of $TEOAE_{nl}$ [right side of panel (b)] has a slope that is inconsistent with this—$TEOAE_{nl}$ phase suggests a single component dominating above 4.5 kHz with the slope consistent with the dominant component being the early OAE component. To be consistent with the $TEOAE_{nl}$ phase, the amplitude of the early component should be greater than the late component above 4.5 kHz.

Pursuant to the observation that the $TEOAE_{nl}$ can be temporally windowed into two components with the amplitude microstructure confined to the late component, Fig. 6 shows examples from three animals of amplitude spectra and phase for $TEOAE_{nl}$ windowed[5] with the amplitude microstructure largely confined to the late component. In each case, the stimulus had a bandwidth of 4–10 kHz. The early OAE component spectral amplitude is predominantly greater than the late OAE component in panels (a) and (b), panel (c) revealing a different mix of the two components (the late OAE component is larger than the early OAE component below 5.8 kHz and from 9 to 9.8 kHz). The right side of panels (a) and (c) show $TEOAE_{nl}$ phase to have a similar slope to the late OAE component at lower frequencies, becoming similar to the early OAE component phase slope above 5 to 6 kHz. Panel (b) shows $TEOAE_{nl}$ phase slope to be similar to the early component phase slope. A relatively slowly rotating phase for the $TEOAE_{nl}$ emission is not commensurate with a round-trip physical cochlear delay to the 4–10-kHz region of the guinea pig cochlea. For a characteristic frequency of 7 kHz, the slope of the $TEOAE_{nl}$ phase gives delays of approximately 350 μs, too short for either a one-way or a round-trip cochlear delay, i.e., a round-trip cochlear delay of about 1600 μs would be expected.[6] The slope of the phase in panels (a) and (b) of $TEOAE_{nl}$ is consistent with the emission arising predominantly from a wave-fixed mechanism. The steeper slope for $TEOAE_{nl}$ in panel (a) below 5.2 kHz and in panel (c) below 5.8 kHz and from 9 to 9.8 kHz is commensurate with this part of $TEOAE_{nl}$ arising from a place-fixed mechanism.

Figure 7 shows two examples of $TEOAE_{nl}$ evoked by 10–18-kHz acoustic transients, i.e., a more basal cochlear stimulation than for Fig. 6. Amplitude microstructure is less pronounced in panel (a) than is found in panel (b) or in Fig. 6. Indeed, in panel (a), while it was possible to window the

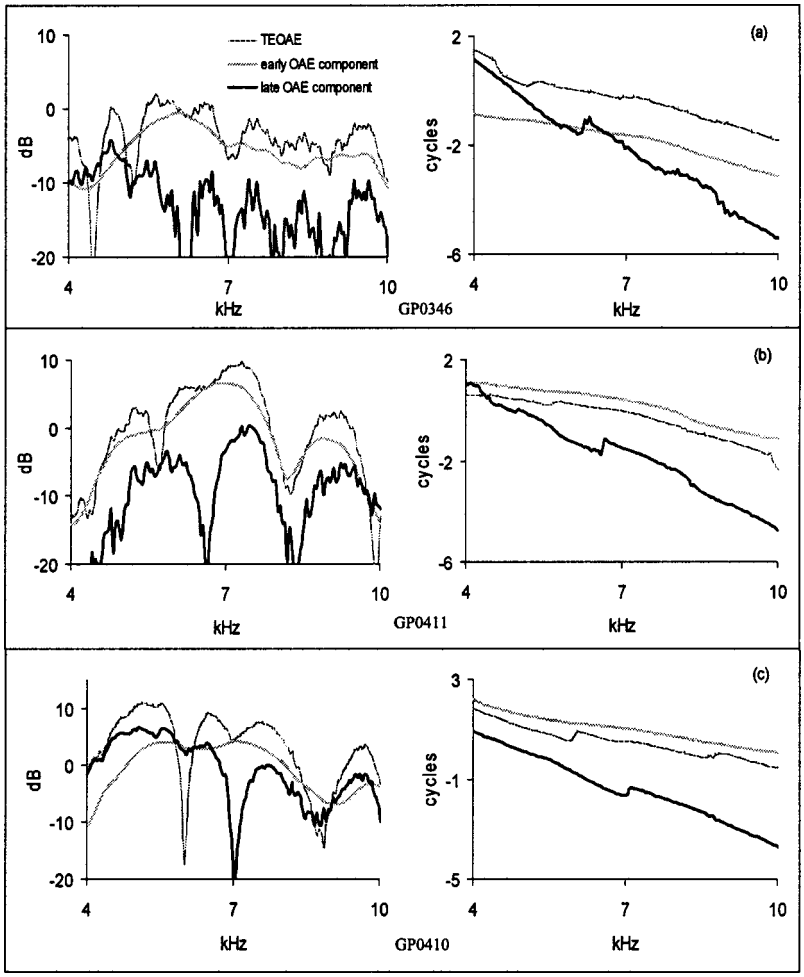


FIG. 6. Examples from three animals of amplitude and phase spectra for $TEOAE_{nl}$ windowed with the amplitude microstructure largely confined to the late component. In each case, the stimulus had a bandwidth of 4–10 kHz. The early OAE component spectral amplitude is predominantly greater than the late OAE component in panels (a) and (b), panel (c) revealing a different mix of the two components (the late OAE component is larger than the early OAE component below 5.8 kHz and from 9 to 9.8 kHz). Amplitude spectra are relatively devoid of microstructure for the early component, the late component having, in each case, pronounced microstructure. Phase responses show an early component with a shallow phase slope and a late component with a much steeper phase.

$TEOAE_{nl}$ into two components with amplitude microstructure confined to the late component, the phases of the two components shown on the right side of panel (a) reveal similar slopes and the amplitude of the late OAE component on the left side of panel (a) is more than 10 dB less than the early component versus frequency. That is, the phases of the slopes of the two components do not differ in panel (a)—it would seem in this case, commensurate with the small am-

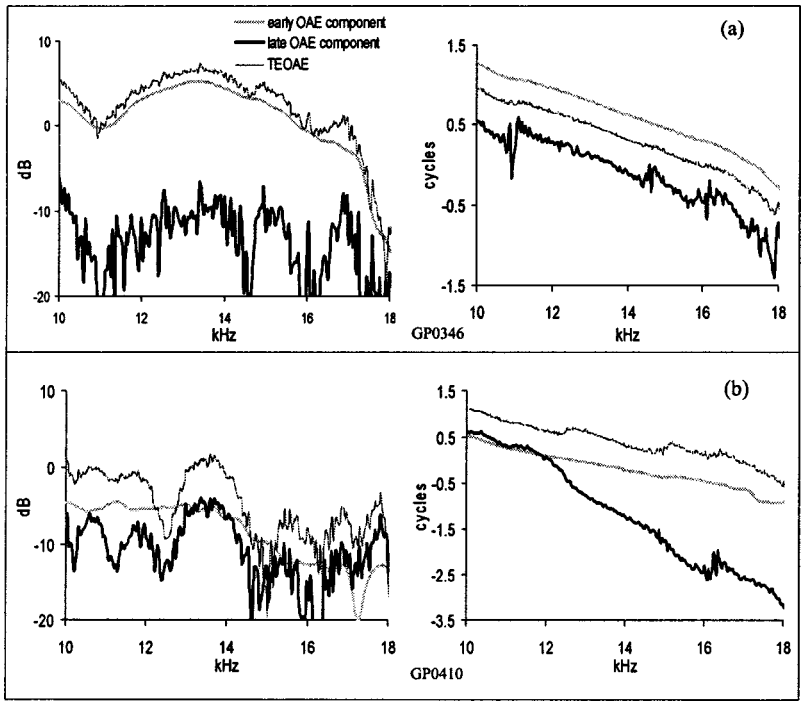


FIG. 7. Two examples of amplitude and phase spectra for $TEOAE_{nl}$ evoked by 10–18-kHz acoustic clicks, i.e., a more basal cochlear stimulation than for Fig. 6. $TEOAE_{nl}$ amplitude microstructure is less pronounced in panel (a) than in panel (b). Indeed, in panel (a), while it was possible to window the $TEOAE_{nl}$ into two components with amplitude microstructure confined to the late component, the phases of the two components shown on the right side of panel (a) reveal similar slopes and the amplitude of the late OAE component on the left side of panel (a) is more than 10 dB less than the early component versus frequency. That is, the phases of the slopes of the two components do not differ in panel (a)—it would seem in this case, commensurate with the small amplitude microstructure, that this $TEOAE_{nl}$ does not have two distinguishable OAE components. The $TEOAE_{nl}$ in panel (b), in contrast, has greater amplitude microstructure and time-domain windowing isolates two components distinguishable based on amplitude and phase.

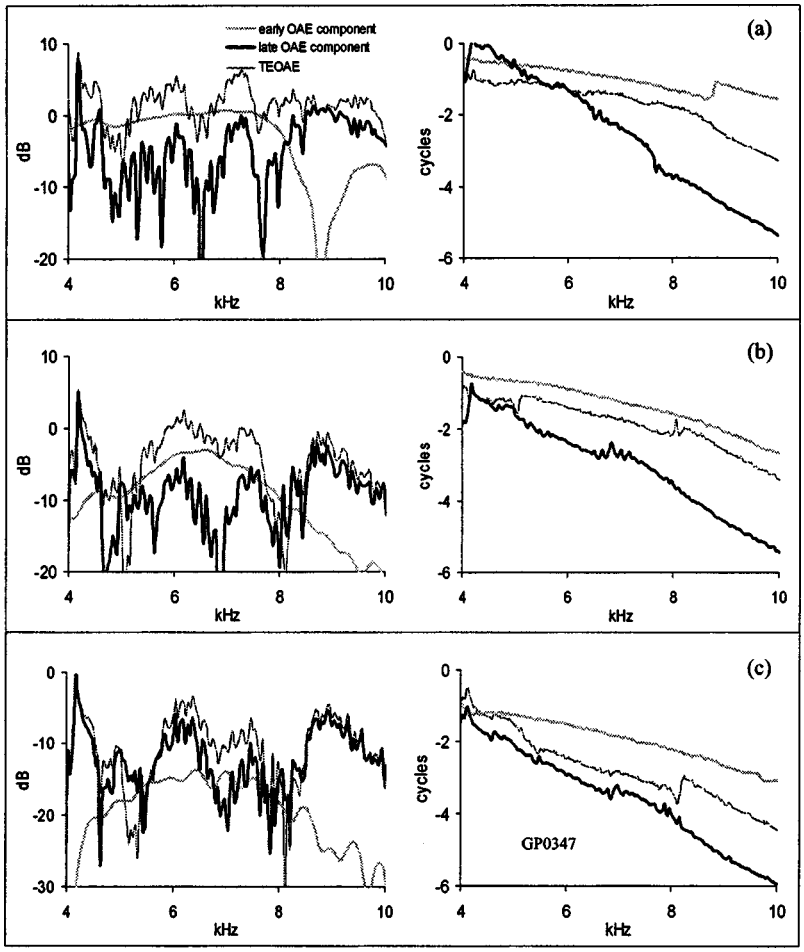


FIG. 8. Amplitude and phase spectra illustrating the stimulus-level dependence of the relative contribution of nonlinear distortion and place-fixed mechanisms to the generation of $TEOAE_{nl}$. Panel (a) is the OAE evoked by the highest level stimulus, panel (c), the lowest level stimulus, stimulus level decreasing in 5-dB steps from (a) to (c). In panel (a), $TEOAE_{nl}$ is dominated by the early component. With decreasing stimulus level, the relative contribution of the early component reduces.

plitude microstructure, that this $TEOAE_{nl}$ does not have two distinguishable OAE components. The $TEOAE_{nl}$ in panel (b), in contrast, has greater amplitude microstructure, and time-domain windowing does isolate two components distinguishable based on amplitude and phase.

B. $TEOAE_{nl}$ versus stimulus level

If the two components of the $TEOAE_{nl}$ in guinea pig represent energy arising from a nonlinear distortion mechanism (the early component) and energy arising from a place-fixed reflection mechanism (the late component), then the relative contribution of these two components should be stimulus-level dependent (Shera and Guinan, 1999; Talmadge *et al.*, 2000; Goodman *et al.*, 2003), i.e., the component arising from a nonlinear distortion or wave-fixed mechanism should become more significant relative to the place-fixed component with increasing stimulus level. This is examined in Fig. 8. Stimulus level decreases from panel (a) to panel (c) in 5-dB steps. The $TEOAE_{nl}$ in panel (a) is dominated by the early OAE component below 8 kHz and the late OAE component above 8 kHz—this is evident from the $TEOAE_{nl}$ phase on the right side of panel (a), there being a shallow phase slope below 8 kHz and a steeper slope above 8 kHz. Panel (c) shows a $TEOAE_{nl}$ with a steeper phase response up to 8 kHz than $TEOAE_{nl}$ in panel (a), consistent with it being predominantly made up of the late OAE component (above 8 kHz the late component dominated the $TEOAE_{nl}$ at all stimulus levels), i.e., this is consistent with the $TEOAE_{nl}$ in panel (c) being dominated by a place-fixed generating mechanism. Windowing of $TEOAE_{nl}$ to isolate an early OAE component with an amplitude spectrum relatively devoid of microstructure [left side of panel (c)] is consistent with this.

Figure 8 presents data consistent with the notion that the early and late OAE components represent OAE generated by a nonlinear distortion mechanism and a place-fixed mechanism, respectively, these two mechanisms being stimulus-level dependent. OAE generated by a nonlinear distortion or wave-fixed mechanism presumably becomes increasingly significant as stimulus level increases.

C. Physiological verification of $TEOAE_{nl}$

The TEOAE was extracted using the nonlinear derived extraction paradigm. The setup for stimulus delivery and response acquisition was an open-field system (see Withnell *et al.*, 1998). An open-field system results in very little nonlinear stimulus-related artifact being present in $TEOAE_{nl}$ and so eliminates the need for time-domain windowing of the early part of the response as is done for human TEOAE recordings where the TEOAE is acquired using a closed system. Figure 9 illustrates the stimulus contamination of $TEOAE_{nl}$ by contrasting the nonlinear derived component of the ear-canal sound-pressure recording extracted pre- and postmortem (stimulus level being matched pre- and postmor-

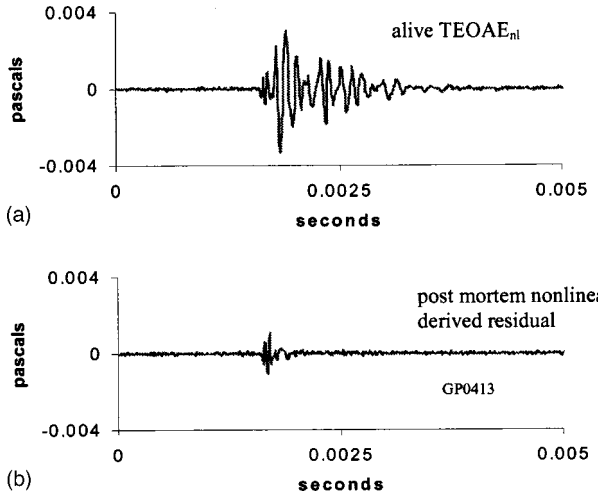


FIG. 9. Verification of the physiological nature of the TEOAE$_{nl}$ response. Panel (a) shows the TEOAE$_{nl}$ obtained from an alive animal, panel (b) the nonlinear derived residual postmortem.

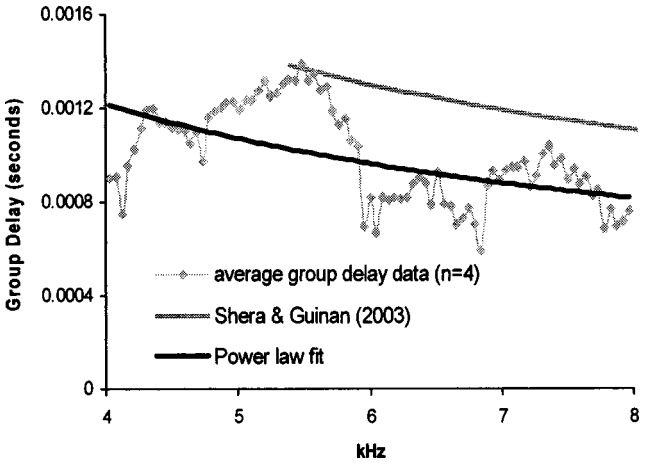


FIG. 10. Average group delay data (phase derivative) for the four animals from Figs. 5(c) and 6, a trend line fitted to this data, and the equation of best fit reported by Shera and Guinan (2003) to their data for SFOAEs evoked by 40-dB SPL stimuli.

tem), and the physiological nature of the TEOAE$_{nl}$ obtained in this study. It is evident that there is very little contamination of TEOAE$_{nl}$ by the stimulus.

IV. DISCUSSION

It would appear that the TEOAE$_{nl}$ in the guinea pig can be time-domain filtered to separate two components with different amplitude and phase properties consistent with different mechanisms of OAE generation. It has previously been shown that intermodulation distortion energy contributes significantly to this TEOAE$_{nl}$ (Yates and Withnell, 1999; Withnell *et al.*, 2000). It is suggested that the earliest part of the TEOAE$_{nl}$ separable by time-domain filtering that is essentially devoid of amplitude microstructure and has a phase slope that is shallow (associated with wave scaling[7]) arises from a nonlinear distortion mechanism. A click stimulus causes the basilar membrane to vibrate analogous to stimulating the cochlea with a range of stimulus levels—the displacement of the basilar membrane is largest soon after the onset of the cochlear response to a high-level click, providing for greatest nonlinear interaction, and then decays over time. The latter part of the TEOAE$_{nl}$ with amplitude microstructure and a phase response that is steeper than the early component presumably arises from a place-fixed mechanism. The amplitude microstructure present in this latter component, in the absence of spontaneous OAEs and multiple intracochlear reflections (Withnell *et al.*, 2003), could be due to variations in cochlear reflectance (Zweig and Shera, 1995; Goodman *et al.*, 2003). The amplitude microstructure of TEOAE$_{nl}$ presumably arises from the complex interaction of the component arising from nonlinear distortion with the component arising from a place-fixed mechanism.

A stimulus-level dependence is at the heart of the perceived mechanisms by which OAEs are generated (Zweig and Shera, 1995; Talmadge *et al.*, 2000; Goodman *et al.*, 2003). Figure 8 suggests that the relative contributions of each of these mechanisms is stimulus-level dependent, the contribution of intermodulation distortion to the TEOAE$_{nl}$ becoming more significant with increasing stimulus level. At the lowest stimulus level shown in Fig. 8 [panel (c)], the TEOAE$_{nl}$ is predominantly comprised of energy arising from a place-fixed mechanism. TEOAE$_{nl}$ phase in panel (c) is much steeper than in panel (a), commensurate with a shift in the mechanism dominating production of TEOAE$_{nl}$.

Time-domain windowing of the TEOAE$_{nl}$ does not always appear to isolate two components distinguishable by their amplitude and phase. Figure 7(a) provides an example of a TEOAE$_{nl}$ with very little amplitude microstructure that appears to arise almost totally from a nonlinear distortion mechanism—in this example, windowing does not isolate two components with different phase slopes.

For wideband inputs, group delay typically has no physical meaning, it being defined for narrow-band inputs (Papoulis, 1962). However, the cochlea separates a wideband input into a series of narrow-band inputs. A TEOAE generated by a linear, place-fixed reflection mechanism may be equivalent to an OAE that is a composite of SFOAEs generated by low-level stimuli with a frequency range encapsulated by the bandwidth of the click stimulus, i.e., it is a wideband response that is the sum of a series of narrow-band responses without nonlinear interaction. With no SOAEs and no multiple internal reflections (Withnell *et al.*, 2003), group delay should then represent a round-trip delay for independent-channel generated TEOAE components. However, the phase response of an OAE arising from a place-fixed mechanism has been shown to be representative of round-trip cochlear delay subject to the effects of variation in cochlear reflectance (Shera and Guinan, 2003). Figure 10 provides the average group delay data (phase derivative) for the four animals from Figs. 5(c) and 6, a trend line fitted to this data, and the equation of best fit reported by Shera and Guinan (2003) to their data for SFOAEs evoked by 40-dB SPL stimuli. It is evident that the group-delay trend line for the latter part of the TEOAE$_{nl}$ component presumed to arise from a place-fixed origin underestimates SFOAE delay reported by Shera and Guinan (2003). The disagreement may be due to any of

(i) the small sample size of our data set;
(ii) the $TEOAE_{nl}$ component described as having a place-fixed origin may still have a component with a nonlinear origin confounding the phase;
(iii) it is not meaningful to calculate group delay for wideband inputs; and
(iv) stimulus-level differences.

Irrespective of the disagreement between the trend line for group-delay data from the latter part of the $TEOAE_{nl}$ versus the findings of Shera and Guinan (2003), the notable differences in amplitude and phase for the two components of the $TEOAE_{nl}$ isolated by time-domain windowing lead us to believe that the latter part of the $TEOAE_{nl}$ does indeed arise from a place-fixed origin.

A. Choice of τ_{cut}

It is not contended nor implied that there is a value of τ_{cut} that perfectly separates $TEOAE_{nl}$ arising from a nonlinear distortion mechanism versus a place-fixed mechanism. It is to be expected, based on a stimulus-level dependence underlying the generation of OAEs, that the $TEOAE_{nl}$ over time represents a continuum of energy arising from both wave-fixed and place-fixed mechanisms. The degree to which each mechanism contributes to the $TEOAE_{nl}$ at any point in time is presumably dependent on the displacement of the BM at that point in time in response to a click stimulus.

B. Noise suppression paradigm to extract the TEOAE from P_{ec}

In Sec. I A, time-domain windowing and nonlinear derived extraction were considered in terms of isolating the TEOAE from the stimulus. A third method of extracting the TEOAE from the ear-canal sound pressure (P_{ec}) is using a noise suppression paradigm. Molenaar *et al.* (2000) examined TEOAEs evoked by clicks with and without pseudorandom noise bursts of alternating phase in humans, the noise suppressing the TEOAE otherwise evoked by the click stimulus. This technique has the limitation that the TEOAE extracted will underestimate the total TEOAE if the noise suppressor does not completely suppress the generation of TEOAE. It does offer, though, an alternative means of extracting the TEOAE from P_{ec} that is worth investigating in the guinea pig.

ACKNOWLEDGMENTS

This work was supported in part by NIH-NIDCD DC04921 (R.H.W.). We thank Dr. Stephen Neely, Dr. Christopher Shera, Dr. Sumit Dhar, and one anonymous reviewer for providing valuable feedback on the manuscript, and Dr. Shawn Goodman for writing the MATLAB code used for data analysis.

[1]In response to a transient acoustic stimulus that has a relatively flat stimulus spectrum, the cochlea distributes the energy along the cochlear partition in a frequency-dependent manner with a delay that is reciprocally related to stimulus frequency. The TEOAE generated will reflect such cochlear delay and the mechanisms underlying generation.

[2]The larger level of noise in panel (b) relative to panel (c) presumably is due to physiological noise.

[3]$TEOAE_{nl}$ for this study was actually a combination of windowing and nonlinear extraction. The nonlinear extraction paradigm using a closed system [see Withnell *et al.* (1998) for discussion of closed versus open systems] in humans has a stimulus artifact that contaminates the early part of the ear-canal sound-pressure recording, and so the first 6 ms of the ear-canal sound-pressure recording was zeroed (see Prieve *et al.*, 1996).

[4]τ_{cut} is defined as that value in time where weighting value of window is 50%.

[5]$\tau_{cut} = 0.002\,52$ s for Fig. 6, panel (a) and 0.002 62 s for Fig. 6, panel (b).

[6]Estimate derived from auditory-nerve fiber threshold tuning data of Tsuji and Liberman (1997), based on a 400-μs delay to the 18-kHz region, and assuming a phase change across filter bandwidth that is independent of bandwidth with a reciprocal relationship between cochlear group delay and filter bandwidth.

[7]The phase versus frequency response is relatively flat due to the impedance perturbation that reflects the incident energy shifting with the traveling wave envelope, i.e., the ''traveling wave envelopes are locally 'shift-similar,' with the number of wavelengths in the traveling wave nearly independent of frequency'' (Shera and Guinan, 1999, p. 786), with the reflection coming from the same region of the traveling wave, regardless of stimulus frequency. Wave scaling is a consequence of scaling symmetry (Shera and Guinan, 1999).

Avan, P., Elbez, M., and Bonfils, P. (**1997**). ''Click-evoked otoacoustic emissions and the influence of high-frequency hearing losses in humans,'' J. Acoust. Soc. Am. **101**, 2771–2777.

Goodman, S. S., Withnell, R. H., and Shera, C. A. (**2003**). ''The origin of SFOAE microstructure in guinea pig,'' Hear. Res. **183**, 7–17.

Kakigi, A., Hirakawa, H., Harel, N., Mount, R. J., and Harrison, R. V. (**1998**). ''Basal cochlear lesions result in increased amplitude of otoacoustic emissions,'' Audiol. Neuro-Otol. **3**, 361–372.

Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: A test of the two mechanism model for DPOAE generation,'' J. Acoust. Soc. Am. **109**, 622–637.

Kalluri, R., and Shera, C. A. (**2004**). ''The relationship between TEOAEs and SFOAEs at low stimulus levels,'' Association for Research in Otolaryngology Mid-Winter Meeting, Daytona Beach, Florida.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1986**). ''Otoacoustic emissions, traveling waves and cochlear mechanisms,'' Hear. Res. **22**, 95–104.

Kemp, D. T., Ryan, S., and Bray, P. (**1990**). ''A guide to the effective use of otoacoustic emissions,'' Ear Hear. **11**, 93–105.

Molenaar, D. G., Shaw, G., and Eggermont, J. J. (**2000**). ''Noise suppression of transient-evoked otoacoustic emissions. I. A comparison with the nonlinear method,'' Hear. Res. **143**, 197–207.

Papoulis (**1962**). *The Fourier Integral and its Applications*. (McGraw-Hill Book Company, New York).

Prieve, B. A., Gorga, M. P., and Neely, S. T. (**1996**). ''Click- and tone-burst-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears,'' J. Acoust. Soc. Am. **99**, 3077–3086.

Ravazzani, P., Tognola, G., and Grandori, F. (**1996**). '''Derived nonlinear' versus 'linear' click-evoked otoacoustic emissions,'' Audiology **35**, 73–86.

Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.

Shera, C. A. (**2003**). ''Wave interference in the generation of reflection- and distortion-source OAEs,'' in *Biophysics of the Cochlea: From Molecule to Model* (World Scientific, Singapore), pp. 439–449.

Shera, C. A., and Guinan, J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.

Shera, C. A., and Guinan, J. J. (**2003**). ''Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning,'' J. Acoust. Soc. Am. **113**, 2762–2772.

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.

Shera, C. A., Tubis, A. Talmadge, C. L., and Guinan, J. J. (**2004**). ''The dual effect of 'suppressor' tones on stimulus-frequency otoacoustic emissions,'' Association for Research in Otolaryngology Mid-Winter Meeting, Daytona Beach, Florida.

Shera, C. A., and Zweig, G. (**1993**). ''Noninvasive measurement of the cochlear traveling wave ratio,'' J. Acoust. Soc. Am. **93**, 3333–3352.
Strickland, E. A., Burns, E. M., and Tubis, A. (**1985**). ''Incidence of spontaneous otoacoustic emissions in children and infants,'' J. Acoust. Soc. Am. **78**, 931–935.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). ''Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure,'' J. Acoust. Soc. Am. **108**, 2911–2932
Tsuji, J. and Liberman, M. C. (**1997**). ''Intracellular labeling of auditory nerve fibers in guinea pig: Central and peripheral projections,'' J. Comp. Neurol. **381**, 188–202.
Withnell, R. H., Kirk, D. L., and Yates, G. K. (**1998**). ''Otoacoustic emissions measured with a physically open recording system,'' J. Acoust. Soc. Am. **104**, 350–355.
Withnell, R. H., and Yates, G. K. (**1998**). ''Onset of basilar membrane nonlinearity reflected in cubic distortion tone input–output functions,'' Hear. Res. **123**, 87–96.
Withnell, R. H., Yates, G. K., and Kirk, D. L. (**2000**). ''Changes to low-frequency components of the TEOAE following acoustic trauma to the base of the cochlea,'' Hear. Res. **139**, 1–12.
Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). ''Generation of DPOAEs in the guinea pig,'' Hear. Res. **178**, 106–117.
Yates, G. K., and Withnell, R. H. (**1999**). ''The role of intermodulation distortion in transient-evoked otoacoustic emissions,'' Hear. Res. **136**, 49–64.
Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of evoked otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task

Douglas S. Brungart[a] and Brian D. Simpson
Air Force Research Laboratory, AFRL/HECB, 2610 Seventh Street, WPAFB, Ohio 45433

Christopher J. Darwin
University of Sussex, Falmer, BN1 9QH, England

Tanya L. Arbogast and Gerald Kidd, Jr.
Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 5 March 2004; revised 1 November 2004; accepted 2 November 2004)

Recent results have shown that listeners attending to the quieter of two speech signals in one ear (the target ear) are highly susceptible to interference from normal or time-reversed speech signals presented in the unattended ear. However, speech-shaped noise signals have little impact on the segregation of speech in the opposite ear. This suggests that there is a fundamental difference between the across-ear interference effects of speech and nonspeech signals. In this experiment, the intelligibility and contralateral-ear masking characteristics of three synthetic speech signals with parametrically adjustable speech-like properties were examined: (1) a modulated noise-band (MNB) speech signal composed of fixed-frequency bands of envelope-modulated noise; (2) a modulated sine-band (MSB) speech signal composed of fixed-frequency amplitude-modulated sinewaves; and (3) a ''sinewave speech'' signal composed of sine waves tracking the first four formants of speech. In all three cases, a systematic decrease in performance in the two-talker target-ear listening task was found as the number of bands in the contralateral speech-like masker increased. These results suggest that speech-like fluctuations in the spectral envelope of a signal play an important role in determining the amount of across-ear interference that a signal will produce in a dichotic cocktail-party listening task. [DOI: 10.1121/1.1835509]

PACS numbers: 43.66.Pn, 43.66.Rq, 43.71.Gv [AK] Pages: 292–304

I. INTRODUCTION

Of all the difficult acoustic environments that occur in the everyday lives of human listeners, some of the most challenging involve the so-called ''cocktail party problem'' of listening to what one talker is saying when other talkers are speaking at the same time (Cherry, 1953). From a signal processing standpoint, this problem is extremely difficult, and even after years of intensive research the designers of automatic speech recognition systems still have not developed adequately robust algorithms for segregating speech in a wide variety of multitalker environments (Stern, 1998). Yet, normal-hearing human listeners are generally quite capable of understanding speech even in extremely complex situations that involve multiple simultaneous talkers in a reverberant environment.

Over the past 50 years, a great deal of research has been devoted to determining how listeners are able to achieve this success [see Yost (1997), Bronkhorst (2000), and Ebata (2003) for recent reviews of this literature]. In part, the answer lies in the inherent ability of human listeners to exploit differences in the voice characteristics of the different talkers, either in terms of fundamental frequency and intonation (Brokx and Nooteboom, 1982; Darwin and Hukin, 2000; de Cheveigne, 1993), vocal tract length (Darwin *et al.*, 2003), or overall speaking level (Egan *et al.*, 1954; Brungart, 2001b). In most situations, however, these monaural speech segregation cues are augmented by the binaural interaural level differences (ILDs) and interaural phase differences (IPDs) that occur when the target and interfering speech signals originate from different spatial locations relative to the listener (Bronkhorst and Plomp, 1988). These binaural difference cues enhance multitalker speech segregation in two ways: first, they introduce acoustic differences in the signals at the two ears that can be equivalent to as much as a 6–10-dB increase in the effective signal-to-noise ratio (SNR) of the target speech [e.g., see Zurek (1993)]; and second, they cause the target and masking signals to appear to originate from different locations in space, thus making it easier to selectively attend to one of the two speech signals (Freyman *et al.*, 1999).

In real-world listening environments, it is difficult to determine relative contributions these two types of binaural segregation cues make to the spatial unmasking of speech. Because all sound sources in realistic environments transmit some energy to each of the listener's two ears, some portion of the target speech signal will always be acoustically masked out by the interfering speech no matter how far apart the two sources are located. Thus, to the extent that listeners are unable to segregate widely separated speech signals in the free field, we cannot be sure whether the reason is because some portion of the target signal was obscured by the masker or because the two talkers did not ''sound'' far enough apart for the listener to perfectly segregate them.

[a]Electronic mail: douglas.brungart@wpafb.af.mil

There is, however, a somewhat artificial experimental manipulation that can be used to by-pass this inherent problem in real-world speech segregation. By presenting the target and masking signals ''dichotically'' over headphones (i.e., with one talker in one ear and one talker in the other ear), it is possible to generate a stimulus with two talkers who appear to originate from different places but have no acoustic overlap that could lead to energetic masking of the target.

Most of the experiments that have been conducted in these kinds of dichotic listening situations have shown that audio signals presented in one ear have little or no impact on the ability of normal-hearing adults to selectively attend to unrelated audio signals in the other ear. For example, Cherry (1953) has shown that a listener's ability to attend to a monaurally presented speech signal is unaffected by the presence of a distracting speech signal in the opposite ear. Other researchers have found similar results for the perception of dichotically separated speech signals (Drullman and Bronkhorst, 2000) and for the detection of tones in the presence of contralaterally presented random-frequency informational maskers (Neff, 1995; Wightman *et al.*, 2003). However, recent results have shown that the ability to ignore a distracting sound in the unattended ear can break down when a second distracting sound is also present in the same ear as the target signal. For example, Kidd and his colleagues (Kidd *et al.*, 2003) have shown that the presence of a random-frequency masker in the listener's unattended ear can sometimes impair the detection of a monaurally presented tone in the opposite ear when a second random-frequency masker is simultaneously presented in the same ear as the target tone. Similarly, Brungart and Simpson (2002) have shown that the presence of an interfering speech signal in the unattended ear can substantially impair the comprehension of a target speech signal in the opposite ear when a second independent interfering signal is simultaneously presented in the same ear as the target speech. Although other studies of dichotic speech perception have shown that listeners who are instructed to attend to a monaurally presented speech signal can be distracted by speech signals in the unattended ear that contain information that is surprising, unexpected, and/or relevant to the listener [such as an unexpected occurrence of the listener's first name (Moray, 1959; Wood and Cowan, 1995; Conway *et al.*, 2001)] or related in some way to the speech signal in the target ear [such as a midsentence swap between the signals in the target and unattended ears (Treisman, 1960)], historically there has been little evidence that irrelevant speech signals generate substantial amounts of across-ear interference in dichotic speech perception. The significance of Brungart and Simpson's (2002) finding is that it indicates that listeners in a dichotic listening task can be distracted by speech signals presented in the unattended ear even when those signals are unrelated to the target speech signals and completely devoid of any information that might be of interest to the listener outside the scope of the experimental task.

One intriguing aspect of Brungart and Simpson's dichotic speech segregation experiment was that significant across-ear interference occurred only for contralateral signals that were qualitatively ''speech-like:'' single-talker speech, multiple-talker speech, and time-reversed speech all caused across-ear interference, but speech-shaped noise did not. Furthermore, when the signal-to-noise ratio (SNR) in the target ear was less than 0 dB, time-reversed speech actually caused just as much across-ear interference as normal speech. Thus, it appears that, despite their obvious dissimilarities, normal speech signals and time-reversed speech signals share a common set of acoustic features that (a) interfere in some way with central speech processing, and (b) are not present in Gaussian noise. This conclusion suggests that some important insights into the processes that listeners use to segregate competing speech signals could be obtained by identifying the acoustic characteristics that cause audio signals to produce across-ear interference in dichotic listening. Furthermore, there is reason to believe that the underlying mechanisms that cause across-ear interference to occur for contralateral speech maskers in Brungart and Simpson's dichotic task might also extend into more realistic binaural listening situations where the target and masking signals are presented in different directions relative to the listener rather than in completely different ears. Indeed, such an effect might explain the relatively larger degradations in performance that have been shown to occur when a second speech masker is added to a stimulus containing two spatially separated competing speech sigals opposed to when a second noise masker is added to a stimulus containing a speech signal masked by a spatially separated noise source. Peissig and Kollmeier (1997), for example, found a 6.2-dB increase in speech reception threshold (SRT) when a second interfering talker was added to a speech signal masked by one competing talker, but only a 2-dB increase in SRT when a second interfering noise was added to a speech signal masked by one competing noise source. In a similar study, Hawley *et al.* (2004) reported a 9-dB increase in SRT with the addition of a second speech competitor to a stimulus containing two spatially separated speech signals, but only a 4-dB increase with the addition of a second noise competitor to a stimulus containing a target speech signal masked by a single spatially separated noise. Relatively large degradations in performance have also been shown to occur when a second interfering talker is added to a monaural stimulus containing two competing talkers (Brungart *et al.*, 2001; Hawley *et al.*, 2004). All of these results might be closely related to the Brungart and Simpson finding that listeners are able to use spatial location to segregate a target speech signal from one competing talker, but that they are unable to use location to segregate a speech signal from two competing talkers at different locations at the same time.

In this paper, we attempt to further explore the acoustic characteristics that cause a signal to interfere with dichotic speech segregation by examining the across-ear interference effects of three different types of highly intelligible but qualitatively unnatural synthetic speech signals and comparing them to the across-ear interference effects of normal speech. The results are discussed in terms of their implications for human speech segregation.

II. GENERAL METHODOLOGY

All three of the experiments conducted in this study were based on the coordinate response measure (CRM) for multitalker communications research, a call-sign, color, and number-based intelligibility test (Moore, 1981) that is particularly well suited for listening tasks that involve more than one simultaneous speech signal (Moore, 1981; Brungart *et al.*, 2001; Brungart and Simpson, 2002). In a typical trial in the CRM task, a listener is presented with one or more sentences of the form ''Ready (call sign) go to (color) (number) now'' and asked to identify the color and number combination that was directly addressed to a preassigned ''target'' call sign (usually ''baron''). In this series of experiments, the CRM phrases were drawn from a publicly available corpus (Bolia *et al.*, 2000) that consists of CRM phrases spoken by four male and four female talkers with all possible combinations of eight call signs (''arrow,'' ''baron,'' ''charlie,'' ''eagle,'' ''hopper,'' ''laker,'' ''ringo,'' ''tiger''), four colors (''blue,'' ''green,'' ''red,'' ''white''), and eight numbers (1–8), for a total of 2048 unique sentences.

Two different types of experiments were conducted with each of the three different synthetic speech signals examined in this study. Both involved listeners who were seated at one of three identical Windows-based PC computers located in three different quiet listening rooms. The first type of experiment was a straightforward single-talker listening experiment that examined the overall intelligibility of the different synthetic CRM speech signals. In each trial of these intelligibility experiments, a target phrase was randomly selected from all the available synthetic phrases containing the target call sign ''baron,'' scaled to a comfortable listening level (roughly 70 dB SPL), and presented to the listener over headphones (AKG240) through a 24-bit sound card (Creative Labs Audigy). The listener's task was simply to use the computer mouse to select the color and number combination contained in the stimulus from a grid of colored digits displayed on the CRT of the control computer.

The second type of experiment was a replication of the dichotic CRM listening task first used by Brungart and Simpson (2002). In each trial of this task, the signal presented to the right (target) ear always consisted of a mixture of two simultaneous phrases from the unprocessed natural-speech CRM corpus: a target phrase, which was randomly selected from the phrases containing the call sign ''baron,'' and a masking phrase, which was randomly selected from all the phrases spoken by a different same-sex talker that contained a different call sign, color, and number from the target phrase. The rms level of the target phrase was also scaled relative to the masking phrase to produce one of five different signal-to-noise ratios (−8, −4, 0, 4, or 8 dB).

The signal presented to the left (unattended) ear consisted of (a) silence; (b) a second masking phrase randomly selected from all the phrases in the standard CRM corpus spoken by a different talker of the same sex as the target talker that contained a different call sign, color, and number than either of the two phrases in the target ear; or (c) a synthetic CRM speech signal that was generated according to the procedures outlined in the following sections.

The participants in this dichotic CRM task were instructed to listen in the right ear for the target phrase containing the call sign ''baron'' and respond by selecting the color and number coordinates contained in that target phrase from the array of colored digits displayed on the screen of the control computer.

The next three sections describe how these experiments were implemented with the three different types of synthetic speech signals that were examined in this investigation of dichotic cocktail-party listening.

III. MODULATED NOISE-BAND SPEECH

One example of a stimulus that is qualitatively much different from speech but still highly intelligible is modulated noise-band (MNB) speech. MNB speech consists of fixed-frequency bands of noise that are independently amplitude modulated to match the envelopes of the corresponding frequency regions in an arbitrary target speech signal (Shannon *et al.*, 1995). When MNB speech is generated from a relatively large number of independently modulated bands of noise, it closely resembles whispered or unvoiced speech. However, as the number of modulated bands is reduced, the spectral detail in the target speech signal is lost and the MNB speech becomes progressively less similar to normal speech. Previous research has shown that MNB speech produces near-perfect vowel intelligibility with eight or more frequency bands, and near-perfect sentence intelligibility with five or more frequency bands (Dorman *et al.*, 1997). As the number of bands is reduced below five, intelligibility systematically decreases until it approaches chance performance in the one-band case where the stimulus is reduced to an amplitude-modulated broadband noise.

As discussed earlier, previous experiments have shown that continuous noise produces little or no across-ear interference in dichotic listening, but that speech does. Because MNB speech systematically changes from a qualitatively noise-like stimulus to a more speech-like stimulus as the number of frequency bands increases, one might also expect the number of frequency bands in MNB speech to influence the amount of across-ear interference it causes in dichotic listening. Experiment 1 was conducted to test this hypothesis. The experiment was divided into two parts. Experiment 1a examined MNB speech intelligibility as a function of the number of independently modulated frequency bands in the stimulus. Experiment 1b examined the contralateral interference effects these MNB stimuli caused in a dichotic cocktail-party listening task.

A. Experiment 1a: Intelligibility

1. Methods

a. Listeners. Nine paid volunteer listeners (four male and five female) participated in the experiment. All had clinically normal hearing (thresholds less than 15 dB HL from 500 Hz to 8 kHz), and their ages ranged from 19–53 years. All of the listeners had participated in previous experiments that utilized the speech materials used in this study.

b. MNB speech materials. For the purposes of this study, only a subset of the standard CRM corpus was processed to generate MNB speech. This subset consisted of all the phrases containing the call signs ''tiger,'' ''eagle,'' and

TABLE I. Cutoff frequencies (in kHz) of the independent frequency bands used to generate the MNB speech in experiment 1.

Number of bands	Start	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.05	4.00														
2	0.05	0.86	4.00													
3	0.05	0.47	1.45	4.00												
5	0.05	0.26	0.61	1.18	2.17	4.00										
10	0.05	0.14	0.26	0.41	0.61	0.86	1.18	1.61	2.17	2.94	4.00					
15	0.05	0.11	0.18	0.26	0.35	0.47	0.61	0.77	0.96	1.18	1.45	1.78	2.17	2.66	3.25	4.00

''baron'' spoken by two male talkers (talkers 2 and 3 from the corpus) and two female talkers (talkers 6 and 7 from the corpus), for a total of 384 phrases.

The phrases were converted to MNB stimuli with the PRAAT speech processing software package (Boersma, 1993). The phrases were first downsampled to 20 kHz and low-pass filtered at 4 kHz. They were then converted into the frequency domain with an FFT, divided into the required number of subbands,[1] and converted back in the time domain where the intensity contours of each subband were extracted by squaring the signals and convolving them with a 64-ms Kaiser window. A pink-noise excitation signal was then converted into the frequency domain, divided into the same number of subbands as the speech stimulus, and converted back into the time domain. Each subband of this noise stimulus was amplitude modulated with the intensity contour extracted from the corresponding subband of the speech signal, and the resulting amplitude-modulated noise bands were added together to construct the final MNB speech signal.

Six different MNB stimuli were generated for each phrase in the reduced corpus, each with a different number of independently modulated frequency bands (1, 2, 3, 5, 10, and 15). Thus, a total of 2304 sentences was available for use in the experiment. Note that the frequency bands were equally spaced on an ERB scale in the range from 50 Hz to 4 kHz, as illustrated in Table I.

c. Procedure. The experiment was conducted according to the procedures for CRM intelligibility testing outlined in Sec. II. The data collection was divided into six blocks of 60 trials, with each block containing ten trials for each of the six possible numbers of bands in the MNB corpus (1, 2, 3, 5, 10, or 15). Thus, each listener participated in a total of 60 trials for each number of bands tested in the experiment.

2. Results and discussion

The results of experiment 1a are shown in the left panel of Fig. 1. The intelligibility of the MNB speech increased systematically from around 15% to near 100% as the number of bands increased from one to five. For comparison, we have also replotted the results for the two speech corpora (out of a total of five tested) that produced the best and worst overall performance in Dorman *et al.*'s (1997) evaluation of the intelligibility of MNB speech: the Iowa Consonant Test of 16 consonants in an /aCa/ format spoken by a single male talker (Tyler *et al.*, 1986) [which was also the speech corpus used in the earlier study by Shannon (1995)]; and a multitalker vowel intelligibility test comprised of the 11 vowels in the words ''hawed, heed, hid, hayed, head, had, hod, hood, hoed, who'd, and heard'' spoken by three men, three women, and three girls (Hillenbrand *et al.*, 1995). These results show that the intelligibility levels obtained with the CRM corpus used in this experiment were roughly comparable to those reported for the relatively easy Iowa Consonant Test used in earlier MNB experiments by Shannon (1995) and Dorman *et al.* (1997).

B. Experiment 1b: Across-ear interference

1. Methods

a. Listeners. The same nine listeners who participated in experiment 1a also participated in experiment 1b.

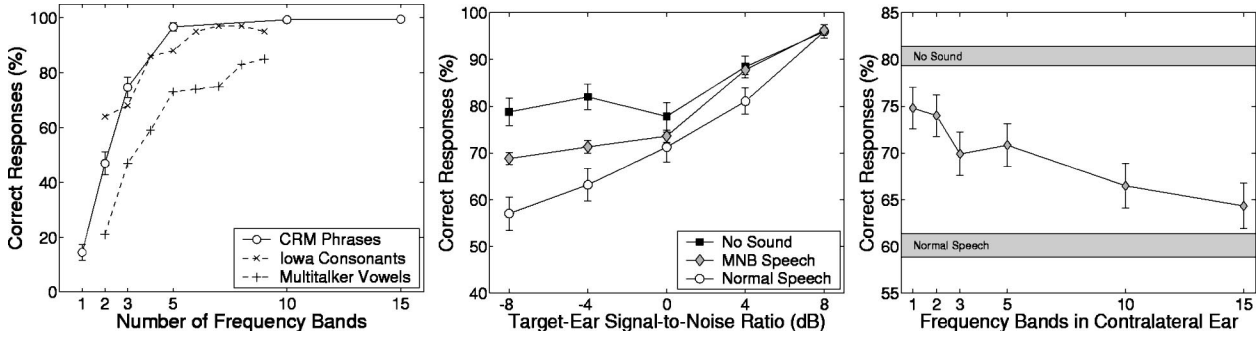


FIG. 1. The open circles in the left panel show the percentage of trials in which the listeners correctly identified both the color and number coordinates in experiment 1a, which measured speech intelligibility as a function of the number of frequency bands in the MNB speech stimuli. For comparision, the results obtained by Dorman *et al.* (1997) for similarly processed Iowa consonants and multitalker vowels have also been replotted in this panel. The center panel shows the percentage of correct color and number identifications in experiment 1b as a function of target-ear SNR. The black squares and open circles show performance in the control conditions where there was no contralateral masker or a normal-speech masker. The shaded diamonds show performance averaged across all the conditions with a contralateral MNB speech masker. The right panel shows the percentage of correct color and number identifications in the negative target-ear SNR conditions of experiment 1b. The shaded bars in that panel show mean performance ±1 standard error in the no-sound and normal-speech control conditions. The error bars represent the 95% confidence intervals for each data point.

b. Procedure. The experiment was conducted according to the procedures for the dichotic CRM task outlined in Sec. II. In the conditions where the masking phrase presented in the left ear consisted of synthetic speech, that masking phrase was randomly selected from the MNB-processed CRM phrases that contained a different call sign, color, and number than either of the two phrases in the target ear.[2] When the normal speech phrase was used in the unattended ear, it was low-pass filtered to 4 kHz to match the bandwidth of the MNB-processed speech stimuli and then scaled to match the rms level of the masking talker in the target ear. When the MNB speech was used in the unattended ear, it was also scaled to match the overall rms level of the masking talker in the target ear.

The data collection was divided into 40 blocks of 80 trials, with two repetitions of each of the eight possible contralateral masking conditions (silence, normal speech, or 1-, 2-, 3-, 5-, 10-, or 15-band MNB speech) at each of the five target-ear SNR values in each block. Thus, each of the nine listeners participated in a total of 80 trials for each combination of contralateral masker and target-ear SNR tested in the experiment.

2. Results and discussion

The results of the experiment are shown in the middle and right-hand panels of Fig. 1. The middle panel shows performance as a function of the SNR in the target ear for the conditions with no sound, MNB speech, or normal speech in the contralateral ear. For simplicity, all of the different MNB conditions have been averaged together to create the middle curve in the panel. In the no-sound and normal-speech control conditions, the results were similar to those in an earlier experiment that used the same CRM stimuli and the same dichotic listening task used in this experiment (Brungart and Simpson, 2002). In the condition with no contralateral masker (black squares), performance increased as the SNR increased above 0 dB, but plateaued at approximately 80% correct responses for SNR values at or below 0 dB. In the condition with a normal speech contralateral masker (open circles), performance was similar to the no-sound condition when the SNR was +8 dB, but it decreased much more rapidly with decreasing SNR. As a consequence, performance at −8-dB SNR was roughly 20 percentage points worse with a contralateral speech masker than it was with no contralateral masking signal. The gray diamonds show performance averaged across the six MNB speech conditions of the experiment. As we hypothesized, the results for the MNB speech consistently fell between those for the no-sound and normal-speech contralateral masking conditions. This suggests that MNB speech causes more contralateral interference than no masker, but less interference than a normal speech masker.

The middle panel of Fig. 1 also indicates that the contralateral maskers had the greatest impact on performance when the target-ear SNR was less than 0 dB. Consequently, the right panel of Fig. 1 focuses on the differences between the MNB-speech conditions in trials where the target-ear SNR was negative. For comparison, shaded regions of the figure show mean performance ±1 standard error in the no-sound and normal-speech control conditions of the experiment. These results show that there was indeed a systematic decrease in performance as the number of frequency bands in the MNB speech increased. A one-factor within-subjects ANOVA on the arcsine-transformed results of the individual subjects for each of the eight contralateral masking conditions (no sound, normal speech, and 1-, 2-, 3-, 5-, 10-, or 15-band MNB speech) indicated that this effect was statistically significant ($F_{(7,56)}=15.58$, $p<0.0001$), and a subsequent *post hoc* test (Fisher LSD, $p<0.05$) indicated the following significant results:

(1) All the MNB speech conditions were significantly worse than the no-sound control condition.
(2) All the MNB speech conditions except the 15-band condition were significantly better than the normal-speech control condition.

Thus, it seems that even the single-band MNB speech distractor, which scored only slightly better than chance in the intelligibility test in experiment 1a, produced a significant amount of across-ear interference in the dichotic listening task of experiment 1b. As the number of frequency bands increased, so did the across-ear interference caused by the MNB speech. However, the amount of interference did not plateau at the 5-band level where intelligibility reached near 100% performance in experiment 1a. Rather, it continued to increase until the 15-band point, where the MNB speech was producing nearly as much contralateral interference as normal speech.

IV. MODULATED SINE-BAND SPEECH

Modulated noise-band speech is qualitatively much different from normal voiced speech, but when it consists of a large number of frequency channels it can sound similar to whispered or unvoiced speech. Thus, it is conceivable that the increase in across-ear masking that occurred in the 15-band condition of experiment 1 could be directly related to the similarity of the speech in that condition to natural whispered speech. It is possible, however, to generate a stimulus that contains the spectral information similar to MNB speech but sounds unnatural even when it contains a large number of frequency channels. This speech is generated by replacing the amplitude-modulated noise bands in MNB speech with amplitude-modulated sine waves fixed at the center frequencies of those bands. Previous experiments that have compared this type of modulated sine-band (MSB) speech to MNB speech have found very little difference in intelligibility between the two types of simulated speech (Dorman *et al.*, 1997), despite the large qualitative difference between the two types of speech signals. Experiment 2 was conducted to evaluate the amount of across-ear interference generated by MSB speech in a dichotic cocktail-party listening task.

A. Experiment 2a: Intelligibility

1. Methods

a. Listeners. Eight paid volunteer listeners with clinically normal hearing (five male and three female) participated in the experiment. Six of the listeners were also participants in experiments 1a and 1b.

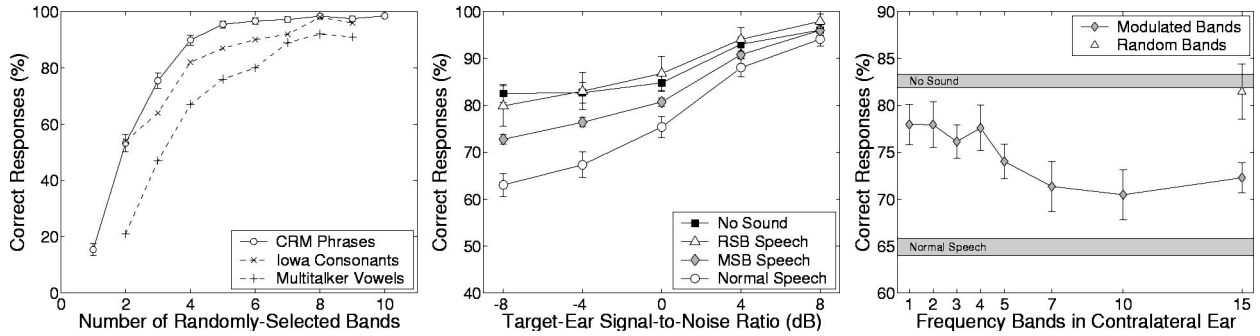


FIG. 2. The left panel shows the percentage of correct color and number identifications in experiment 2a, which measured speech intelligibility as a function of the number of frequency bands in the MSB speech stimuli. As in Fig. 1, the intelligibility results obtained by Dorman *et al.* (1997) for MSB-processed Iowa consonants and multitalker vowels have been replotted in this panel for comparison. The center panel shows the percentage of correct color and number identifications in experiment 2b as a function of target-ear SNR. The black squares and open circles show performance in the control conditions where there was no signal in the contralateral ear (squares) or a normal-speech signal in the contralateral ear (circles). The shaded diamonds show performance averaged across all the conditions with a contralateral MNB speech masker. The right panel shows the percentage of correct color and number identifications in the negative target-ear SNR conditions of experiment 1b. The shaded bars in that panel show mean performance ±1 standard error in the no-sound and normal-talker control conditions. The error bars represent the 95% confidence intervals for each data point.

b. Speech materials. The MSB speech stimuli were derived from the male-talker sentences from the same CRM speech corpus used in experiment 1.[3] These stimuli were processed with a technique that Arbogast *et al.* (2002) adapted from cochlear implant simulation software originally developed by the House Ear Institute. The sentences in the CRM corpus were first downsampled from 40 to 20 kHz. Then, they were high-pass filtered at 1200 Hz with a first-order Butterworth filter and processed with a bank of 15 fourth-order 1/3rd-octave Butterworth filters with logarithmically spaced center frequencies ranging from 215 to 4891 Hz with a ratio of successive center frequencies of 1.25. The envelopes of each of these channels were extracted by half-wave rectifying the bandpass-filtered signals and low-pass filtering them at 50 Hz. Then, these envelopes were used to modulate pure tones with zero starting phases and center frequencies at the midpoints of each filter band. Individual sound files were created for each of these 15 bands for the 256 CRM phrases spoken by each of the male talkers in the CRM corpus, and the stimuli used in the experiment were generated by randomly selecting 1–10 of these individual bands from the same original CRM phrase and summing them together electronically.[4]

c. Procedure. Other than the method used to generate the speech stimuli, the experimental procedure was essentially identical to the one used in experiment 1a. Each block of trials in the experiment consisted of 12 repetitions of each of the 10 MSB speech conditions of the experiment (i.e., 1–10 individual randomly selected bands). Each listener participated in 10 blocks of trials, so a total of 960 trials was collected in each of the 10 conditions of the experiment (8 listeners×10 blocks×12 repetitions).

2. Results and discussion

The left panel of Fig. 2 shows the intelligibility results from experiment 2a. Intelligibility was poor (<20%) in the one-band condition, but it increased systematically with the number of bands, plateauing at near 100% performance when five independent frequency bands were present in the stimulus. Overall, this performance function is very similar to the one obtained with the MNB-processed CRM stimuli in experiment 1a (plotted in the left panel of Fig. 1). The intelligibility scores were, however, slightly higher than those reported for the MSB-processed Iowa consonants in the earlier experiment by Dorman *et al.* (1997), which have been replotted in the figure for comparison. Comparing Figs. 1 and 2, it is apparent that the CRM stimuli used in this experiment produced intelligibility levels that were very similar to those obtained for the Iowa consonants in the MNB processing condition, but somewhat better than those obtained for the Iowa consonants in the MSB condition. This difference may, in part, be due to the fact that Dorman and his colleagues generated their MSB stimuli with modulated sinewave bands that were always evenly distributed across the speech spectrum, while the stimuli in this experiment were generated with modulated sinewave bands that were randomly selected from the 1/3rd-octave bands that were available in the 15-band MSB processed speech. The difference might also simply be due to the semantic differences between the two speech corpora. In either case, the results shown in Fig. 2 indicate that the random-frequency MSB speech used in experiment 2a produced intelligibility in the CRM task that was comparable to that obtained for MNB speech generated with the same number of frequency bands in experiment 1a.

B. Experiment 2b: Across-ear interference

1. Methods

a. Listeners. Seven of the eight listeners who participated in experiment 2a also participated in experiment 2b.

b. Speech materials. The MSB conditions of experiment 2b used the same stimulus processing as described in experiment 2a. In addition to these MSB speech conditions, a 15-band random sine-band (RSB) speech control condition was also tested. The RSB speech was produced by randomizing the phase component of a standard 15-band MSB speech signal. This was accomplished by multiplying the long-term complex spectrum (FFT) of a randomly selected 15-band MSB speech signal by the long-term complex spectrum of a broadband Gaussian noise and taking the inverse FFT of this

multiplied frequency-domain signal (Arbogast *et al.*, 2002). This processing resulted in an unintelligible waveform that was spectrally identical to the MSB speech but contained no phonetic information about the original utterance.

c. Procedure. Experiment 2b used the same dichotic CRM task used in experiment 1b, with the exception that only two of the talkers were used as target talkers (the male talker 1 and the female talker 6) with the same target talker used in every stimulus presentation within the same block of trials. In the conditions where the masking phrase presented in the left ear consisted of synthetic speech, that masking phrase consisted of MSB speech with 1, 2, 3, 5, 7, 10, or 15 bands or RSB speech with 15 bands. In all cases, the masking speech signal was selected to have a different color and number than either of the two phrases in the target ear.

The data collection was divided into blocks of approximately 70 trials with each subject participating in roughly 100 blocks, for a total of 6864 trials per subject or 48 048 trials in the experiment. All subjects participated in all conditions, and the total number of trials per condition ranged from 1698 trials for the 15-band RSB speech condition to 7305 trials for the 15-band MSB speech condition.

2. Results and discussion

The middle and right-hand panels of Fig. 2 show the overall results of experiment 2b. The middle panel shows performance as a function of the SNR in the target ear for the conditions with no sound, RSB speech, MSB speech, or normal speech in the contralateral ear. Again, the different MSB conditions have been averaged together to simplify the visual presentation of the data in this panel. The results show that the no-sound (black squares) and normal-speech (open circles) control conditions were essentially identical to the corresponding conditions of experiment 1b (shown in Fig. 1). Also, as with the MNB speech in experiment 1b, the results with the MSB speech in experiment 2b consistently fell between these two control conditions. In contrast, performance with the 15-band RSB speech (open triangles) was essentially identical to the no-sound control condition.

The right panel of Fig. 2 shows performance in the different MSB-speech conditions averaged across trials where the target-ear SNR was less than 0 dB. Again, the shaded regions of the figure show mean performance ± 1 standard error in the no-sound and normal-speech control conditions of the experiment. Performance in the 15-band RSB condition is also shown by the white triangle. The arcsine-transformed data from the individual subjects in each of the 11 contralateral masking conditions (no sound, normal speech, 1-, 2-, 3-, 4-, 5-, 7-, 10-, or 15-band MSB speech or RSB speech) were also subjected to a one-factor within-subjects ANOVA, which indicated that the main effect of the contralateral masking condition was statistically significant ($F_{(10,60)}=9.52$, $p<0.0001$). A subsequent *post hoc* test (Fisher LSD, $p<0.05$) revealed the following significant effects:

(1) All the MSB speech conditions except the 1-band condition were significantly worse than the no-sound control condition.[5]
(2) All the MSB speech conditions were significantly better than the normal-speech control condition.
(3) The 1-, 2-, and 4-band conditions were significantly better than the 7-, 10-, and 15-band conditions.
(4) There was no significant difference between the 15-band RSB condition and the no-audio control condition.

As in the MNB condition, the results show a general trend of increasing across-ear interference with an increasing number of frequency bands. However, in the limiting 15-band case, performance appeared to be slightly better relative to the normal-speech control condition with MSB speech. This may reflect the fact that 15-band MNB speech sounds similar to natural whispered speech, while MSB speech sounds decidedly unnatural even with 15 frequency bands.

It is also interesting to note that the RSB speech failed to produce any measurable across-ear interference even though it contained all 15 possible frequency bands. The long-term magnitude spectrum of this RSB speech signal was identical to that of the 15-band MSB speech, so it seems that the across-ear interference caused by the MSB speech cannot be explained by spectral content alone. Rather, it seems that the speech-like temporal modulations in the individual bands of the MSB speech were critical to the across-ear interference effects that occurred with those stimuli. This seems to be consistent with our earlier finding that the contralateral noise that was shaped to match the long-term rms spectrum of CRM speech produced little or no across-ear interference in the dichotic CRM task (Brungart and Simpson, 2002). It is also consistent with the results of Arbogast *et al.* (2002), who also found a substantial difference between the masking properties of MSB and RSB speech in normal binaural listening environments. In their experiment, they randomly selected 8 of the 15 bands for use in the target speech signal, and allocated 6 of the remaining bands either to an MSB speech masker or an RSB speech masker. Their results showed that the speech reception threshold (SRT) was 22 dB lower with RSB masking speech than it was with MSB masking speech, presumably because the speech-like MSB masker was more easily confused with the target speech signal. Our results show that this masking difference between MSB and RSB speech extends to the case where the target and masking speech signals are presented to different ears.

V. SINEWAVE SPEECH

An additional type of ''speech-like'' stimulus that is qualitatively different from speech but still highly intelligible is so-called ''sinewave speech,'' which consists of a small number of time-varying amplitude-modulated sine waves that track the formant frequencies of a speech signal (Remez *et al.*, 1981). Experiment 3 was conducted to determine whether this kind of stimulus also produces across-ear interference in a dichotic cocktail-party listening task.

A. Experiment 3a: Intelligibility

1. Methods

a. Listeners. Nine paid volunteer listeners with clinically normal hearing (four male and five female) participated

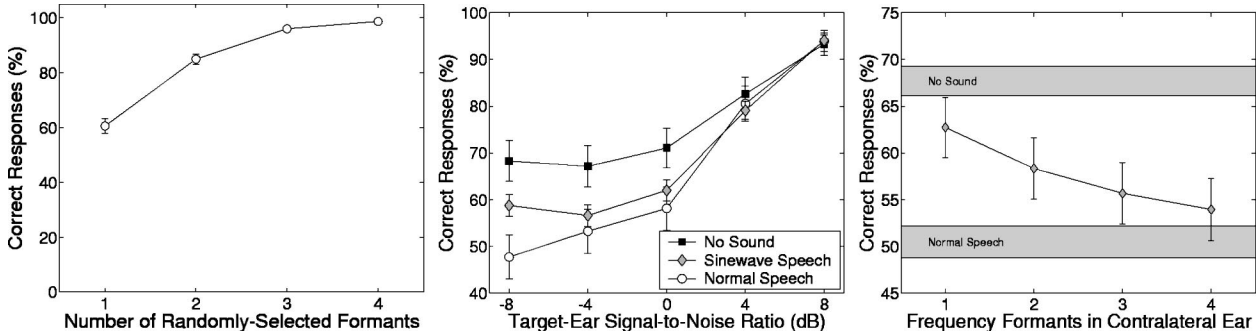


FIG. 3. The left panel shows the percentage of correct color and number identifications in experiment 3a, which measured speech intelligibility as a function of the number of formants in the sinewave speech stimuli. The center panel shows the percentage of correct color and number identifications in experiment 3b as a function of target-ear SNR. The right panel shows the percentage of correct color and number identifications in the negative target-ear SNR conditions of experiment 3b, which measured the effects of a contralateral sinewave speech interferer on two-talker segregation performance in the listener's right ear. The error bars represent the 95% confidence intervals for each data point.

in the experiment. Six of the nine listeners had previously participated in experiment 1, and four had previously participated in experiment 2.

b. Speech Materials. The sinewave speech stimuli were processed directly from the CRM speech corpus using LPC-based MATLAB script files that have been made publicly available on the Internet by Ellis (2003). These scripts estimate the magnitudes and frequencies of the first four formants in each 2.6-ms frame from the filter pole positions derived from an LPC analysis. The CRM sentences were resampled to an 8-kHz rate prior to performing this LPC analysis, resynthesized into sinewave speech, and then resampled to a 50-kHz rate prior to presentation to the listeners. This processing was done in real time within each trial of the experiment.

c. Procedure. Again, the procedure used in experiment 3a was essentially identical to the one used in experiments 1a and 2a. In each trial of the experiment, a target phrase was randomly selected from all the phrases containing the target call sign "baron" in the CRM corpus. This target phrase was processed into sinewave speech, and then one, two, three, or four of the first four formants were randomly selected for inclusion in the stimulus. The data collection was divided into 10 blocks of 60 trials, with each block containing 15 trials for each of the four possible numbers of formants (1, 2, 3, or 4). Thus, each listener participated in a total of 600 trials in the experiment.

2. Results and discussion

The left panel of Fig. 3 shows the intelligibility results from experiment 3a. The major difference between these results and the earlier results with the MNB and MSB speech signals in experiments 1a and 2a is the much higher intelligibility score that was achieved with just a single randomly selected formant (near 60%, versus less than 20% for the other two stimulus types). This reflects the fact that the sinewave speech adapts itself to track variations in the frequencies of the formants, while the MSB and MNB stimuli provide spectral information only in fixed frequency regions. Note that intelligibility approaches 100% for sinewave speech stimuli comprised of three or more randomly selected formants.

B. Experiment 3b: Across-ear interference

1. Methods

a. Listeners. The same nine listeners who participated in experiment 3a also participated in experiment 3b.

b. Procedure. The procedure used in experiment 3b was essentially identical to the one used in experiment 1b. When a synthetic speech signal was presented in the left ear, it consisted of sinewave speech that was generated with 1, 2, 3, or 4 randomly selected formants using the procedure outlined in the previous section. When a natural speech phrase was presented in the unattended ear, it was low-pass filtered to 4 kHz to match the maximum bandwidth of the sinewave speech stimuli. In all cases, the interfering speech signal in the contralateral ear was scaled to match the rms level of the masking talker in the target ear.

The data collection was divided into 24 blocks of 60 trials, with two repetitions of each of the six possible contralateral masking conditions (silence, normal speech, or 1-, 2-, 3-, or 4-band sinewave speech) at each of the five target-ear SNR values in each block. Thus, each of the nine listeners participated in 1440 trials in the experiment, for a total of 432 trials for each combination of target-ear SNR and contralateral-ear masker tested in the experiment.

2. Results and discussion

The results of experiment 3b are shown in the right two panels of Fig. 3. The middle panel of the figure shows performance as a function of the target-ear SNR. Again, the four sinewave-speech conditions have been averaged together into a single curve (shaded diamonds) to allow an easy comparison to the no-sound (black circles) and normal-speech (open circle) control conditions. Although performance in these control conditions was markedly lower than in experiments 1b and 2b (presumably because of the different mix of subjects), the overall pattern of performance was the same: a plateauing in performance at negative SNR values in the no-sound condition, and a roughly 20-percentage point decrease in performance in the normal-speech condition at an SNR of −8 dB.

Performance with the sinewave-speech contralateral maskers (gray diamonds) again fell between these two control conditions, with the largest decrease relative to the no-

sound condition occurring at negative target-ear SNR values. The right panel of Fig. 3 shows performance as a function of the number of formant frequencies in the contralateral sinewave speech masker averaged across trials where the target-ear SNR was less than 0 dB. As before, the arcsine-transformed data from the individual subjects in each of the six contralateral masking conditions (no sound, normal speech, and 1-, 2-, 3-, or 4-formant sinewave speech) were analyzed by a within-subjects ANOVA, which indicated that the main effect of the contralateral masking condition was statistically significant ($F_{(5,40)}=11.13$, $p<0.0001$). A subsequent *post hoc* test (Fisher LSD, $p<0.05$) found the following significant differences:

(1) All the sinewave speech conditions except the 1-formant condition were significantly worse than the no-sound control condition.
(2) All the sinewave speech conditions except the 4-formant condition were significantly better than the normal-speech control condition.
(3) Performance in the 1-formant condition was significantly better than the 3- and 4-formant conditions.

Thus, we see that, as with the other types of simulated speech signals tested in these experiments, sinewave speech tends to produce more across-ear interference than noise in dichotic listening, but less interference than normal speech. Also, the data suggest that sinewave speech may be somewhat more efficient at generating across-ear interference than MSB or MNB speech. Sinewave speech produced almost as much interference as normal speech with just 4 formant frequency bands, a level of interference that required 15 bands for the MNB speech and never occurred with the MSB speech. However, it should be noted that, like MSB speech, the sinewave speech stimuli never sounded remotely similar to any type of natural speech even with the largest number of frequency bands tested. Thus, it seems that the difference in across-ear interference that occurred between experiment 1b and experiment 2b cannot be accounted for solely by the whisper-like characteristics of MNB speech when it contains a large number of frequency bands.

VI. GENERAL DISCUSSION

This paper has presented the results of three experiments comparing the across-ear interference generated by three distinct types of simulated speech to the amount of across-ear interference that occurs with a normal speech signal. Although the three types of simulated speech were qualitatively much different, their contralateral masking characteristics were similar: (1) all produced some amount of contralateral interference when they contained only one or two frequency bands; (2) the amount of contralateral interference increased systematically with the number of frequency bands; and (3) performance for the maximum number of frequency bands tested approached the normal-speech control condition.

The results of the experiments described in this paper, along with those of our earlier study examining the effects of a contralateral masker on dichotic speech perception (Brungart and Simpson, 2002), allow us to answer a number of important questions regarding the across-ear interference that occurs in dichotic cocktail-party listening.

(1) *Is there a threshold level of similarity to speech that must be reached in order for a speech-like signal to generate across-ear interference in a dichotic cocktail-party listening task*? No. With all three of the synthetic speech stimuli we tested, the amount of across-ear interference increased gradually as the number of bands increased. Similarly, in our earlier experiment, there was a gradual decrease in across-ear interference as the speech signal in the contralateral ear was masked with noise (Brungart and Simpson, 2002). This argues against the existence of a ''threshold'' level of speech-like attributes that must be reached in order for a contralaterally presented speech signal to interfere with speech perception in the opposite ear.
(2) *Is long-term spectral similarity to speech necessary or sufficient for a signal to generate across-ear interference in a dichotic cocktail-party listening task*? No. In our earlier experiment, we showed that Gaussian noise that was spectrally shaped to match the long-term spectrum of speech caused little or no across-ear interference in dichotic listening. In this series of experiments, we demonstrated that at least two types of signals with long-term spectra that differed dramatically from normal speech (MSB speech and sinewave speech) generated substantial amounts of across-ear interference. From these two results, we can conclude that spectral similarity to speech is neither necessary nor sufficient for a sound to produce across-ear interference in dichotic listening. Further evidence for the relatively minor role that long-term spectrum plays in contralateral masking was provided by the results of experiment 2b: the long-term spectrum of the 15-band RSB speech contralateral masker used in that experiment was identical to the spectrum of the 15-band MSB speech, but the RSB speech produced far less contralateral interference than the MSB speech masker. Again, this suggests that overall spectrum is a relatively unimportant parameter in determining the amount of across-ear interference a contralateral masking signal will generate.
(3) *Is intelligibility necessary for a signal to generate across-ear interference in a dichotic cocktail-party listening task*? No. In our earlier experiment, we demonstrated that time-reversed speech produced just as much across-ear interference as normal speech when the target-ear signal-to-noise ratio was less than 0 dB. Thus, it appears that unintelligible signals can produce just as much contralateral interference as intelligible signals in dichotic listening.
(4) *Is intelligibility sufficient for a signal to generate across-ear interference in a dichotic cocktail-party listening task*? Probably. We have not tested all of the synthetic signals that could conceivably be used to generate intelligible speech, but we have examined three of the least speech-like signals that have been demonstrated to contain usable verbal information, and we have shown that

all three produce significant amounts of across-ear interference in dichotic listening. This leads us to suspect that any signal capable of conveying the useful phonetic information contained in normal speech will produce some across-ear interference in a dichotic cocktail-party task. However, we should point out that, to this point, we have only tested signals that have been gated on and off at approximately the same time as the target speech. It is possible that adaptation might allow listeners to perform better in the dichotic listening task if the contralateral masker were a continuous speech signal that was turned on some time before the onset of the target speech.

(5) *Are speech-like temporal modulations in the spectral envelope of a signal sufficient to generate across-ear interference in a dichotic cocktail-party listening task*? Yes. The MNB speech differed from broadband speech-shaped noise only in terms of the introduction of speech-like modulations in the spectral envelope, and these modulations were sufficient to generate a substantial amount of contralateral interference in the dichotic listening task. Similarly, the MSB speech differed from the RSB speech only in terms of its envelope modulations, and these modulations were sufficient to generate a substantial amount of across-ear interference. However, it is important to note that the modulations that appear to be most critical to the across-ear interference effects demonstrated in these experiments are the varying narrowband temporal modulations that occur in speech, and that the contralateral masking effects of these modulations are probably limited to listening tasks where the target signal is also speech-like. Listening tasks involving nonspeech target signals may be more sensitive to contralateral interference from signals with different qualitative characteristics and different modulation patterns. Kidd *et al.* (2003), for example, examined performance in a nonspeech dichotic listening task that required listeners to detect fixed-frequency pulsed tone targets in the presence of tone or noise maskers and found that contralaterally presented fixed-frequency tone complexes that were coherently gated with the target produced significant amounts of across-ear interference, but that contralaterally presented notch-filtered noise that was coherently gated with the target did not. Thus, in that case, significant across-ear interference only occurred when the contralateral masking signal was synchronously gated with *and* qualitatively similar to the target signal. Consequently, it is likely that the contralaterally presented synthetic speech signals that caused significant across-ear interference in this experiment would have little or no effect on performance in the dichotic tone-detection task examined by Kidd and his colleagues. Thus, while speech-like modulations appear to be sufficient to produce across-ear interference in dichotic speech perception tasks, other factors—such as qualitative target–masker similarity—can strongly influence the across-ear interference effects that occur in other kinds of listening tasks.

(6) *Are speech-like temporal modulations in the spectral envelope of a signal necessary to generate across-ear interference in a dichotic cocktail-party listening task*? Possibly. We have not yet tested any signals that generate a substantial amount of across-ear interference and do not have speech-like temporal envelope fluctuations. Thus, while we cannot rule out the possibility that such signals exist, we do not yet have any evidence to demonstrate that signals without speech-like envelope fluctuations can cause across-ear interference in dichotic cocktail-party listening.

(7) *What are the requirements for modulations in the spectral envelope of a signal to be "speech-like" in the sense that they will produce significant amounts of across-ear interference in a dichotic cocktail-party listening task*? This is perhaps the most interesting remaining research question related to the contralateral interference effects we have demonstrated in our dichotic listening experiments. All of our experiments to this point suggest that certain types of contralaterally presented audio signals are identified as ''speech-like'' by some preattentive central auditory processing mechanism, and that signals that fall into this category interfere with a listener's ability to segregate speech signals presented in the opposite ear. The results of this experiment strongly suggest that speech-like modulations in the spectral envelope play an important role in determining what kinds of signals are identified as speech-like by this central processing. Furthermore, our earlier results have shown that these speech-like fluctuations do not necessarily have to be intelligible to cause interference: time-reversed speech, which is unintelligible but has envelope fluctuations similar to those in normal speech, produces nearly as much across-ear interference as normal speech. At this point, however, it is not clear what the parameters are that determine whether or not these envelope fluctuations are speech-like. What range of modulation frequencies will generate this type of interference? Do the modulation frequencies have to vary over time like they do in natural speech, or will constant envelope modulations cause the same amount of contralateral interference? Do the modulations have to be correlated across frequency as they are in natural speech, or do independent speech-like envelope modulations (such as those that would occur with a stimulus matching the envelopes of different utterances at different frequency regions) also interfere? The answers to these questions are important, because they have the potential to provide valuable insights into the processing methods that listeners unconsciously use to segregate complicated auditory scenes containing more than one simultaneous speech signal. This information might also provide some new ideas about how to produce machine listening devices capable of segregating multiple-talker listening environments using the same strategies that human listeners use for these segregation tasks. At this point, however, only further research can provide the answers to these important questions about dichotic speech perception.

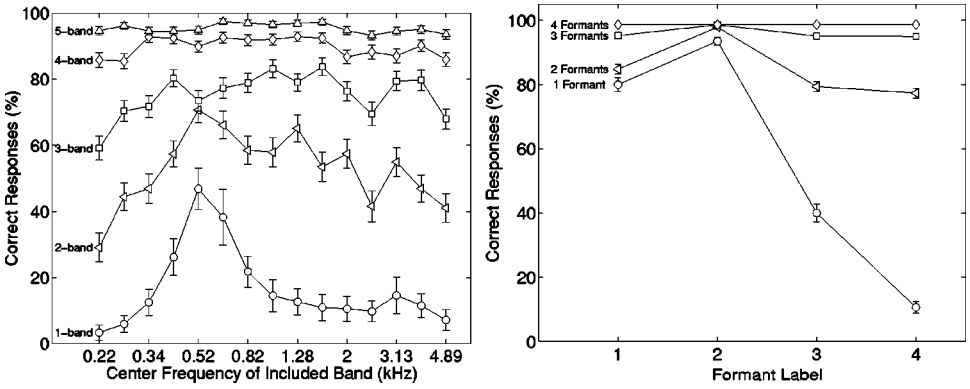


FIG. 4. Each curve in the left panel shows the percentage of correct responses in experiment 2a averaged across all the stimuli with the same number of frequency components that contained the indicated frequency band. Similarly, the curves in the right panel show the percentage of correct responses in experiment 3a averaged across all the stimuli with the same number of frequency components that contained the indicated (labeled) formant. The error bars represent ±1 standard error around each data point.

VII. SUMMARY AND CONCLUSIONS

In this series of three experiments, we have demonstrated that three subjectively very different types of synthetic speech signals (MNB speech, MSB speech, and sinewave speech) have similar effects on speech intelligibility when they are presented to the unattended ear in a dichotic cocktail-party listening task. In all three cases, there was a systematic decrease in performance in the two-talker target-ear listening task when the number of frequency bands in the contralateral speech-like masker increased. These results suggest that speech-like fluctuations in the spectral envelope of a signal play an important role in determining the amount of across-ear interference that signal will produce in a dichotic cocktail-party listening task.

In closing, it is perhaps useful to take a step back and consider how this finding relates to our more general understanding of how listeners process multiple simultaneous speech signals in real-world cocktail party listening environments. Clearly, the stimuli examined in this experiment are artificial in the sense that they would never occur in real-world listening. Indeed, even the more general realm of dichotic listening is somewhat unrealistic, because real-world speech signals are almost always perceived binaurally rather than monaurally. However, what these results do allow us to do is begin to gain some insights into the point at which the auditory system starts to make a distinction between signals that are speech-like and should be processed when the listener is performing a speech perception task and those that are ''noise-like'' and should be discarded. In the long term, these insights might also help us understand the acoustic features that make it difficult for listeners to segregate simultaneously presented speech signals that, from a purely acoustic standpoint, should individually be clearly audible [a concept sometimes referred to as informational masking (Kidd *et al.*, 1998; Freyman *et al.*, 2001, 1999; Brungart, 2001b)]. Further research is now needed to fully examine the relationship between the temporal fluctuations that occur in the envelopes of a speech-like masking signal and the amount of masking such a signal will produce when it is presented in the unattended ear in a dichotic cocktail-party listening task, and to determine the extent to which a similar kind of interference might occur in more realistic binaural cocktail-party listening tasks that more accurately represent the difficulties listeners encounter in real-world verbal communication.

ACKNOWLEDGMENTS

Portions of this research were supported by AFOSR Grant 01-HE-01-COR and NIH/NICDC Grants DC00100, DC045045, and DC04663.

APPENDIX: FREQUENCY WEIGHTING WITH MSB AND SINEWAVE SPEECH

In speech perception, different frequency regions vary in terms of their relative contributions to overall intelligibility. This attribute of speech perception is one of the foundations of the Articulation Index (AI), which assigns different weights to each 1/3rd-octave band to account for differences in the relative importance of each band in the perception of phonetically balanced speech (French and Steinberg, 1947). However, the coordinate response measure speech materials used in these experiments are not phonetically balanced, so their frequency-dependent intelligibility characteristics may differ from those that would ordinarily occur with traditional speech perception tasks (Brungart, 2001a). Thus, it may be useful to analyze the results of experiments 2a and 3a to examine the contributions that different frequency regions made to the overall perception of the CRM stimuli.

Figure 4 shows how performance varied across the possible frequency component combinations that could occur with the MSB stimuli in experiment 2a and with the sinewave speech stimuli in experiment 3a. In the left panel, each curve represents mean performance across all the MSB speech trials in experiment 2a that contained the indicated number of frequency bands. Within each curve, the data points represent mean performance across all the stimuli with that particular number of bands that contained the frequency component indicated by the abscissa. Thus, in the one-band curve (circles), each data point represents performance in stimulus presentations that contained only the designated frequency component. In the two-band curve (left-pointing triangles), each data point represents mean performance across all the trials that contained the designated band plus one other randomly selected band. And, in the five-band curve, each data point represents mean performance across all the trials that contained the designated band plus four other randomly selected bands.

From the one-band curve, it is immediately apparent that the most important frequency component for overall intelligibility in the CRM task was the modulated sinewave at 520

Hz. In fact, the listeners were able to correctly identify both the color and the number in the stimulus almost half the time when the 520-Hz component was the only frequency component present in the stimulus. In comparison, most of the other frequency bands generated only about 10% correct performance when they were presented in isolation. Interestingly, the 520-Hz band is much lower in frequency than the most highly weighted 1/3rd-octave band in the calculation of the AI, which is centered at 2 kHz (French and Steinberg, 1947). This suggests that the phonetic information in the CRM speech corpus is concentrated in a lower frequency range than the phonetic information in long-term running speech.

As additional frequency bands were added to the MSB CRM stimuli, the specific bands contained in the stimuli became increasingly less important until, in the four-band case, the presence of any particular band no longer had any meaningful impact on the overall intelligibility of the stimuli. This suggests that highly intelligible MSB speech can be generated from any five randomly selected bands out of the 15 frequency bands tested in experiment 2.

The right panel of Fig. 4 shows the relative intelligibility contributions of each of the four formant frequencies tested in the sinewave speech stimuli of experiment 3a. As before, each curve represents a different number of formants, and each data point represents average performance across all the trials with the designated number of sinewave components that happened to contain the formant indicated by the abscissa. Again, these results show that there were large differences across the different formant combinations when the number of frequency components was low. Indeed, the data show that intelligibility was close to 100% when just the second formant was present in the stimulus, but was less than 10% when only the fourth formant was present in the stimulus. The performance variations across the different possible combinations largely disappeared when a second formant was added to the stimulus.

The dramatic variations in performance across the different formant combinations in experiment 3a suggest that perhaps there might also have been a significant variation in the amount of contralateral interference caused by these different sinewave speech stimuli. In order to test this hypothesis, the overall performance level was calculated for each of the 15 possible combinations of 1, 2, 3, or 4 formants in the contralateral masking conditions of experiment 3b, and the correlation coefficient was calculated between the scores in these 15 conditions and the intelligibility scores in the 15 corresponding conditions of experiment 3a. The resulting r value was -0.009, suggesting that intelligibility may be a relatively poor predictor of the across-ear interference that will occur when a speech-like masking signal is presented in the unattended ear in a dichotic cocktail party listening task.

[1]This division was accomplished with the Filter (Pass Hann) command in PRAAT. See the PRAAT documentation (Boersma, 1993) for more details.

[2]The MNB speech could be male or female independent of the sex of the target talker. However, because all voicing information is removed, there is little or no apparent difference between male and female MNB speech.

[3]Again, because all of the voicing information was removed, there was little or no discernible difference between male and female talkers in the MSB speech.

[4]In comparing this technique to the one used to produce the stimuli in the earlier study by Dorman *et al.* (1997), it is important to note that this processing technique involves the exclusion of some speech envelope information when the number of bands is reduced, while with Dorman's technique the speech envelope information is not excluded but rather averaged over a larger bandwidth when the number of bands decreases.

[5]Note that on the surface this seems to contrast with the mean results shown in Fig. 2, which show a mean for the 1-band condition that is comparable to the 2-band condition with a slightly smaller error bar. However, a *post hoc* LSD test on the arcsine-transformed data indicates that the 1-band MSB speech signal is not significantly different from the no-sound control condition ($p=0.0615$). The discrepancy reflects the fact that the ANOVA evaluated the arcsine-transformed data of the individual listeners, while Fig. 2 shows the mean data pooled across all the listeners.

Arbogast, T., Mason, C., and Kidd, G. (**2002**). ''The effect of spatial separation on informational and energetic masking of speech,'' J. Acoust. Soc. Am. **112**, 2086–2098.

Boersma, P. (**1993**). ''Accurate short term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,'' Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, 97–110.

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

Brokx, J., and Nooteboom, S. (**1982**). ''Intonation and the perceptual separation of simultaneous voices,'' J. Phonetics **10**, 23–36.

Bronkhorst, A. (**2000**). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acustica **86**, 117–128.

Bronkhorst, A., and Plomp, R. (**1988**). ''The effect of head-induced interaural time and level difference on speech intelligibility in noise,'' J. Acoust. Soc. Am. **83**, 1508–1516.

Brungart, D. (**2001a**). ''Evaluation of speech intelligibility with the coordinate response measure,'' J. Acoust. Soc. Am. **109**, 2276–2279.

Brungart, D. (**2001b**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D., and Simpson, B. (**2002**). ''Within-ear and across-ear interference in a cocktail-party listening task,'' J. Acoust. Soc. Am. **112**, 2985–2995.

Brungart, D., Simpson, B., Ericson, M., and Scott, K. (**2001**). ''Informational and energetic masking effects in the perception of multiple simultaneous talkers,'' J. Acoust. Soc. Am. **110**, 2527–2538.

Cherry, E. (**1953**). ''Some experiments on the recognition of speech, with one and two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Conway, R., Cowan, N., and Bunting, M. (**2001**). ''The cocktail party phenomenon revisited: The importance of working memory capacity,'' Psychon. Bull. Rev. **8**, 331–335.

Darwin, C., and Hukin, R. (**2000**). ''Effectiveness of spatial cues, prosody, and talker characteristics in selective attention,'' J. Acoust. Soc. Am. **107**, 970–977.

Darwin, C., Brungart, D., and Simpson, B. (**2003**). ''Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers,'' J. Acoust. Soc. Am. **114**, 2913–2922.

de Cheveigne, A. (**1993**). ''Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing,'' J. Acoust. Soc. Am. **93**, 3271–3290.

Dorman, M., Loizou, P., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Drullman, R., and Bronkhorst, A. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.

Ebata, M. (**2003**). ''Spatial unmasking and attention related to the cocktail party problem,'' Acous. Sci. Technol. **24**, 208–219.

Egan, J., Carterette, E., and Thwing, E. (**1954**). ''Factors affecting multichannel listening,'' J. Acoust. Soc. Am. **26**, 774–782.

Ellis, D. (**2003**). ''Sinewave Speech Analysis/Synthesis in MATLAB,'' http://www.ee.columbia.edu/dpwe/resources/matlab/sws/

French, N., and Steinberg, J. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119.
Freyman, R., Balakrishnan, U., and Helfer, K. (**2001**). ''Spatial release from informational masking in speech recognition,'' J. Acoust. Soc. Am. **109**, 2112–2122.
Freyman, R., Helfer, K., McCall, D., and Clifton, R. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3587.
Hawley, M., Litovsky, R., and Culling, J. (**2004**). ''The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer,'' J. Acoust. Soc. Am. **115**, 833–843.
Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Kidd, G. J., Mason, C., Rohtla, T., and Deliwala, P. (**1998**). ''Release from informational masking due to the spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.
Kidd, G. J., Mason, C., Arbogast, T., Brungart, D., and Simpson, B. (**2003**). ''Informational masking caused by contralateral stimulation,'' J. Acoust. Soc. Am. **113**, 1594–1603.
Moore, T. (**1981**). ''Voice communication jamming research,'' in AGARD Conference Proceedings 331: Aural Communication in Aviation, pp. 2:1–2:6. Neuilly-Sur-Seine, France.
Moray, N. (**1959**). ''Attention in dichotic listening: Affective cues and the influence of instructions,'' Q. J. Exp. Psychol. **9**, 56–60.
Neff, D. (**1995**). ''Signal properties that reduce masking by simultaneous random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.
Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **35**, 1660–1670.
Remez, R., Rubin, P., Pisoni, D., and Carrell, T. (**1981**). ''Speech perception without traditional speech cues,'' Science **212**, 947–950.
Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Stern, R. (**1998**). ''Robust speech recognition,'' in *Survey of the State of the Art in Human Language Technology*, edited by R. Cole, J. Mariani, H. Uszkoreit, G. Varile, A. Zaemen, A. Zampolli, and V. Zue (Cambridge University Press, Cambridge).
Treisman, A. (**1960**). ''Contextual cues in selective listening,'' Q. J. Exp. Psychol. **12**, 242–248.
Tyler, R., Preece, J., and Tye-Murray, N. (**1986**). ''The Iowa audiovisual speech perception laser videodisk,'' in *Laser Videsodisc and Laboratory Report* (Department of Otolaryngology, Head and Neck Surgery, University of Iowa Hospital and Clinics, Iowa City, IA).
Wightman, F., Callahan, M., Lutfi, R., Kistler, D., and Oh, E. (**2003**). ''Children's detection of pure-tone signals: Informational masking with contralateral maskers,'' J. Acoust. Soc. Am. **113**, 3297–3305.
Wood, N., and Cowan, N. (**1995**). ''The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry,'' J. Exp. Psychol. **124**, 243–262.
Yost, W. (**1997**). ''The cocktail party problem: Forty years later,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, N.J.), pp. 329–348.
Zurek, P. M. (**1993**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. Studebaker and I. Hochberg (Allyn and Bacon, Portland).

The processing and perception of size information in speech sounds[a)]

David R. R. Smith, Roy D. Patterson, and Richard Turner
Centre for Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

Hideki Kawahara and Toshio Irino
Faculty of Systems Engineering, Wakayama University, 930 Sakaedani, Wakayama 640-8510, Japan

(Received 20 April 2004; revised 11 August 2004; accepted 14 October 2004)

There is information in speech sounds about the length of the vocal tract; specifically, as a child grows, the resonators in the vocal tract grow and the formant frequencies of the vowels decrease. It has been hypothesized that the auditory system applies a scale transform to all sounds to segregate size information from resonator shape information, and thereby enhance both size perception and speech recognition [Irino and Patterson, Speech Commun. **36**, 181–203 (2002)]. This paper describes size discrimination experiments and vowel recognition experiments designed to provide evidence for an auditory scaling mechanism. Vowels were scaled to represent people with vocal tracts much longer and shorter than normal, and with pitches much higher and lower than normal. The results of the discrimination experiments show that listeners can make fine judgments about the relative size of speakers, and they can do so for vowels scaled well beyond the normal range. Similarly, the recognition experiments show good performance for vowels in the normal range, and for vowels scaled well beyond the normal range of experience. Together, the experiments support the hypothesis that the auditory system automatically normalizes for the size information in communication sounds. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1828637]

PACS numbers: 43.66.Lj, 43.66.Ba, 43.71.Bp, 43.72.Ar [DDO] [DOS] Pages: 305–318

I. INTRODUCTION

Most animals generate their communication sounds by exciting resonant cavities in the body with a stream of sharp acoustic pulses. The resonators grow as the animal grows, and as a result, the calls of animals contain information about the size of the individual. Behavioral studies show that, when animals vocalize to attract or repel suitors, or to establish and defend territories, the size of the sender is an important part of the communication. The effect of size has been documented for many species: for example, frogs (Fairchild, 1981; Narins and Smith, 1986), dogs (Riede and Fitch, 1999), deer (Clutton-Brock and Albon, 1979), and monkeys (Fitch, 1997). Cohen (1993) has argued that size is a dimension of sound much like frequency and time, and he has developed a version of the affine Mellin transform (Titchmarsh, 1948) that can segregate the size information in a sound from the size-invariant information. The implication is that animals have evolved a physiological form of this Mellin transform which operates at a relatively early point in the auditory system, and that this is the basis of the size processing observed in animal behavior. The purpose of the current paper was to investigate the perception of size information in speech sounds to see if it is compatible with the hypothesis that the auditory system applies a size-normalizing transform to all sounds prior to the commencement of speech-specific processing.

The physiological mechanism that humans use to produce speech sounds is the same as that used by all mammals to produce their calls; the vocal cords in the larynx produce glottal pulses which excite resonances in the vocal tract beyond the larynx. As a child grows into an adult, there is an increase in vocal-tract length (VTL) (Fitch and Giedd, 1999), and as a result, the formant frequencies of the vowels decrease (Fant, 1960; Fitch and Giedd, 1999; Huber *et al.*, 1999), and this is an important form of size information in speech. A second source of size information is glottal pulse rate (GPR) measured in Hz. GPR is determined largely by the length and mass of the vocal folds (Titze, 1989), both of which increase with sex and age. In males, at puberty, there is an additional spurt in VTL and a sudden drop in GPR which complicates the interpretation of VTL and GPR information with respect to speaker size. This is an important issue but it is not the topic of this paper, and care is taken in the design of the experiments to avoid complications associated with speaker sex and the abrupt changes that occur at puberty in males.

Irino and Patterson (1997, 1999a, b, c, 2002) have developed a two-dimensional, pitch-synchronous version of the Mellin transform to simulate the processing of size information in speech. They have shown that the transform can segregate VTL information from information about vocal-tract shape (vowel type). It is assumed that the VTL information is used to evaluate speaker size, and that the normalized shape information facilitates vowel recognition. Both Turner *et al.* (2004) and Welling and Ney (2002) have demonstrated the advantage of normalization in machine recognition of vowels sounds. This paper focuses on the two complimentary as-

[a)]Portions of this work were presented in ''The perception of scale in vowel sounds,'' British Society of Audiology, Nottingham, United Kingdom, 2003, and ''The existence region of scaled vowels in pitch-VTL space,'' 18th Int. Conference on Acoustics, Kyoto, Japan, 2004.

pects of size processing: (a) the ability to discriminate a change in the size of the speaker, and (b) the ability to normalize across size differences in order to extract the size-invariant properties of vowels (i.e., vowel type).

A. Speaker size discrimination

If size functions as a dimension of sound for humans, then we might expect to find that listeners can readily make fine discriminations about speaker size just as they can for loudness or brightness. Moreover, if it is a general mechanism, we should expect to find that they make size judgments even when the vowel sounds are scaled to simulate humans much larger and smaller than those normally encountered. Kawahara, Masuda-Kasuse, and de Cheveigne (1999) have recently developed a high-quality vocoder, referred to as STRAIGHT, that uses the classical source-filter theory of speech Dudley (1939) to segregate GPR information from the spectral-envelope information associated with the shape and length of the vocal tract. Liu and Kewley-Port (2004) have reviewed STRAIGHT and commented favorably on its ability to manipulate formant-related information. STRAIGHT produces a pitch-independent spectral envelope that accurately tracks the motion of the vocal tract throughout the utterance. Subsequently, the utterance can be resynthesized with arbitrary changes in GPR and VTL; so, for example, the utterance of a man can be readily transformed to sound like a women or a child (Kawahara, 2003; Kawahara and Matsui, 2003). The utterance can also be scaled well beyond the normal range of GPR and VTL values encountered in everyday speech. We used STRAIGHT to scale vowels which could then be used to measure size discrimination, that is, the smallest change in VTL required to reliably discriminate a change in speaker size. The experiments were performed both within the normal range of the human voice and in a much larger region of the GPR-VTL space surrounding that range.

B. Vowel normalization

If the size-processing hypothesis is correct, then it is likely that vowel normalization is a natural by-product of size processing. Vowel normalization refers to the fact that humans readily recognize that the sounds produced by men, women and children saying a given vowel, such as /e/, are indeed the same vowel, despite gross differences in the waveforms. If size processing is applied to all sounds at a relatively early stage in auditory processing, and if it is the basis of vowel normalization, then we should find that vowel recognition, like size discrimination, is largely immune to the scaling of vowel sounds beyond the normal speech range. Assmann *et al.* (2002, 2003) have recently measured vowel recognition within the normal speech range and somewhat beyond in humans, using vowels scaled by STRAIGHT. They argue that the auditory system employs a form of neural net to learn which formant frequencies (spectra) go with which vowel type, and that it learns the connections for all values of GPR in the normal speech range. They interpret the deterioration in recognition performance for their more extreme stimuli as evidence of the neural net failing to generalize beyond the training set. But, they do not consider the possibility of a general-purpose size mechanism as suggested by Cohen (1993) or Irino and Patterson (1999a, b, c).

The experiments of Assmann and colleagues are limited to GPRs in the normal speech range and above it. In this paper, we report complementary experiments in which vowel recognition is measured in the normal speech range and for vowels in a three-octave range below that which is normally experienced. In the discussion, we compare the results from the two sets of studies and evaluate the relative merits of ''general-purpose normalization'' and the ''neural net'' hypothesis as explanations of the observed recognition performance.

II. METHOD

A. Stimuli

We recorded the English vowels (/a/, /e/, /i/, /o/, /u/) as spoken by author RP in natural /hVd/ sequences (i.e., *haad, hayed, heed, hoed, who'd*) using a high-quality microphone (Shure SM58-LCE) and a 44.1-kHz sampling rate. The vowels were sustained (e.g., *haaaad*) and the natural onset of the vowel was preserved while avoiding the aspiration of the preceding /h/. A cosine-squared amplitude function was used to gate the vowels on over 5 ms and off over 30 ms. The central plateau was 565 ms, so the total duration of each vowel was 600 ms. The vowels were normalized to the same rms level (0.1, relative to maximum ± 1). The pitch of the vowels was scaled to 113 Hz, which is near to the average for men. These five vowels comprise what is referred to as the ''canonical'' vowels.

The scaling of the vowels was performed using STRAIGHT (Kawahara *et al.*, 1999). It is referred to as a vocoder (voice coder), but it is actually a sophisticated speech processing package that dissects and analyzes an utterance with glottal cycle resolution. It produces a pitch-independent spectral envelope that represents the vocal-tract information independent of glottal pulse sampling. Vocal-tract shape determines which vowel type is encoded by the shape of the spectral envelope; vocal-tract length determines the scale of the pattern, and this is the form of the VTL information. Once STRAIGHT has segregated a vowel into glottal pulse rate and spectral envelope frames, the vowel can be resynthesized with the spectral-envelope dimension (frequency) expanded or contracted, and the glottal-pulse-rate dimension (time) expanded or contracted, and the operations are largely independent. Utterances recorded from a man can be transformed to sound like a women or a child; examples are provided on our web page.[1] The distinctive advantage of STRAIGHT is that the spectral envelope of the speech that carries the vocal-tract information is smoothed as it is extracted, to remove the harmonic structure associated with the original GPR, *and* the instantaneous zeros produced by the interaction of GPR and the frame rate of the analysis window. As a result, the resynthesized utterances are of extremely high quality even when the speech is resynthesized with GPR and VTLvalues well beyond the normal range of human speech (provided the GPR is not far above the first formant).

The scaling of GPR is simply a matter of expanding or

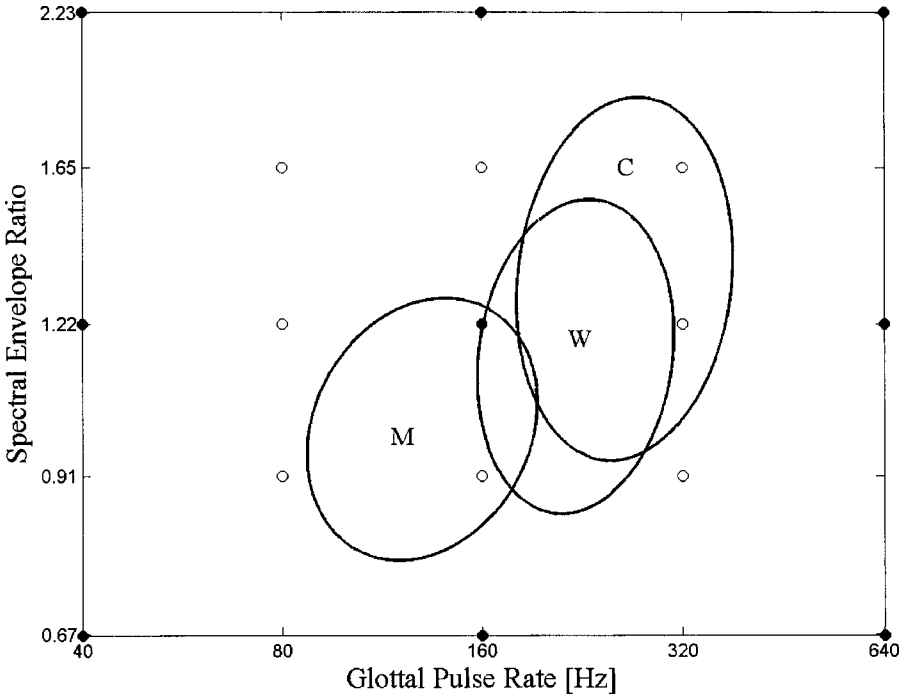


FIG. 1. Combinations of GPR and SER values used in the experiments on speaker size discrimination. Discrimination performance was measured at nine points (solid circles) in the single-vowel experiment, and at 17 points (solid and open circles) in the speech-like experiment. The three ellipses show estimates of the normal range of GPR and SER values in speech for men (M), women (W), and children (C), derived from the data of Peterson and Barney (1952). In each category 90% of individuals would be expected to fall within the respective ellipse.

contracting the time axis of the sequence of glottal events. The scaling of VTL is accomplished by compressing or expanding the spectral envelope of the speech linearly along a linear frequency axis. On a logarithmic frequency axis, the spectral envelope shifts along the axis as a unit. The change in VTL is described by the spectral envelope ratio (SER), that is, the ratio of the unit on the new frequency axis to that of the axis associated with the original recording. Note that values of SER less than unity indicate lengthening of the vocal tract to simulate larger men, and SERs greater than unity indicate shortening of the vocal tract to simulate smaller men, women, and children.

Following the scaling of GPR and VTL by STRAIGHT, the first 100 ms of the wave was removed because the abruptness of the original gate caused STRAIGHT to overshoot at onset. Then, a cosine-squared gating function (10-ms onset, 30-ms offset, 465-ms plateau) was used to select a stationary part of the vowel. The rms level was set to 0.025 (relative to maximum ± 1). The stimuli were played by a 24-bit sound card (Audigy 2, Sound Blaster) through a TDT antialiasing filter with a sharp cutoff at 10 kHz and a final attenuator. The stimuli were presented binaurally to the listener over AKG K240DF headphones. Listeners were seated in a double-walled, IAC, sound-attenuating booth. The sound level of the vowels was 66 dB SPL.

B. Procedures and listeners

1. Discrimination procedures

The just-noticeable difference (jnd) in speaker size was measured in two discrimination experiments. One measured discrimination performance using single vowels, and the other used a more speech-like sequence of four vowels. The two paradigms are referred to as the *single-vowel* and *speech-like* discrimination tasks. Both discrimination experiments employed a two interval, two-alternative, forced-choice paradigm (2AFC) with the method of constant stimuli. One interval contained the standard stimulus, the other contained the test stimulus in which the simulated VTL of the speaker was either larger or smaller. The order of the test and standard intervals in each trial was chosen randomly and the listener's task was to choose the interval in which the vowel(s) were spoken by the smaller speaker. The listeners were given written instructions explaining the task in terms of speaker size. Most listeners considered it a natural task to judge the size of speaker of the vowel sounds. One listener initially maintained that the task would be too difficult but was in fact easily able to do the task by ''thinking of the speakers as cartoon characters.'' *No feedback was given in either experiment.*

In the first discrimination experiment, the two temporal intervals of a trial each contained a single vowel of the same type (i.e., /a/ was compared with /a/). The listener was required to choose the interval in which the vowel was spoken by the smaller speaker. Psychometric functions for VTL discrimination were gathered with this *single-vowel* paradigm at nine widely spaced points on the GPR–SER plane—the combinations formed by three GPRs (40, 160, and 640 Hz) and three SERs (0.67, 1.22, and 2.23), cf. Fig. 1. At each of the nine GPR–SER points, a psychometric function was measured with six test SER (VTL) values bracketing the standard SER (VTL) value. A psychometric function was collected for each of the five vowels. The psychometric functions around one GPR–SER standard point were collected in one run (taking 30 min per listener), and consisted of ten blocks of 30 trials, where each block contained the six tests from the five vowels. Each listener thus provided 60 trials per psychometric function for each of the five vowels.

By its nature, a change in vocal-tract length produces a shift of the vowel spectrum along the frequency axis, and so it might be possible for a listener to focus on one formant peak and perform the task by noting whether the peak shifts

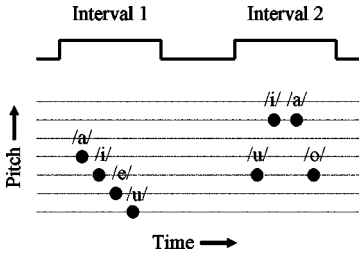


FIG. 2. Schematic paradigm for the experiment to measure speaker size discrimination with speech-like stimuli.

up or down in the second interval. It is our impression that this was not what the listeners did, and that it would be difficult, if not impossible, to do. Nevertheless, it is a logical possibility, and so, in a second discrimination experiment, we used a more *speech-like* paradigm, which effectively precluded the possibility of using a simple spectral cue. The paradigm is presented in quasimusical notation in Fig. 2. Each temporal interval of the 2AFC trial contained a sequence of four of the five vowels chosen randomly without replacement, and the vowels were presented with one of four pitch contours (rising, falling, up–across–down, down–across–up). The duration of the vowels was 400 ms (15-ms onset, 90-ms offset), and each interval gave the impression of a person carefully pronouncing a sequence of vowels at a reasonably natural rate (four vowels in 1.6 s). The pitch values were drawn from an equal-temperament, quarter-tone musical scale, in which the pitch of each note differed from its neighbors by just under 3%. The starting point for the pitch contour was varied randomly over a 9% range, and the level of the vowels in a given interval was roved in intensity over a 6-dB range. The only fixed parameter within an interval was simulated speaker VTL, and the only consistent change between intervals was speaker VTL. The listeners' task was to choose the interval in which the vowels were spoken by the smaller speaker, independent of the pattern, pitch, or loudness of the vowels. In this paradigm, the listener cannot do the task by choosing a single spectral component in one vowel and noting whether it goes up or down when the same vowel occurs in the second interval. For each combination of SER and GPR, a six-point psychometric function was measured with ten trials per point per listener. A total of 17 psychometric functions was gathered at widely spaced points on the GPR–SER plane (cf. Fig. 1).

Perceptually, this paradigm prompts the listener to think of the sounds in the two intervals as coming from two different speakers; the natural prosody of the sequences discourages listening for spectral peaks. The steps in the pitch contours were limited to quarter tones because larger steps made the sequences less speech-like. Randomizing the starting point of the contour precludes the possibility of tracking a single harmonic in one vowel from the first interval to the second. The level was not varied within interval because it made the sequences less speech-like. The fixed level difference between intervals reinforced the impression that the speakers in the two intervals were different.

For comparison, pitch discrimination was measured for five of the nine GPR–SER combinations, using the same paradigm as used to measure VTL discrimination for single vowels. In this case, the VTL of the vowels was fixed and the pitch was varied between intervals. The listener had to choose the interval containing the vowel with the higher pitch. Psychometric functions for pitch discrimination were collected at the center of the normal range (160 Hz, 1.22 SER), and the four extreme discrimination points with GPRs of 40 and 640 Hz, and SERs of 0.67 and 2.23 (cf. Fig. 1).

2. Recognition procedures

The vowel identification experiments were performed using a single-interval, five-alternative, forced-choice paradigm (5AFC). The listener heard a scaled version of one of five stationary English vowels (/a/, /e/, /i/, /o/, /u/), and had to identify the vowel spoken by selecting the appropriate button on a response box displayed on a monitor in the booth. Vowel duration was 500 ms. At the start of the first session only, to ensure that the listeners understood which button corresponded to which vowel sound, we played 100 scaled vowels from within the range of everyday experience with feedback. The particular combinations of GPR and VTL in this set were not used in the vowel identification experiments. In the main experiments, to minimize training effects, there was no feedback. The recognition data were gathered with two distinct experimental paradigms whose names refer to the combinations of GPR and SER used to construct the stimuli.

The "surface" paradigm involved a rectangular grid with all 49 combinations of 7 GPR and 7 SER values as shown in the upper panel of Fig. 3 (large open circles); it was designed to measure the surface of recognition performance over a wide range of GPR and SER values to reveal where performance begins to deteriorate. The GPR values ranged from 10 to 640 Hz in doublings; that is from more than an octave below the lower limit of periodicity pitch (32 Hz, Pressnitzer *et al.*, 2001), to about an octave above the pitch of young children. The SER values ranged from 0.5 to 3.0 in seven equal steps on a logarithmic scale; that is, the longest of the simulated vocal tracts was about 32 cm and the shortest was about 5 cm.[2] If we extrapolate from the known relationship between VTL and body height, VTLs of 32 and 5 cm correspond, respectively, to a man 14 feet tall and a newborn baby (1 foot long). Each run contained one trial of all conditions for each vowel (a total of 7 GPRs $\times$7 SERs$\times$5 vowels, or 245 trials). Each listener contributed ten runs to the surface map of vowel identification. The listeners were reminded of the five canonical vowels at the start of the run, and at 100-trial intervals thereafter. Each run took approximately 15 min to complete. No feedback was given during data collection.

The "strip" paradigm was intended to provide detailed information about the deterioration in performance along 8 spokes radiating from the center of the recognition surface; the specific combinations of GPR and SER for each strip are shown in Fig. 3 (lower panel). Each strip consisted of nine combinations of GPR and SER, making a total of 9 (sample points)$\times$5 (vowels)$\times$10 (repetitions), or 450 trials per strip. For three of the listeners in strips 1, 8, and 4 an extra (easier) point was added to ensure that these listeners' correct recognition scores approached 100%. The stimuli

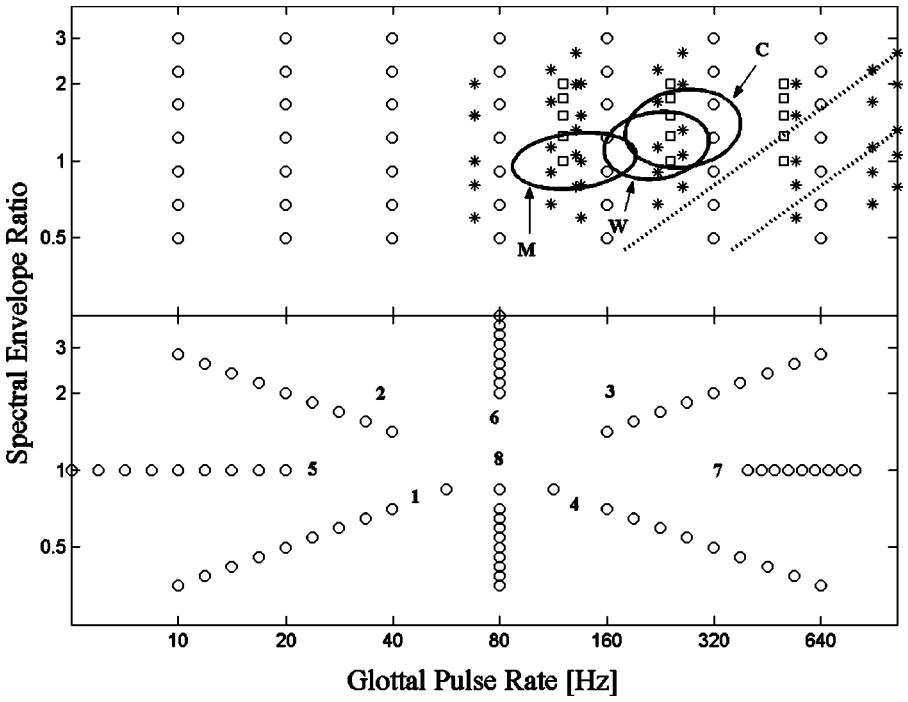


FIG. 3. Combinations of GPR and SER values used in the vowel recognition experiments. The SER determines the contraction or dilation of the spectral envelope applied by STRAIGHT during resynthesis (small SER values indicate lengthening of the VTL to simulate larger men; large SER values indicate shortening of the VTL to simulate women and children). The open circles in the top panel show the 7×7 sample points used in the *surface* recognition experiment (GPR values of 10, 20, 40, 80, 160, 320, and 640 Hz; SER values of 0.5, 0.67, 0.91, 1.22, 1.65, 2.23, and 3.0). The three ellipses show the range of GPR and SER in speech for men, women, and children (derived from Peterson and Barney, 1952). The open squares show the GPR–SER values used in Assmann *et al.* (2002) and the asterisks show the GPR-SER values used in Assmann and Nearey (2003). The upper and lower diagonal lines (dotted) show where the fundamental, $F0$, equals the first formant frequency, $F1$, or twice $F1$, for the vowel /e/, respectively. As the GPR/$F1$ ratio increases across the region between the dotted lines towards higher GPRs or smaller SERs (longer vocal tracts), the distinctiveness of the vowel deteriorates, although the stimulus still sounds vowel-like. The bottom panel shows the eight *strips* of GPR–SER combinations used in the second vowel recognition experiment.

were presented in pseudorandom order in blocks of 45 trials (9 sample points $\times5$ vowels), and the 10 replications were presented sequentially within a session. Each strip took approximately 30 min to complete. As a reminder of the mapping of vowel sound to button, the five canonical vowels were presented with feedback at the start of each run and thereafter at 100-trial intervals. The order of the eight runs required to gather the data for the eight strips was varied between listeners to balance the effects of experience and/or fatigue. No feedback was given during data collection.

3. Limit on the GPR/*F*1 ratio

There is a limitation on the GPR that can be used when producing vowels with long vocal tracts. As GPR increases, the fundamental, $F0$, eventually becomes greater than the frequency of the first format, $F1$, and as this happens, the first formant becomes very weak relative to the higher formants and the distinctiveness of the vowel deteriorates. The limitation is illustrated for the vowel /e/ by the dotted diagonal lines in the lower right-hand corner of Fig. 3 (upper panel). The upper dotted line shows the combinations of GPR and SER where $F0$ coincides with $F1$ (GPR/$F1=1$); the lower dotted line shows the combinations of GPR and SER where $F0$ coincides with $2F1$ (GPR/$F1=2$). As the GPR/$F1$ ratio increases across the region between the dotted lines towards higher GPRs or smaller SERs (longer vocal tracts), the distinctiveness of the vowel deteriorates, although the stimulus still sounds vowel-like. Thus, normalization is still applied to these sounds, but the perceptual information used to specify the vowel type gradually fades out of the representation.

For the vowels /i/ and /u/, the corresponding dotted lines would be a little higher because $F1$ is lower in these vowels; for the vowels /o/ and /a/, the corresponding dotted lines would be a little lower because $F1$ is higher in these vowels. Since the limitation varies with vowel type, and the level of $F1$ decreases continuously in the region of the GPR/$F1$ limit, discrimination of speaker size and vowel recognition would be expected to deteriorate gradually across this region rather than abruptly. Note, however, that the normal range of vowels (the ellipses) is not far above the upper dotted line, particularly for women and children, and the limitation is highly asymmetric. That is, there is no corresponding limitation to vowel production on the other side of the normal speech range. We will return to this limitation in Secs. III and IV.

4. Listeners

The recognition experiments were performed before the discrimination experiments. There were five listeners in each of the four experiments. Two listeners participated in all the experiments. Another two listeners took part in all the experiments except the speech-like discrimination task; they had left Cambridge by the time this experiment was designed. So, one new listener was recruited to participate in the single-vowel discrimination experiment only, to make a

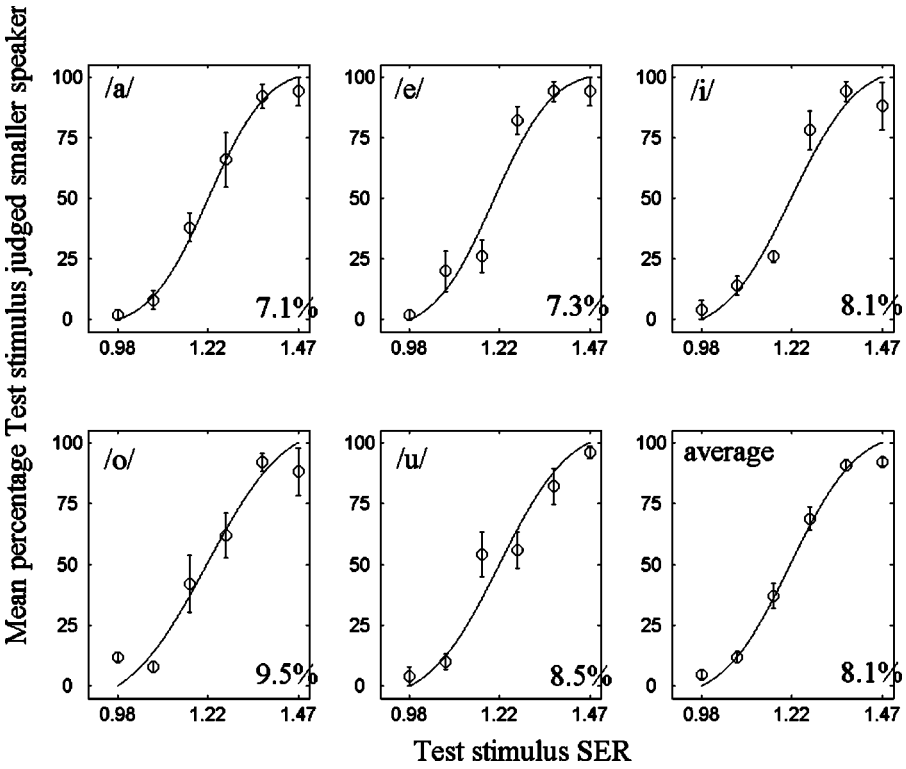


FIG. 4. Psychometric functions for speaker size discrimination in the center of the normal speech range (cf. Fig. 1, 160 Hz, 1.22 SER). Mean percentage of times the test stimulus was judged to be spoken by the smaller speaker, as a function of the SER of the test stimulus. The smooth curves are best-fitting cumulative Gaussians (Foster and Bischof, 1997). The data are shown for each vowel separately, and averaged across all five vowels (bottom-right panel). The means are based on the data of five listeners. Each point on the psychometric function for an individual vowel is based on 50 trials (10 trials from each listener). Error bars show the standard error of the mean. For the data averaged across all five vowels (bottom-right panel), each data point is based on 250 trials (50 trials from each listener). The jnd calculated from the fitted curve is shown on the bottom right of each panel.

total of five listeners for that experiment. Then, three new listeners were recruited to participate in the speech-like discrimination experiment. The listeners ranged in age from 20 to 52 years, and were paid student volunteers. All had normal absolute thresholds at 0.5, 1, 2, 4, and 8 kHz. One listener was an author (DS); all other listeners were naive to the purpose of the experiments.

III. RESULTS

The results show that detecting a change in speaker size based on a change in VTL (SER) is a relatively easy task for a wide range of VTLs. Performance remains above threshold for a range of GPR and SER values far exceeding those associated with everyday speech. Similarly, vowel recognition performance remains above threshold for a range of GPR and SER values far greater than the normal speech range. The results are presented briefly in this section and interpreted with respect to the underlying mechanisms in Sec. IV. In both the discrimination and recognition experiments, the pattern of results was similar across listeners and the levels of performance were comparable, and so, the results will be presented in terms of the average over the five listeners in all cases.

A. Discrimination of VTL and GPR

Listeners naturally hear changes in VTL as changes in speaker size. The JND for speaker size was initially measured using the *single-vowel discrimination* paradigm at nine points in the GPR-SER plane (Fig. 1). The psychometric functions for size discrimination in the center of the normal range (160 Hz, 1.22 SER) are presented separately for the five different vowels and their average (Fig. 4). The mean percentage of trials on which the test stimulus was judged to be spoken by the smaller speaker is shown as a function of the SER of the test stimulus. The group psychometric functions are similar for all vowels; they are monotonic and have relatively steep slopes, confirming that performance is similar across listeners. Cumulative Gaussian functions were fitted to the psychometric functions (Foster and Bischof, 1997). The SER value at the midpoint of the function (50%) is referred to as the point of subjective equality; it was typically within 1% of the standard used to generate the psychometric function which indicates that the perception of speaker size (SER or VTL) was unbiased. The jnd was defined as the difference in SER between the values associated with 50 and 76 percent correct ($d'=1$ in this 2AFC task) relative to the perceived SER (50 percent correct) of the standard, expressed as a percentage. The average jnd was 8.1% (with a standard deviation of $\pm 1.0\%$) in this region of the GPR–SER plane, that is, [(1.319−1.22)/1.22]*100. It was a little larger for the most difficult vowel /o/ (9.5%) and a little smaller for the easiest vowels /a/ (7.1%) and /e/ (7.3%). For comparison, the jnd's for noise level (loudness), light level (brightness), and chemical density (odor) are about 10%, 15%, and 25%, respectively (Miller, 1947; Cornsweet and Pinsker, 1956; Gescheider, 1976).

The jnd for speaker size is approximately the same for the six conditions where the GPR is either 40 or 160 Hz (two left columns of the design, cf. Fig. 1), being on average about 9%. When the pitch is 640 Hz, the jnd doubles to about 18% when the SER is greater than 1. For the condition with a high pitch and a long VTL (Fig. 1, bottom-right point), the task essentially breaks down. The jnd was measurable for three of the listeners but it rose to 50%, and two of the listeners could not do the task. The pattern of size discrimination performance is related to recognition performance (cf. subsection C). Briefly, discrimination performance is good

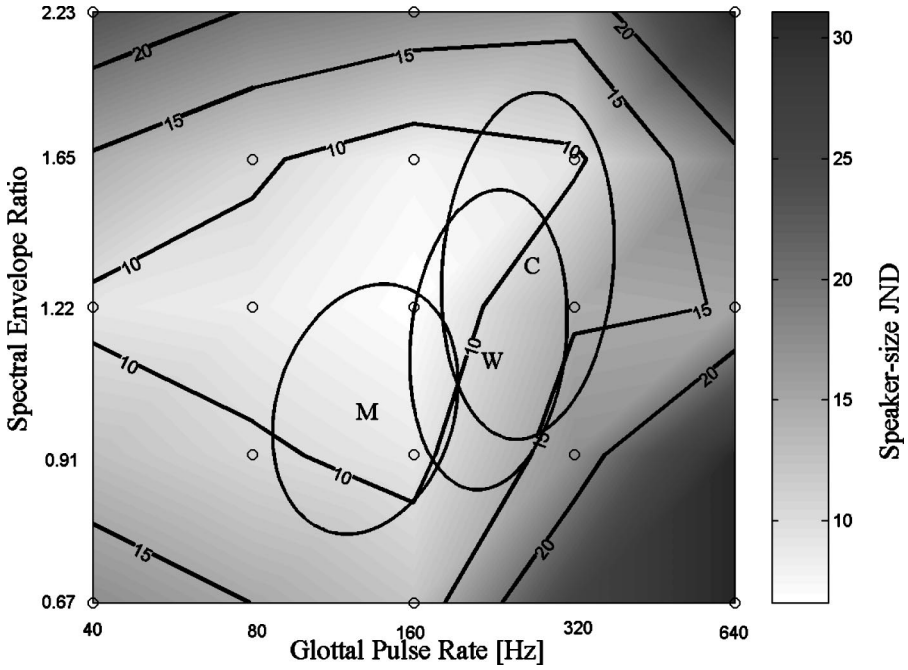


FIG. 5. jnd contours (expressed as a percentage of the SER) for speaker size in the speech-like experiment. The jnd's are presented as a 2D surface plot with gray tone showing discrimination performance. The jnd was measured at the points shown by the circles, and the surface was interpolated between the data points. Each jnd is based on a psychometric function fitted to 300 trials (60 from each of 5 listeners). The thick black lines show the contours for jnd's of 10%, 15%, and 20%. The three ellipses show the range of GPR and SER in speech for men, women, and children (derived from data of Peterson and Barney, 1952).

where recognition performance is good, and vice versa. Notably, discrimination does not deteriorate as soon as the GPR and VTL values exceed the normal speech range (shown by the ellipse in Fig. 1).

For comparison, the jnd for *pitch discrimination* was measured for these same vowels at five of the nine points where size discrimination was measured—the central point and the four corners (cf. Fig. 1). The jnd for GPR discrimination was less than 2% when the GPR was greater than 100 Hz, independent of VTL. The jnd rises to about 9% when the GPR is 40 Hz, both for long and short vocal tracts. In this region, the GPR is approaching the lower limit of melodic pitch, which is about 32 Hz (Pressnitzer *et al.*, 2001), and the jnd for discrimination of a change in the rate of clicks in a click train rises to values of about 7% (for a recent review, see Krumbholtz *et al.*, 2000). The jnd for the frequency of a sinusoid also rises as frequency decreases below 250 Hz (e.g., Sek and Moore, 1995). Thus, GPR and size discrimination show different patterns of variation across the GPR–SER plane. GPR discrimination deteriorates in a region where size discrimination remains good (i.e., the region of low pitch), and GPR discrimination remains good in the region where size discrimination deteriorates (i.e., for combinations of high pitch and long vocal tract).

In the *single-vowel discrimination* experiment, the listeners reported using speaker size as the cue for the discrimination rather than the pitch of a sinusoid. Nevertheless, it would be possible to do the task using a simple spectral cue if one could reliably identify the spectral peak associated with one of the formants in the first interval and check to see which way it shifted in the second. It is our opinion that this is not possible when the vowel and pitch change on every trial. The *speech-like discrimination* task effectively precludes the possibility of making the size discrimination on the basis of a simple spectral cue. The jnd for speaker size was measured with the speech-like paradigm with four vowels in each interval and varying pitch contours in the two intervals (Fig. 2). The size jnd was measured at the center point of the normal range and for two concentric squares of 8 points centered on the first point; the outer square is composed of the 8 points in the single-vowel experiment; the inner square has a positive diagonal that just encompasses the normal range of GPR–SER values in speech. The circles in Fig. 1 show the specific combinations.

The jnd for speaker size was measured at 17 points in the GPR–SER plane, which is sufficient to make a contour map of resolution. Figure 5 shows the speaker-size jnd (SER jnd in percent) as a function of GPR and SER with logarithmic axes using a 2D surface plot in which gray tone shows resolution. Small jnd's (better resolution) are plotted in grays approaching white, and large jnd's (worse performance) are plotted in grays approaching black. The actual sample points are shown as circles; the contours are derived by interpolation between the data points. The three ellipses show estimates of the normal range of GPR and SER in speech for men, women, and children (Peterson and Barney, 1952). In each case, the ellipse encompasses 90% of the individuals in the Peterson and Barney data for that category of speaker (man, woman, or child).

Figure 5 and Table I show that discrimination performance is excellent, with jnd's less than 10%, in a triangular region of the GPR–SER space that includes about half of the normal region for women and children and most of the normal region for men. The 15% and 20% contours shows that discrimination performance remains high for SERs well beyond the normal range, provided the GPR is below about 200 Hz. Above 200 Hz, the 15% and 20% contours are well outside the normal range for short vocal tracts, but the 15% contour encroaches on the normal range for women and children when the vocal tract is relatively long. We were unable to measure the jnd in the bottom, right-hand condition (640 Hz, 0.67 SER). In this case, the $F0$ of the GPR is above the first formant, causing the vowel quality to deteriorate. To anchor the contour map in the bottom-right corner, we used

TABLE I. GPR and SER values of the 17 points in the speech-like discrimination experiment where speaker-size jnd's were measured. Each cell contains the speaker-size jnd and (in brackets) the vowel recognition score for that particular combination of GPR and SER as measured in the surface vowel recognition experiment (cf. percent-correct values in Fig. 6, lower right-hand panel).

	GPR				
SER	40	80	160	320	640
2.23	23.4 (70)		17.6 (73.2)		31.1 (72.8)
1.65		10.6 (93.2)	7.6 (96)	9.3 (97.6)	
1.22	9.1 (96)	8.3 (98.4)	6.6 (99.6)	14.5 (94)	15.1 (69.6)
0.91		10.5 (98.4)	8.9 (91.2)	17.2 (82.8)	
0.67	17.5 (83.2)		12.3 (86.4)		52[a] (42.8)

[a]jnd value from single-vowel size discrimination experiment.

the jnd from the single-vowel discrimination experiment. Typically, jnd values obtained with the single-vowel paradigm were slightly better than in the speech-like paradigm, so performance in this corner may be even worse than that shown.

B. Vowel recognition

The vowel recognition data obtained with the *surface paradigm* are presented separately for the individual vowels in Fig. 6, along with the average for the five vowels (lower right-hand panel). The abscissa is GPR and the ordinate is SER plotted on logarithmic axes; the percent correct is given by the tone of gray. The points where performance was measured are shown by the circles. The gray-scale tone and contours were created by interpolation between the data points. The heavy black line shows threshold, that is, the 50-percent-correct identification contour where d' is 1.0 in this 5AFC paradigm. The figure shows that performance was surprisingly good and only drops below threshold for the more extreme values of GPR and SER. To reveal the regions where performance drops below ceiling levels more clearly, the data are also presented as 3D surfaces in Fig. 7 plotted above a plane showing the sample points. The bold lines show the threshold contours as in Fig. 6.

With regard to VTL, the worst performance is associated with the vowels /o/ and /u/, where both the upper and lower threshold contours fall just within the range of measured values. For /i/, the upper bound falls within the measured range; for /a/, the lower bound falls within the measured range. For /e/, performance only drops below threshold when low SER values occur in combination with either low or high GPR values. With regard to GPR, recognition performance remains near ceiling levels as GPR decreases below the range of voice pitch (~64 Hz), to the limit of melodic pitch (~32 Hz) and beyond. At 10 Hz, although there is no pitch sensation and one hears a stream of individual glottal cycles, the vowel quality is readily perceived. As GPR increases above the normal speech range to 640 Hz, performance remains near ceiling levels for /a/, /i/, and /u/. Performance drops to threshold irregularly along the upper GPR boundary for /e/ and drops reliably below threshold for one vowel, /o/. The average data present a reasonable summary of recognition performance for the five vowels (lower right-hand

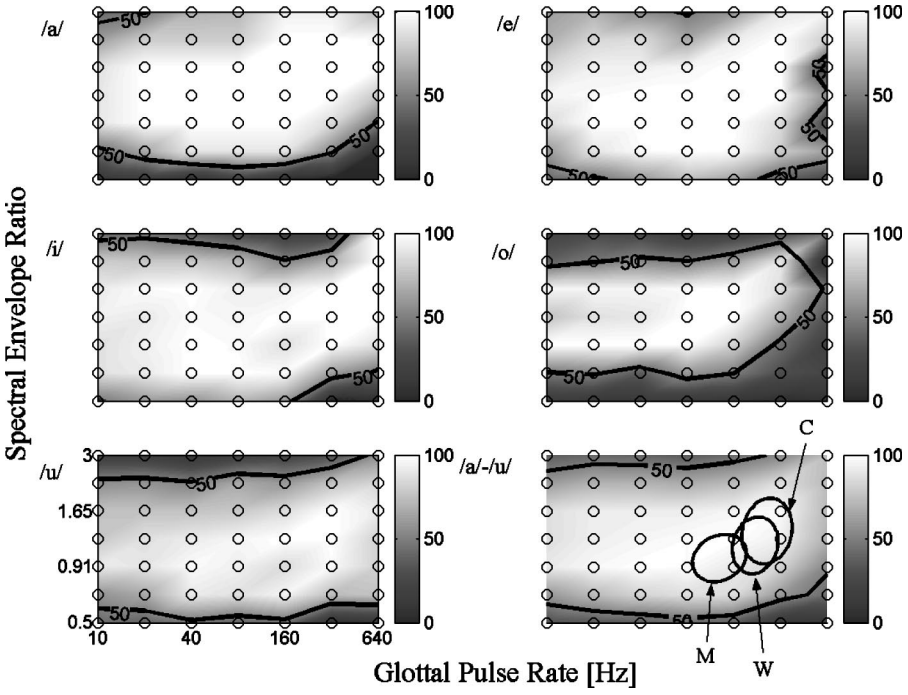


FIG. 6. Vowel recognition performance using the *surface* paradigm. The data are presented as a 2D surface plot with gray tone showing mean percent correct. Sample points are shown as circles with interpolation between data points. The means represent 10 trials from each of 5 listeners. The data averaged across all five vowels are shown in the bottom-right panel (250 trials/point). The thick black contour marks recognition threshold (50%, $d'=1.0$) in our 5AFC experiment. The three ellipses (bottom-right panel) show the range of GPR and SER in speech for men, women, and children (derived from Peterson and Barney, 1952).

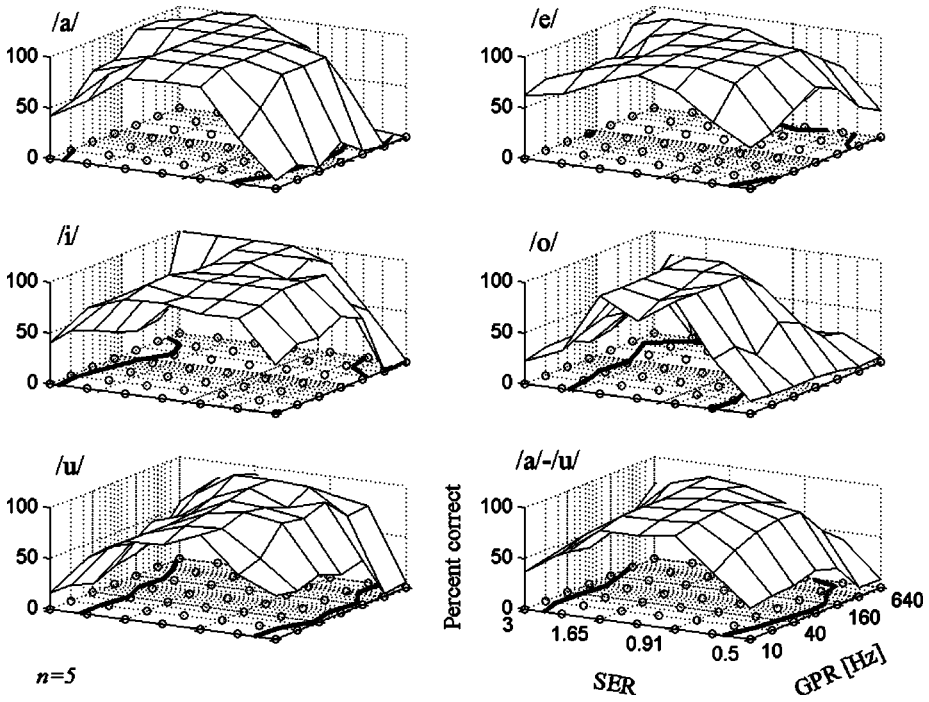


FIG. 7. Vowel recognition performance using the *surface* paradigm. The data are presented as a 3D wire-mesh surface (no interpolation), with height showing mean percent correct. The GPR and SER combinations used in the experiment are shown by the circles on the 2D projection plane lying below the 3D surface. Recognition threshold (50%, $d' = 1.0$ in 5AFC) is shown by the thick black contour on the 2D plane. For other details cf. Fig. 6.

panel). Performance drops to threshold when the SER value decreases to ~0.6 or when it increases to ~2.8, and this is largely independent of GPR. Between the threshold SER contours, performance is similar for GPR values throughout the range from 10 to 640 Hz.

The recognition data obtained with the *strip paradigm* are presented separately for the eight strips in the panels around the circumference of Fig. 8; the center panel shows the GPR–SER values for each strip. The data are averaged over vowel type. The ordinate is mean percent correct in all of the data panels. The abscissa is GPR for the panels in the left-hand and right-hand columns of the figure; the abscissa is SER in the two data panels in the central column of the figure. Threshold for these psychometric functions is 50 percent correct, where d' is 1.0 in this 5AFC paradigm. The data panels in the *central row* of the figure show that for a central SER value, performance stays above threshold throughout the full range of GPR values from 5 to 640 Hz; indeed, at the lower GPR values, performance is essentially perfect even in the region below the lower limit of pitch. The data panels in the *central column* of the figure show that for a pitch of 80 Hz (very low male), performance stays above threshold down to an SER value of 0.55 and up to an SER value of 2.8. If the recognition surface were elliptical in shape, reflecting the shapes of the normal ranges for men women and children, then the psychometric functions across

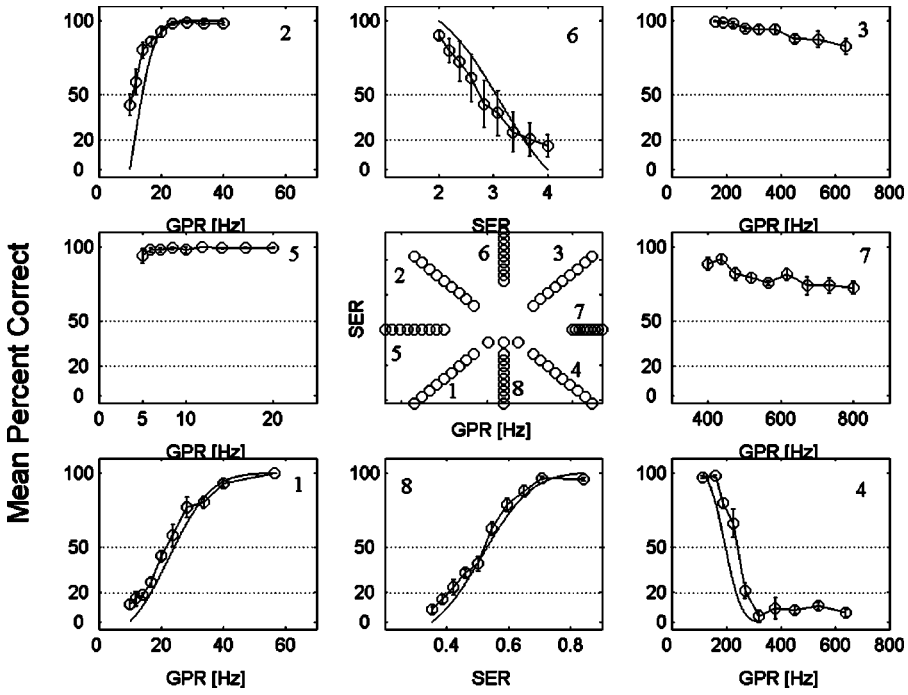


FIG. 8. Vowel recognition performance using the *strip* paradigm. Data collapsed across all five vowels and all five listeners. Each data point is based on 250 trials. Smooth curves are best-fitting cumulative Gaussians and have been used where appropriate. The center panel shows the GPR–SER values for all eight strips. For other details cf. Fig. 3.

the top and bottom rows would have the same form and drop to threshold in approximately the same region of the figure. By and large, they do not. In the top row, performance remains well above threshold for the high GPRs in the right-hand panel, and it only just drops to threshold for the low GPRs in the left-hand panel. For short VTLs, then, the shape of the recognition surface is more rectangular than elliptical. In the bottom row, left-hand panel, the psychometric function falls below threshold in about the same region as the psychometric function in the central panel, indicating that GPR and SER interact when both values are small to produce a larger reduction in performance than either would on its own; this means that the corner of the recognition surface is rounded in this case. In the bottom right-hand panel, the psychometric function falls below threshold even sooner than the psychometric function in the central panel, indicating that GPR and SER interact more strongly here and produce a much larger reduction in performance than either would on its own; this means that this corner of the recognition surface is more rounded than the surface in Fig. 6 might initially indicate. This is because it is difficult to produce vowels with a well-defined first formant when the GPR is high.

C. Speaker-size discrimination and vowel recognition performance

Table I shows that speaker size discrimination and vowel recognition performance are related; when discrimination performance is good, vowel recognition performance is good. As we move away from the normal speech range (cf. the ellipses in Figs. 5 and 6) performance starts to drop off in a similar way for both perceptual tasks. The Pearson product-moment coefficient of correlation, r, between these two performance measures is -0.91 and it is highly significant ($p \ll 0.001$, one-tailed, $N=17$). The negative correlation is because high vowel recognition scores go with low speaker-size jnd's.

IV. DISCUSSION

The discrimination experiments show that listeners can make fine judgments about the relative size of two speakers, and that they can make size judgments for vowels scaled well beyond the normal range in both VTL and GPR (Fig. 5). The jnd for SER is less than 10% over a wide area of the GPR–SER plane, and when the GPR is 160 Hz, there are approximately 10 jnd's in speaker size between the bounding SER values of 0.67 and 2.23. The recognition experiments show that listeners can identify vowels manipulated to simulate speakers with GPRs and VTLs well beyond the normal speech range (Fig. 6). Recognition performance was above threshold for an area approximately ten times greater than the normal speech range.

A. Speaker size discrimination

The most relevant data on size perception come from some simple studies performed by Lass and Davis (1976) and Fitch (1994). Lass and Davis (1976) asked listeners to judge the height of 30 men and women reading a standard prose passage on a four-category scale. Categorization performance was better than chance. However, no attempt was made to control for the average pitch difference between men and woman, and the range of heights was limited. Fitch (1994) used computerized vowel sounds. He made the assumption that formant frequencies are a linear function of VTL and scaled the formant values for an 18-cm vocal tract to produce proportional values for vocal tracts of 17, 16, and 15 cm. For each vocal tract, he synthesized the ''schwa'' vowel at two GPRs, 100 and 150 Hz. The vowels were presented one at a time to a group of listeners who rated the size of the speaker on a 7-point scale. Despite the simplicity of the experiment and the limited range of VTL values, the data show significant main effects of both GPR and VTL on the size ratings for this schwa vowel. The dissertation does not, however, measure size discrimination or vowel recognition, and the vowels are limited to the normal range for men.

The acoustic basis for size discrimination is clear; formant frequencies decrease as VTL increases. Research on speech production indicates that, over the full range of size from children to adults, the relationship between formant frequency and VTL is almost linear (Fant, 1960). Measurements with magnetic resonance imaging (Fitch and Giedd, 1999) show that VTL is highly correlated with height (Fitch and Giedd, 1999). There is also a highly significant correlation between formant frequency and age (Huber *et al.*, 1999). Recently, González (2004) has reported that there is even a weak relationship between formant frequency and size within a group of adult men and within a group of adult women. Turner and Patterson (2003) have recently used quantitative clustering to reanalyze the classic data of Peterson and Barney (1952) and show that within a given vowel cluster, speaker size is the largest source of variation. Finally, it is perhaps worth noting that there is a strong relationship between formant related parameters and body size in rhesus monkeys (Fitch, 1997).

In retrospect, given the importance of body size in human interaction, and the strong correlation between height and vocal tract length (Fitch and Giedd, 1999), it seems odd that the perception of speaker size has received so little attention in hearing and speech research. In spectral terms, the effect of a change in speaker size is theoretically very simple; if the GPR is fixed and the frequency axis is logarithmic, the profile for a given vowel has a fixed shape and VTL changes simply shift the profile as a unit—towards the origin as size increases and away from it as size decreases. The analysis of spectral profiles by the auditory system has been a very popular topic in psychoacoustics since it was introduced by Spiegel, Picardi, and Green (1981). However, in the main, people have elected to follow Spiegel *et al.* and concentrate on profiles constructed from sets of equal-amplitude sinusoids whose frequencies are equally spaced on a *logarithmic* axis. These stimuli are not like the voiced parts of speech; they do not have a regular harmonic structure, the excitation is not pulsive, and they sound nothing like vowels. Moreover, the task in traditional profile analysis (PA) is to detect an increment in one of the sinusoidal components, which is very different from detection of a shift in the spectral location of the profile as a whole.

An excellent overview of PA research is presented in

Drennan (1998); he describes a few PA experiments in which the stimuli are composed of sets of harmonically related components that are intended to simulate vowel sounds to a greater or lesser degree. Leek, Dorman, and Summerfield (1987) generated four ''flat-spectrum vowels'' starting with a set of equal-amplitude harmonics spanning much of the speech range, and incremented pairs of components at the frequencies of the formant peaks. They measured the size of the increment required to recognize the vowel and found it to be consistent with the results of traditional profile studies as reported in Green (1988). Alcántara and Moore (1995) generated six flat-spectrum vowels with the components in cosine phase, as they are in normal vowels, or with the components in random phase. As might be expected, the increment required at the formant frequencies to detect the vowel was consistently smaller in the cosine-phase condition than in the random-phase condition. However, in these and other PA studies, there is no attempt to simulate the filtering action of the vocal tract and produce realistic vowel profiles; nor is there any attempt to simulate changes in VTL or measure sensitivity to coherent spectral shifts.

B. Vowel recognition

Assmann *et al.* (2002) have reported a recognition study similar to those presented in this paper in which the vowels of three men were scaled in GPR and SER using STRAIGHT. The SER was scaled up in five equal steps from 1.0 to 2.0. The GPR was scaled up in octaves from 1 to 2 and 4. The combinations of GPR and SER are presented by open squares in the upper panel of Fig. 3; in positioning the squares, it has been assumed that the average GPR and VTL for the 11 vowels of the three men is near the average GPR and VTL for men in the classic data of Peterson and Barney (1952). So, the bottom square in the left column of squares is near the center of the ellipse[3] for men from the Peterson and Barney (1952) data. The recognition performance for the rectangle of the GPR–SER plane used in Assmann *et al.*'s experiment has the same general form as shown for our average data in the lower right-hand panel of Fig. 6 (labeled ''/a/–/u/''). That is, when the GPR scalar is 1 or 2 and the SER is between 1.0 and 1.5, performance is at ceiling levels; thereafter, as the SER increases to 2.0, performance falls gradually, but it remains well above threshold for both GPR scalars (1 and 2). When the GPR scalar is 4, performance is at ceiling levels for the *larger* SER scalars (1.5–2.0) and it decreases as the SER scalar decreases to 1.0. That is, performance decreases as the stimuli encroach on the region where the definition of the first formant deteriorates (cf. the dotted diagonal lines in Fig. 3, upper panel). The percent-correct values are lower in Assmann *et al.* (2002) than in our study because they used 11 vowel types rather than 5. Nevertheless, in the worst case (GPR scalar=4; SER=1) the d' was 1.24, which is still above the threshold value (1.0) in Fig. 6 (bold solid line).

Assmann *et al.* (2002) interpret the reduction in performance outside the range of GPR–SER combinations normally encountered for men women and children as evidence that the brain learns the combinations of pitch and formant frequencies associated with the different vowels for the normal range of men, women, and children, in much the same way as a neural net would. The combinations of GPR and formant frequency in their experiment go beyond the normal range but not very far, and so in their next study (Assmann and Nearey, 2003) they extended the range, taking the lowest GPR scalar down from 1.0 to 0.5 and the lowest SER down from 1.0 to 0.6. They used the vowels of two men, and extended the design with the vowels from two women and two children (aged 7). Reducing the GPR scalar to 0.5 has essentially no effect on performance relative to that achieved with a GPR scalar of 1.0; this is true for all values of SER and for all three classes of speaker. Similarly, the effect of reducing the SER is small when the GPR scalar is 0.5 or 1.0; there is a reduction in performance for the vowels of the men and women, but it remains well above threshold. The most striking effect is a three-way interaction between GPR scalar, SER, and speaker group. Briefly, when the GPR scalar is 0.5 or 1.0, the reduction in performance observed with the vowels of men as SER rises to 2.0 is accentuated with the vowels of women and children, and when the GPR scalar is increased to 4, the effects of SER and speaker class are amplified and performance drops to chance.

Much of the complexity, however, appears to be the result of using relative measures for GPR and VTL when plotting the data, and ignoring the fact that the base values of GPR and SER are changing substantially across speaker group in the statistical analysis. The asterisks in Fig. 3 show the combinations of GPR and SER for the data of Assmann and Nearey (2003) when we adjust for the fact that the base GPR (1.0) represents a higher pitch for women and children, and the base SER (1.0) represents a shorter vocal tract for women and children. Specifically, we assume that the origin for each speaker group (GPR scalar=1; SER=1) is at the center of the Peterson and Barney ellipse for that group. This shows that when the GPR scalar is 4, the vowels for women and children (right-most pair of asterisk columns) are in the region where the definition of the first formant is deteriorating (cf. the diagonal lines in Fig. 3), and the vowels for men (column of asterisks at GPR $\sim$600 Hz) are encroaching on this region. For the remainder of their conditions, performance is well above threshold, except for the largest SER (2.0) for women and children, and in this region performance is deteriorating in our data as well. In summary, the pattern of recognition performance in the region where the data overlap appears to be comparable in all three of these recognition experiments.

C. Vowel normalization by scale transform and/or statistical learning

There are several aspects of the recognition data which suggest that performance is not primarily determined by learning the statistics of the correspondence between GPR and formant frequencies in natural speech with a neural net, as suggested by Assmann *et al.* (2002). Neural nets have no natural mechanism for extrapolating beyond the training data (LeCun and Bengio, 1995; Wolpert, 1996a, b), so we would expect some deterioration in recognition performance as soon as either the GPR or SER move beyond the normal range. Assmann and colleagues do not provide a clear speci-

fication of the normal range, but it would seem reasonable to assume that theirs would be similar to the one we derived from the Peterson and Barney (1952) data. A comparison of the data from all three recognition experiments with the ellipses of normal GPR and SER values (cf. Figs 3 and 6) shows that recognition performance is near ceiling levels across a region of GPRs and SERs that extends well beyond the normal range. This includes many physiologically implausible combinations that most people would have little or no experience with. Most notably, performance does not drop as GPR decreases down out of the normal range for men, women, or children. It remains at ceiling levels down to the lower limit of voice pitch.

Much of the drop in recognition performance in the studies of Assmann and colleagues occurs, as in our study, in the region where it is not possible to generate vowels with a good definition of the first formant (bottom right-hand corners of the panels in Fig. 6). The formant is only represented by one harmonic (the fundamental) on the upper side of the formant. It seems likely that this plays at least as large a role in the reduction of performance as lack of experience of vowels in this region. While we do not wish to deny a role for experience and training in improving performance in vowel recognition, it is hard to see it explaining the large range over which listeners are able to recognize vowels at near-ceiling levels, particularly when they are given no feedback.

Assmann and colleagues do not consider the possibility that the auditory system applies a scale transform to the internal representation of sound prior to the recognition process as suggested by Cohen (1993) and Irino and Patterson (1999b, 2002), and that the normalization inherent in the scaling transform is the reason why humans can recognize vowels with GPRs and VTLs far beyond the normal speech range. Our data showing that size discrimination and vowel recognition are both possible over a region approximately ten times greater than the normal speech range support the hypothesis that the auditory system applies some form of scaling transform (such as the Mellin transform) to all input sounds prior to speech-specific processing.

There are two complementary advantages provided by scale transforms which segregate the size information associated with vocal-tract length from the shape information associated with vowel type: on the one hand, the normalization renders vowel recognition immune to size distortion and facilitates the problem of dealing with speakers of very different sizes; on the other hand, it concentrates the size information in the representation and facilitates decisions such as whether the speaker is a man, woman, or child. Recently, the auditory image model (AIM) of Patterson *et al.* (1992, 1995, 2000) has been extended to include a stage that normalizes the auditory images produced by AIM and converts them to Mellin images which are scale invariant. The system was tested with the aid of a simple vowel classifier and a large range of scaled vowels like those in the current experiments (Turner *et al.*, 2004). The tests showed that the recognition of scaled vowels is enhanced by the addition of the Mellin image stage, and the range of suprathreshold performance is compatible with our vowel recognition data. This adds further support to the hypothesis that the auditory system has a scaling mechanism, and that it plays an important role in vowel normalization.

D. Relative versus absolute size

It is important to distinguish between judging *relative* size and *absolute* size. Our discrimination task only requires a relative size judgment; moreover, the two sounds are presented in a paradigm designed to favor the immediate comparison of two internal representations of sound and minimize the memory load. It is like judging which of two weights is heavier by lifting one and then the other; you do not need to know what the absolute weights are, simply that the second feels lighter or heavier when you pick it up. Judgments about absolute size are probably much harder to make, and it seems likely that you need to know something about the source to judge its absolute size. The general problem of the relationship between the perception of relative and absolute size is beyond the scope of this paper.

V. SUMMARY AND CONCLUSIONS

The Mellin transform has been used to develop a signal-processing model of vowel normalization (Cohen, 1993) and an auditory model of vowel normalization (Irino and Patterson, 2002). The implication is that size is a dimension of sound, and that the size information can be segregated automatically from the shape information. The current paper presents psychophysical experiments which suggest that size is a dimension of auditory perception as well as a dimension of sound itself, and that vowel normalization is based on a scale transform. Glottal pulse rate and vocal-tract length were manipulated independently over a large range using a high-quality vocoder (STRAIGHT). Human listeners were able to make discriminations about speaker size, and to recognize scaled vowels, over a range of GPRs and SERs ten times greater than that encountered in normal speech (Figs. 5 and 6). The results support the hypothesis that the auditory system includes some form of scale transform that automatically segregates size and shape information in the sound.

ACKNOWLEDGMENTS

This research was supported in part by the following institutions: (authors DS, RP, RT): the UK Medical Research Council (G9901257) and the US EOARD (Contract No. SPC-024083); (author HK): the National Institute of Information and Communications Technology of Japan; (author TI): faculty of systems engineering, Wakayama University, and the Japanese Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (B), 15300061, 2003. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of EOARD. Pilot data from this research program were reported in Turner *et al.* (2004). Some of the work has been reported in abstract form (Smith, Patterson, and Jefferis, 2003; Smith and Patterson, 2004).

[1]http://www.mrc-cbu.cam.ac.uk/cnbh/web2002/framesets/Soundsframeset.htm. Click on ''Scaled vowels.''

[2]An estimate of the size of speaker for a given SER can be derived by extrapolating from the VTL versus height data in Fitch and Giedd (1999). The average VTL for 7 men aged 19 to 25 in Fitch and Giedd (1999) was 16 cm. An SER of 0.5 means that the spectrum envelope of the initial input vowel has been compressed by a factor of 2, while an SER of 3.0 means that the spectrum envelope has been dilated by a factor of 3. Assuming linear scaling between formant position and VTL, our SERs are equivalent to VTLs of 32 cm (giants) and 5.3 cm (tiny children). Given the correlation between VTL and height [Fitch and Giedd, 1999; cf. Fig. 2(a)], our smallest SER of 0.5 would simulate the sound of a speaker 4.3 m (14 feet) tall, and our largest SER of 3.0 would simulate a baby just 0.35 m (1 foot 2 inches) tall.

[3]The open square symbols for Assmann *et al.* (2002) have been slightly displaced to the left by 15, 30, and 40 Hz for the first, second, and third columns, respectively, to distinguish them clearly from some of the asterisks symbols used for Assmann and Nearey (2003).

Alcántara, J. I., and Moore, B. C. J. (**1995**). ''The identification of vowel-like harmonic complexes: Effect of component phase, level, and fundamental frequency,'' J. Acoust. Soc. Am. **97**, 3813–3824.

Assmann, P. F., Nearey, T. M., and Scott, J. M. (**2002**). ''Modeling the perception of frequency-shifted vowels,'' in ''Proceedings of the 7th Int. Conference on Spoken Language Perception,'' ICSLP, 425–428.

Assmann, P. F., and Nearey, T. M. (**2003**). ''Frequency shifts and vowel identification,'' in ''Proceedings of the 15th Int. Congress of Phonetic Sciences,'' Barcelona ICPhS.

Clutton-Brock, T. H., and Albon, S. D. (**1979**). ''The roaring of red deer and the evolution of honest advertising,'' Behaviour **69**, 145–170.

Cohen, L. (**1993**). ''The scale transform,'' IEEE Trans. Acoust., Speech, Signal Process. **41**, 3275–3292.

Cornsweet, T. N., and Pinsker, H. M. (**1965**). ''Luminance discrimination of brief flashes under various conditions of adaptation,'' J. Physiol. (London) **176**, 294–310.

Drennan, W. (**1998**). ''Sources of variation in profile analysis: Individual differences, extended training, roving level, component spacing, and dynamic contour,'' Ph.D. dissertation, Indiana University.

Dudley, H. (**1939**). ''Remaking speech,'' J. Acoust. Soc. Am. **11**, 169–177.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fairchild, L. (**1981**). ''Mate selection and behavioural thermoregulation in Fowler's toads,'' Science **212**, 950–951.

Fitch, W. T. (**1994**). ''Vocal tract length perception and the evolution of language,'' Ph.D. dissertation, Brown University.

Fitch, W. T. (**1997**). ''Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques,'' J. Acoust. Soc. Am. **102**, 1213–1222.

Fitch, W. T., and Giedd, J. (**1999**). ''Morphology and development of the human vocal tract: A study using magnetic resonance imaging,'' J. Acoust. Soc. Am. **106**, 1511–1522.

Foster, D. H., and Bischof, W. F. (**1997**). ''Bootstrap estimates of the statistical accuracy of thresholds obtained from psychometric functions,'' Spatial Vis. **11**, 135–139.

Gescheider, G. A. (**1976**). *Psychophysics; Method and Theory* (Erlbaum, Hillsdale, NJ).

González, J. (**2004**). ''Formant frequencies and body size of speaker: A weak relationship in adult humans,'' J. Phonetics **32**, 277–287.

Green, D. M. (**1988**). *Profile Analysis* (Oxford University Press, London).

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). ''Formants of children, women, and men: The effects of vocal intensity variation,'' J. Acoust. Soc. Am. **106**, 1532–1542.

Irino, T., and Patterson, R. D. (**1997**). ''A time-domain. level-dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.

Irino, T., and Patterson, R. D. (**1999a**). ''Extracting size and shape information of sound sources in an optimal auditory processing model,'' CASA Workshop, IJCAI-99, Stockholm, 1–4 Aug., 1999.

Irino, T., and Patterson, R. D. (**1999b**). ''Stabilised wavelet Mellin transform: An auditory strategy for normalising sound-source size,'' Eurospeech 99, Budapest, Hungary, Sept., 1899–1902.

Irino, T., and Patterson, R. D. (**1999c**). ''An auditory strategy for separating size and shape information of sound sources,'' Japan Soc. Artificial Intell., Tech. Rep., SIG-Challenge-9907-6, 33–38.

Irino, T., and Patterson, R. D. (**2002**). ''Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilized wavelet-Mellin transform,'' Speech Commun. **36**, 181–203.

Kawahara, H., Masuda-Kasuse, I., and de Cheveigne, A. (**1999**). ''Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based $F0$ extraction: Possible role of repetitive structure in sounds,'' Speech Commun. **27**(3–4), 187–207.

Kawahara, H., and Matsui, H. (**2003**). ''Auditory morphing based on an elastic perceptual distance metric in an interference-free, time-frequency representation,'' in Proceedings IEEE Int. Conference on Acoustics, Speech & Signal Processing (ICASSP '03) **1**, 256–259.

Kawahara, H. (**2003**). ''Exemplar-based voice quality analysis and control using a high quality auditory morphing procedure based on STRAIGHT,'' VOQUAL'03, ESCA Tutorial and Research Workshop, Geneva, 27–29 August, 2003, 109–114.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). ''The lower limit of pitch as determined by rate discrimination,'' J. Acoust. Soc. Am. **108**, 1170–1180.

Lass, N. J., and Davis, M. (**1976**). ''An investigation of speaker height and weight identification,'' J. Acoust. Soc. Am. **60**, 700–703.

LeCun, Y., and Bengio, Y. (**1995**). ''Convolutional networks for images, speech, and time-series,'' in *The Handbook of Brain Theory and Neural Networks*, edited by M. A. Arbib (MIT Press, Cambridge, MA).

Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). ''Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 148–154.

Liu, C., and Kewley-Port, D. (**2004**). ''STRAIGHT: a new speech synthesizer for vowel formant discrimination,'' ARLO **5**, 31–36.

Miller, G. A. (**1947**). ''Sensitivity to changes in the intensity of white noise and its relation to masking and loudness,'' J. Acoust. Soc. Am. **19**, 609–619.

Narins, P. M., and Smith, S. L. (**1986**). ''Clinal variation in anuran advertisement calls—basis for acoustic isolation,'' Behav. Ecol. Sociobiol. **19**, 135–141.

Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception, Proceedings of the 9th International Symposium on Hearing*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.

Patterson, R. D., Allerhand, M. H., 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. (**2000**). ''Auditory images: How complex sounds are represented in the auditory system,'' J. Acoust. Soc. Jpn. (E) **21**, 183–190.

Peterson, G. E., and Barney, H. I. (**1952**). ''Control methods used in the study of vowels,'' J. Acoust. Soc. Am. **24**, 75–184.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). ''The lower limit of pitch,'' J. Acoust. Soc. Am. **109**, 2074–2084.

Riede, T., and Fitch, W. T. (**1999**). ''Vocal tract length and acoustics of vocalization in the domestic dog, *Canis familiris*,'' J. Exp. Biol. **202**, 2859–2869.

Sek, A., and Moore, B. C. J. (**1995**). ''Frequency discrimination as a function of frequency, measured in several ways,'' J. Acoust. Soc. Am. **97**, 2479–2486.

Smith, D. R. R., Patterson, R. D., and Jefferis, J. (**2003**). ''The perception of scale in vowel sounds,'' British Society of Audiology, Nottingham **P35**.

Smith, D. R. R., and Patterson, R. D. (**2004**). ''The existence region of scaled vowels in pitch-VTL space,'' 18th Int. Conference on Acoustics, Kyoto Japan, Vol. I, 453–456.

Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). ''Signal and masker uncertainty in intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1015–1019.

Titchmarsh, E. C. (**1948**). *Introduction to the Theory of Fourier Integrals*, 2nd ed. (Oxford University Press, London).

Titze, I. R. (**1989**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85**, 1699–1707.

Turner, R. E., and Patterson, R. D. (**2003**). ''An analysis of the size information in classical formant data: Peterson and Barney (1952) revisited,'' J. Acoust. Soc. Jpn. **33**, 585–589.

Turner, R. E., Al-Hames, M. A., Smith, D. R. R., Kawahara, H., Irino, T., and Patterson, R. D. (**2004**) ''Vowel normalisation: Time-domain processing of the internal dynamics of speech,'' in *Dynamics of Speech Production and Perception*, edited by P. Divenyi (in press).
Wolpert, D. H. (**1996a**). ''The lack of *a priori* distinctions between learning algorithms,'' Neural Comput. **8**, 1341–1390.
Wolpert, D. H. (**1996b**). ''The existence of *a priori* distinctions between learning algorithms,'' Neural Comput. **8**, 1391–1420.
Welling, L., and Ney, H. (**2002**). ''Speaker adaptive modeling by vocal tract normalization,'' IEEE Trans. Speech Audio Process. **10**, 415–426.

Perceptual differences between low and high rates of stimulation on single electrodes for cochlear implantees

David M. Landsberger[a] and Colette M. McKay
The University of Melbourne, Department of Otolaryngology, 384-388 Albert St. East, Melbourne 3002 Australia

(Received 18 November 2003; revised 21 October 2004; accepted 22 October 2004)

Previous research has shown that increases in the rate of stimulation on a single electrode yield changes in pitch perception until the rate is increased beyond a given critical rate, after which changes in rate are only perceived as changes in loudness. The critical rate beyond which a rate increase no longer elicits a pitch change in most subjects is approximately 300 Hz, although a small number of subjects have been observed to have critical rates up to approximately 1000 Hz. In this article, we sought to determine if increasing the rate of stimulation beyond the critical rate (up to 12.8 kHz) would eventually result in new changes of perception (other than loudness.) Our data replicate the previously observed results that rates between approximately 300 and 1500 Hz are indistinguishable from each other. However, we observed the finding that a rate of stimulation well above the critical rate (starting between 1500 Hz and 12.8 kHz, depending on electrode and subject) can elicit changes in perception. The perceptual differences between these high rates were sometimes but not always labeled as pitch changes. This phenomenon needs further research to assess its potential relevance to speech perception using high rates of stimulation. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1830672]

PACS numbers: 43.66.Ts, 43.66.Fe, 43.66.Hg [NFV] Pages: 319–327

I. INTRODUCTION

Contemporary cochlear implants and speech-processing strategies are designed to provide electrical stimulation at high pulse rates. By increasing the rate of stimulation, certain elements of a signal can be encoded with finer resolution (such as amplitude or frequency modulation). It has therefore been assumed that higher rates of stimulation are inherently better. However, data examining the relationship between the pulse rate of speech-processing strategies and resulting speech perception performance yield conflicting results across subjects and experiments. The changes in performance from high rates of stimulation could be caused by changes of rate on electrodes individually. Alternatively, higher rates may cause more interaction between electrodes. For example, refractory effects caused by stimulation on one electrode can affect the response to stimulation of nearby electrodes. The experiments presented in this article investigated the perceptual differences between single-electrode stimuli using high rates. The hypothesis was that some subjects experience a change in percept at high rates on individual electrodes. If the hypothesis is confirmed, it might explain part of the variability of speech perception at different stimulation rates.

Loizou *et al.* (2000) studied speech perception performance using a six-channel CIS strategy (Wilson *et al.*, 1991, 1993). They reported that average performance for word and consonant recognition was better at 2100 Hz than at rates below 800 Hz. However, of the six subjects tested, four reached peak performance at either 800 or 1400 Hz. These results suggest that, although higher rates are generally better than low rates, there is a rate of stimulation above which additional rate increments do not affect performance. Wilson *et al.* (2000) found similar results when he examined the effect of rate on a six-channel CIS implementation with four subjects. One subject showed significant improvement in performance with increasing rate only up to approximately 500 Hz/channel, while another showed improvements in performance for rates up to approximately 3500 Hz/channel.

Using the ACE strategy [an n-of-m speech processing strategy (Vandali *et al.*, 2000)] Holden *et al.* (2002) compared speech performance at various sound-pressure levels in subjects using either 720 Hz/channel or 1800 Hz/channel. Subjects' performance was not significantly different for 720 Hz/channel and 1800 Hz/channel with CNC words, CNC phonemes, and CUNY sentences presented at 60 or 70dB. However, for CUNY sentences and CNC phonemes at 50 dB SPL, performance improved from 720 Hz/channel to 1800 Hz/channel. An examination of individual subjects' data reveals that some subjects performed better with the 720-Hz/channel rate, while others performed better with the 1800-Hz/channel rate.

Vandali *et al.* (2000) reported contrary results. Using an n-of-m strategy (a prototype of the ACE strategy), they compared speech comprehension at 65 dBA using speech processor rates of 250, 807, and 1615 Hz/channel. Using a fixed number of maxima of eight (the subset of channels selected in each stimulation cycle), no differences in performance were found between the 250-Hz/channel and 807-Hz/channel conditions. However, significantly poorer performance was found for the 1615-Hz/channel condition. Although the paper indicates that the poorer results for the 1615-Hz/channel condition are mostly caused by one subject, the data do not

[a]Electronic mail: d.landsberger@proxy.medoto.unimelb.edu.au

support the hypothesis that higher rates of stimulation yield improved performance.

Other experiments have been performed that examine speech perception performance as a function of rate of stimulation. However, these experiments covaried other parameters with rate of stimulation. For example, some experiments reduced the number of channels as they increased the rate of stimulation per channel in order to maintain a total rate of stimulation below the maximal rates of which the implants are capable. Brill *et al.* (1997) examined speech performance with a CIS strategy, varying the number of channels and rate per channel in order to maintain a fixed overall rate of stimulation. For their three subjects, Brill *et al.* found that the condition for optimal performance for a fixed overall stimulation rate varied across subjects. They concluded that ''trading channels for higher stimulation rate can improve performance, at least for some patients.'' Because the rate of stimulation was not examined independently from the number of channels, it is difficult to draw conclusions about the effect of rate of stimulation alone.

Kiefer *et al.* (2001) compared speech comprehension performance within subjects for the three most common commercial speech processing strategies, SPEAK (Seligman and McDermott, 1995; Skinner *et al.*, 1994; Whitford *et al.*, 1995), ACE, and CIS. They found that performance was best for all subjects using either ACE or CIS, which are both strategies implementing high rates of stimulation. The SPEAK processing strategy which stimulates at rates between 200 and 300 Hz was not as good as the higher rate strategies. However, most of the subjects for whom CIS was the optimal strategy performed better at 1200 Hz than they did at 1800 Hz. An analysis of the rate of stimulation in this experiment is confounded by the same trade-off Brill *et al.* (1997) encountered as the 1200-Hz map utilized 12 channels and the 1800-Hz condition utilized only six channels. It is therefore difficult to draw a strong conclusion from this experiment as to whether or not rate of stimulation as an independent variable affects performance once the rate of stimulation has exceeded the 250-Hz rate of the SPEAK strategy.

The data examining different rates of stimulation for speech-processing strategies provided conflicting conclusions regarding performance improvements with increasing rates of the processing strategies. Nevertheless, the trend has been to increase the rate of processing strategies, with the expectation that higher rates of stimulation are able to carry more information.

The per-channel rates are increased with the assumption that increased rates lead to increased information without a change in the perceptual quality. This assumption is based on data suggesting that there is a rate saturation point on single electrodes after which additional increments in rate do not change the percept of the stimulus except for changes in loudness. The rate representing the saturation point varies across subjects and experiments. Several papers have suggested that the saturation point is about 300 Hz (Blamey *et al.*, 1984; Shannon, 1983; Tong *et al.*, 1983; Zeng, 2002). Simmons *et al.* (1981) suggest the point is 350 Hz. Eddington *et al.* (1978) found the rate saturation point to lie between 70 and 400 Hz. However, a few experiments have reported discrimination of higher rates. Townshend *et al.* (1987) tested three subjects. One of these subjects had a rate saturation point of approximately 175 Hz, while the other two could distinguish between rates of 1000 Hz and below. Wilson *et al.* (2000) showed data from two subjects, one of whom had a saturation point of 500 Hz while the other had a saturation point of 1000 Hz.

While the exact value of the saturation point varies across subjects and experiments, all of these experiments show that there is a rate beyond which no perceptual differences are detected (other than loudness changes.) The procedure for locating these saturation points involves measuring the just-noticeable difference between a given rate and a higher rate. If the just-noticeable difference becomes very large or immeasurable, then the subject is declared to not be able to distinguish that rate from rates above it. The assumption is often made that all rates above this saturation point are indistinguishable. However, it is possible that there is a range of rates beginning at the saturation point which are perceptually indistinguishable from each other, but rates higher than this range would be perceptually different from the saturation rate. With the tendency to increase the per-channel rate of stimulation to rates well above saturation points, it is important to know whether or not there actually are perceptual differences at higher rates. Perceptual differences at rates well above the saturation point could affect performance and perceptual quality of high rates of stimulation on multichannel speech processors.

In the following studies, we examined the ability of implantees to discriminate different rates of stimulation on single electrodes ranging between 100 Hz and 12.8 kHz. The procedures were repeated for a basal, an apical, and a medial electrode for all subjects. Experiments 1 and 2 investigate discrimination of rates in the ranges of 100 to 1500 Hz and 600 Hz to 12.8 kHz, respectively. The experiments investigated whether higher-rate stimuli could be ranked in order of pitch.

II. EXPERIMENT 1

A. Subjects and methods

Seven users of the Nucleus CI24 implant participated in the study. All normally used an n-of-m strategy (either SPEAK or ACE). Details of the etiology of deafness and implant use are presented in Table I.

All stimuli were presented to subjects by means of a SPEAR speech processor (HearWorks Pty. Ltd., 2003) that was controlled by a personal computer using custom-written software. Stimulation consisted of constant-rate biphasic pulses in monopolar mode (in which current flows from an electrode implanted in the cochlea to two extracochlear return electrodes, MP1+2) on a single electrode for 500 ms. All stimuli were presented with a phase duration of 26 μs and an interphase gap of 8.4 μs. Stimuli varied only in which electrode was used, the rate at which the stimulation occurred on that electrode, and the current level needed to maintain the required loudness. For each subject, three electrodes were chosen to represent an apical, basal, and a medial electrode. Electrode choices for each subject are listed in

TABLE I. Subject details. The ''electrode array'' column refers to which Nucleus CI24 device was used. Straight refers to the CI24M and Contour refers to CI24R (CS).

Subject	Age when profoundly deaf	Years of profound deafness	Etiology	Strategy	Rate per channel (Hz)	Electrode array	Electrodes used
AB	61	14	ME infections as a child.	ACE	1800	Contour	6,13,20
BK	19	50	Unknown	ACE	500	Contour	6,13,20
DC	Unknown	Unknown	Noise exposure	ACE	900	Contour	6,13,20
FZ	58	10	Unknown	ACE	1200	Straight	6,13,19
GB	77	1	Noise exposure	SPEAK	250	Straight	6,13,20
JM	57	4	Meniere's	SPEAK	250	Contour	3,12,20
MM	34	5	Unknown	ACE	900	Straight	5,13,19

Table I. Stimulation rates included the range from 100 to 1000 Hz in 100-Hz steps, as well as 1500 Hz.

1. Loudness balancing

Maximum comfortable current levels (C-levels) were measured for all stimuli using an ascending method of adjustment. Current levels were increased for each stimulus until they were just slightly too loud for comfort and then reduced to the point where they were loud but comfortable again.

For each electrode, all the stimuli were loudness balanced to the C level of the 600-Hz stimulus. Two separate procedures were used for the loudness balancing. The first procedure (used for subjects FZ, AB, GB, and JM) consisted of repeatedly playing one stimulus followed by the other stimulus with a 500-ms interstimulus interval (ISI). The second (test) stimulus to be presented was initially much quieter than the first stimulus so that it was easy for the subjects to differentiate between the two sounds. Subjects were instructed to adjust the loudness of the test stimulus using a toggle switch until the loudness of the two stimuli was the same. They were encouraged to raise the loudness of the second stimulus until it was slightly louder than the first stimulus, and then reduce the loudness until the levels matched. This procedure was then repeated, adjusting the current level of the other stimulus. The differences in current levels between the two balanced stimuli were averaged to calculate the current level of the test stimulus that matched the loudness of the 600-Hz stimulus.

Some subjects (MM, BK, and DC) had difficulty with the previous loudness-balancing procedure. For these subjects, the loudness balancing procedure was an adaptive 2-interval, forced-choice task. In this procedure, the reference and test stimuli were presented once with a 500-ms ISI between them. After hearing the two sounds, subjects pressed a button on a response box corresponding to which of the two stimuli was louder. The loudness of the test stimulus was adjusted up or down for the next trial based on whether the subject considered the test stimulus softer or louder than the reference stimulus. After ten reversals had been made, the last six reversals were averaged to calculate the current level for the test stimulus that matched the loudness of the 600-Hz stimulus. The procedure was repeated using the test stimulus as the standard stimulus and the 600-Hz stimulus as the adjustable stimulus. These data were used to calculate the current level of the loudness-balanced test stimulus in the same way as used in the first loudness-balancing procedure.

2. Pulse rate discrimination

Pulse rate discrimination was assessed using a 4-interval, forced-choice task. In each trial, four stimuli were presented on the same electrode; three stimuli were presented at one rate of stimulation while the fourth stimulus was presented at another rate of stimulation. Each stimulus was separated with a 500-ms ISI. The amplitude of each stimulus was set randomly per presentation to be within ± 2 current levels of the loudness balanced levels to minimize loudness cues. A current level step is about 0.18dB. A ± 2 current level jitter should be sufficient to minimize loudness cues, as an analysis of variability of our loudness balancing yielded a 95% confidence interval of ± 0.78 current steps.

Subjects were instructed to identify which of the four sounds was different from the others in any way except loudness. Subjects pressed a button on a response box corresponding to the interval with the different sound. The computer recorded button presses. A block of trials consisted of 110 trials on one electrode where all 11 rates of stimulation were compared to the other ten rates of stimulation. Each block was repeated five times for each of three electrodes. Thus, there were ten comparisons for each rate pair (five in each order).

B. Results

To analyze the data, we looked at the percentage of times a subject was able to correctly identify which stimulus was different for each stimulus pair. The ability to detect a difference was declared to be significant if the subject was able to distinguish between two stimuli 60% or more of the time. With a chance level of 0.25, the probability of getting 60% or more correct by chance is less than 0.02. Many comparisons are made in this experiment, which increase the chances of type I errors. With $\alpha = 0.02$, we expect 1 in 50 indistinguishable differences to be found to be significantly different. This is acceptable because this rate of type I errors

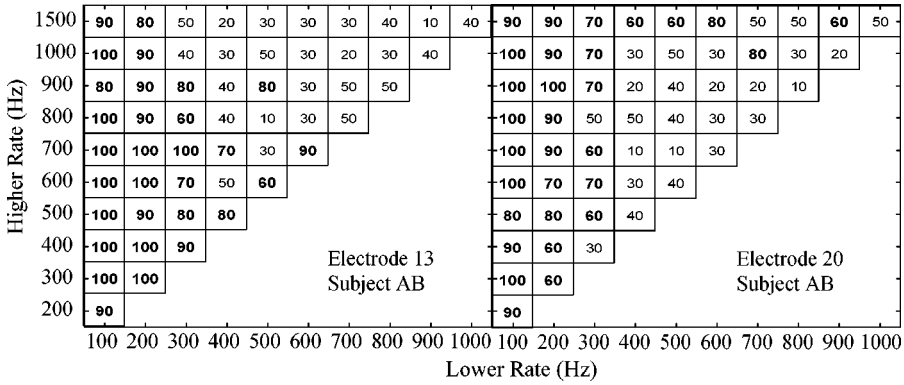


FIG. 1. Data for subject AB in experiment 1. The *x* axis represents the lower rate of the two compared rates and the *y* axis represents the higher rate. The percentage of time the subject was able to identify the correct answer is labeled in the corresponding square. Chance performance is 25%. Significant differences ($p<0.02$) are shown in bold. The figure on the left represents the data for subject AB on his medial electrode (electrode 13). On electrode 13, AB is able to identify rates of 100 and 200 Hz as being different from higher rates. Performance is also good at detecting the difference between 300 Hz and higher rates. These results are typical of what was found on all subjects and all electrodes except for electrode 20 for subject AB. Data for subject AB on electrode 20 are represented on the right, which shows that subject AB was able to detect well above chance the difference between 1500 Hz and most lower rates.

would not alter the pattern of results. It is the pattern of results that is investigated by this experiment, and not the results of any given rate-pair comparison.

With one exception (subject AB, electrode 20), the results for all subjects and electrodes were similar. Figure 1 shows data for subject AB on electrode 13 on the left and electrode 20 on the right. The data show that on electrode 13 for subject AB, 100 and 200 Hz are differentiable from each other, as well as all higher rates with scores of more than 80% correct. Subject AB also performed well on comparisons between 300 Hz and higher rates of stimulation with an average score of 71.25% correct. However, rates of 400 Hz and above did not seem to be systematically different from each other. These results are consistent with previous findings that have suggested that rate of stimulation up to approximately 300 Hz creates a change in pitch and that at rates higher than 300 Hz, there are no perceptual differences other than loudness (Blamey *et al.*, 1984; Shannon, 1983; Simmons *et al.*, 1981; Tong *et al.*, 1983; Zeng, 2002).

For subject AB on electrode 20, rates between 100 and 300 Hz were easily differentiated from each other and higher rates. Rates between 400 and 1000 Hz were not easily differentiated. However, 1500-Hz stimuli were found to be different from lower rates 7 out of 10 times. Subject AB correctly identified the different stimulus 41 out of 70 times ($p<0.0001$ with chance at 25%) when 1500-Hz stimuli were compared to rates between 400 and 1000 Hz on electrode 20.

No differences were observed between the apical, medial, and basal electrodes and all subjects showed similar patterns. We therefore collapsed the data across all subjects and electrodes to create Fig. 2. The values in each cell of Fig. 2 represent the number of times the rate pair was discriminated with 60% accuracy ($p<0.02$). There are 21 data points (7 subjects$\times$3 electrodes) represented in each cell. With $\alpha=0.02$, if subjects were unable to detect any differences, we would expect a type I error in fewer than every other cell. A type I error would manifest itself as an increment of 1 in a cell. Figure 2 shows that 100-Hz stimuli were consistently identified as different than higher rates. Two-hundred-Hz stimuli were identified as different consistently as well, but not quite as often as in the 100-Hz case. Four-hundred-Hz stimuli were not consistently identified as different than higher rates.

The total number of times a subject correctly identified the different stimulus for each electrode for experiment 1 was correlated with the stimulation rate of the subject's clinical speech processor and no relationship was found ($p=0.789$).

III. EXPERIMENT 2

A. Subjects and methods

Six of the seven subjects used in experiment 1 participated in experiment 2. Subject MM was unable to participate in this experiment.

Experiment 2 consisted of a higher-rate replication of the previous experiment. The stimuli were steady rate pulse trains presented on a single electrode at 600, 900, 1800, 2400, 3600, 7200 Hz, and 12.8 kHz. Each stimulus was presented for a 500-ms duration. All other stimulus parameters

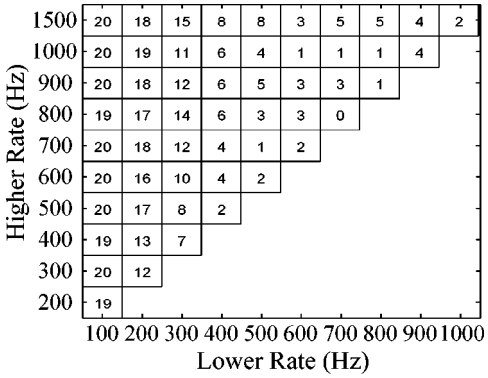


FIG. 2. Summary data collapsed across all subjects and electrodes for experiment 1. The values in each cell represent the number of times that the rate pair has been discriminated significantly (60% or better). There are 21 data points (7 subjects$\times$3 electrodes) represented in each cell. The figure legends are the same as in Fig. 1. One-hundred- and 200-Hz stimuli are detectably different from each other, and all higher rates in almost all cases. Three-hundred-Hz stimuli show a weaker, but similar pattern. Rates between 400 and 1500 Hz are not reliably detected as different from each other.

remained the same from the previous experiment. Loudness was balanced to C levels at 600 Hz for each subject using the same procedure used for that subject in the previous experiment. The ISI was increased to 1000 ms because of technical limitations in our software. In each block of trials all seven stimuli were paired with the remaining six stimuli. This procedure was repeated five times for each of the three electrodes, yielding ten comparisons per rate pair.

B. Results

Because different patterns are present with different subjects and different electrodes, the data for each subject are presented in Fig. 3. While the results vary, all subjects showed that they could distinguish some high-rate stimuli that are well above their measured saturation points in experiment 1. Data from one of the best performers, subject DC, are on the top row of Fig. 3. Subject DC could generally distinguish rates below 900 Hz from rates above 3600 Hz on all three electrodes. Additionally, subject DC could distinguish many other rates on each electrode. The poorest performer's data are on the bottom row of Fig. 3. This subject (FZ) was unable to differentiate any of the rates between 600 and 7200 Hz on any of his electrodes. He was only able to distinguish 600 and 900 Hz from 12.8 kHz on the medial electrode and 12.8 kHz from 600, 900, 2400, 3600, and 7200 Hz on the basal electrode. On the apical electrode, no sounds were differentiated.

The total number of times a subject correctly identified the different stimulus for each electrode for experiment 2 was correlated with the stimulation rate of the subject's clinical speech processor, and no relationship was found ($p = 0.765$). A significant relationship was found between the total number of times a subject correctly identified the different stimulus on each electrode between experiment 1 and 2 ($p = 0.021$).

IV. EXPERIMENT 3

A. Subjects and methods

The third experiment was a qualitative exploration of the differences between the perceptions of the various high-rate single-electrode stimuli. We explored whether or not the differences in percept could be attributed to pitch, similar to the changes in percept found from a change in rates at low stimulation rates.

Five of the six subjects used in experiment 2 participated in experiment 3. Subject FZ was excluded from this study because he did not show any consistent ability to detect differences between the various high-rate stimuli in experiment 2.

Before performing the experiment, we asked all subjects if they understood the concept of pitch. To help explain the concept of a pitch difference we presented subjects with a 100- and a 600-Hz stimulus on the same electrode, and explained that the 600-Hz stimulus had a higher pitch. We also explained that the difference between a man and woman's speaking voice was usually a difference in pitch, and that a woman's voice is usually higher in pitch than a man's voice. Once we were comfortable that the subject properly understood the difference between high and low pitches (most subjects seemed comfortable with the concept before we addressed the issue), we began the experiment.

The stimuli that were used in this experiment were the same set of stimuli used in experiment 2. Stimuli were presented at the same loudnesses used in experiment 2 with a ± 2 current level jitter. In each trial, two different stimuli were presented on the same electrode. The two stimuli were presented with a 1000-ms ISI. Subjects were asked to press one of two buttons on a button box corresponding to the interval containing the higher-pitched stimulus. Each block of trials compared every stimulus with every other stimulus on the same electrode twice for a total of 42 trials in a block. Each trial block was repeated 5 times for each of 3 electrodes totaling 630 trials. Each rate was compared with each other rate a total of 10 times during the 630 trials. An extra trial block was run on subject DC's apical electrode.

B. Results

To analyze the data, we calculated how often each subject rated the higher-rate stimulus as having the higher pitch. If no pitch difference was detected, then expected performance would be at chance level (50%). Scores significantly above chance indicate that higher rates are perceived as having higher pitches, and scores significantly below chance indicate a pitch reversal (that higher rates are perceived as having a lower pitch.) For our first analysis, all pairs of stimuli were analyzed, regardless of whether or not a given pair of stimuli was found to be perceptually different for a given subject in the previous experiment. In examining our data, we found higher rates generating higher pitches, higher rates generating lower pitches, and higher rates producing no pitch changes. Sometimes we found all three patterns in the same subject as was the case for subject DC. Subject DC's data is summarized in Fig. 4.

A strong relationship for higher rates yielding a perception of a higher pitch was observed for subject DC's basal and medial electrodes. Subject DC identified the higher rate stimulation as having the higher pitch 89.05% of the time for the basal electrode and 86.19% of the time for the medial electrode. On the basal electrode, subject DC identified the higher rate of stimulation as having the higher pitch 100% of the time for 12 out of the 21 compared pairs of stimuli. Most of the other rate comparisons on this electrode yielded higher pitch percepts for higher rate stimulation well above the chance level. Subject DC's results for the medial electrode showed similar patterns. For all pairs of stimuli for which there was at least one stimulus with a rate in between the two, subject DC identified the higher rate as having the higher pitch at least 90% of the time.

However, this pattern was not found across all subjects and electrodes. The relationship between rate and pitch observed for subject DC's basal and medial electrodes was not present for his apical electrode. When rates between 600 and 1800 Hz were compared with each other or higher rates, either stimulus was approximately equally likely to be rated as having a higher pitch. However, for rates above 2400 Hz, subject DC perceived higher rate stimulation as having a lower pitch on this electrode.

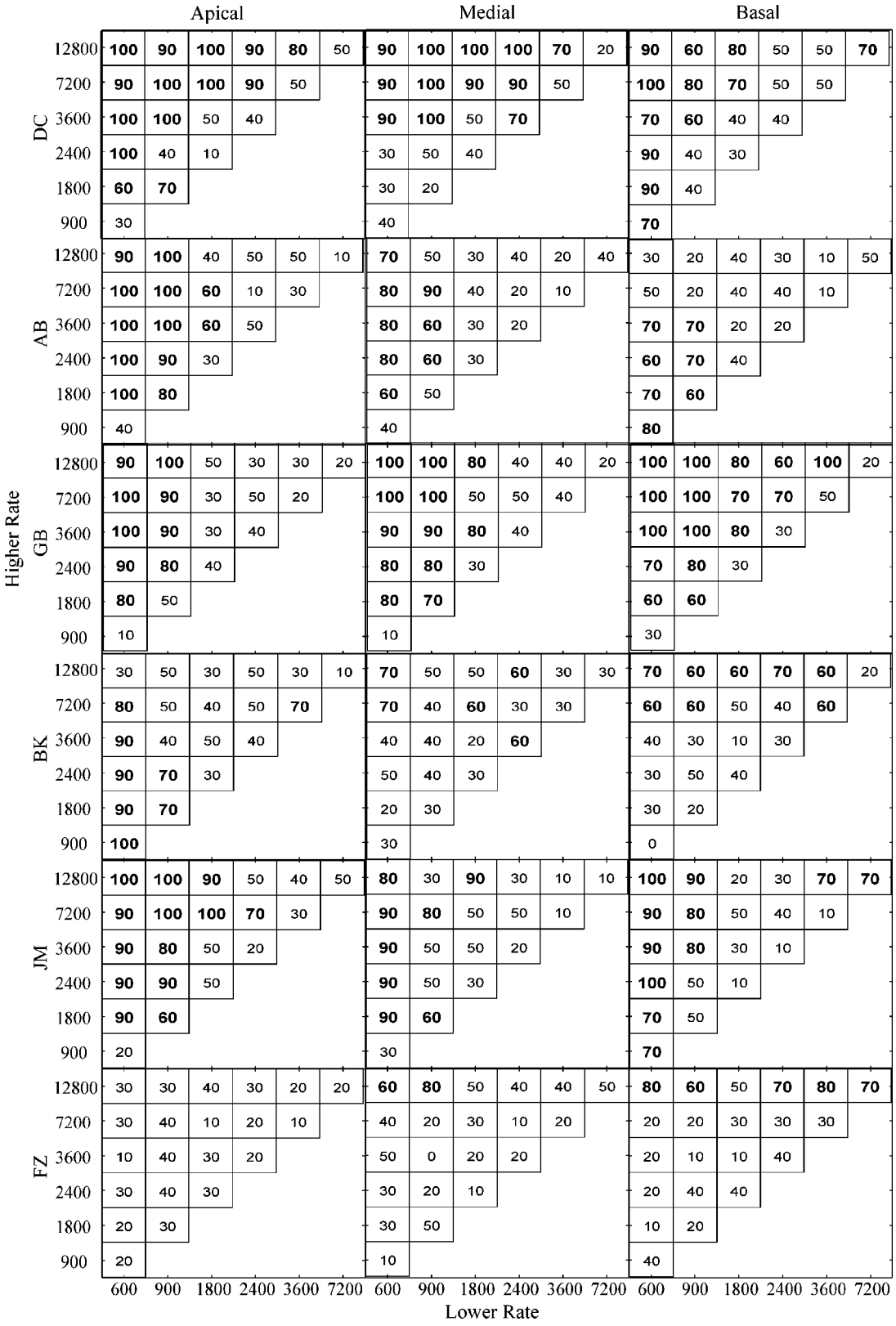


FIG. 3. Data for all six subjects and all three electrodes. The three columns of data represent the basal, medial, and apical electrodes, respectively. The six rows of data represent the six subjects in the experiment. The values in each of the cells indicate the percentage of comparisons where the subject was able to correctly identify the different stimulus. Significant differences ($p<0.02$) are shown in bold.

While each electrode for each subject showed one of three patterns (higher rate yields higher pitch, higher rate yields lower pitch, rate changes do not affect pitch), there were no consistent patterns across electrodes or subjects. A summary of these data is presented in Fig. 5. Figure 5 consists of a series of box plots for all three electrodes for each subject. Data points were only included in the box plots if, for the same comparison by the same subject, they were

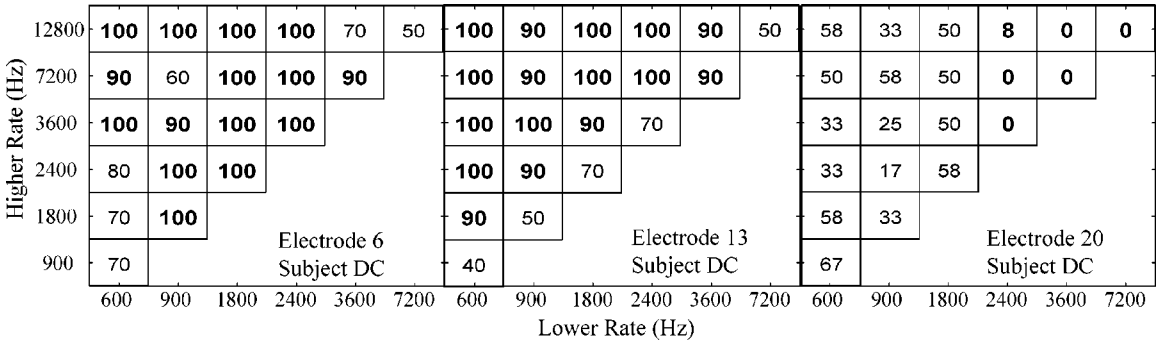


FIG. 4. Experiment 3 data for subject DC. The figures from left to right represent results for subject DC's basal, medial, and apical electrodes, respectively. The values in each cell represent the percentage of times the higher rate stimulus was rated as having the higher pitch. There are 10 comparisons per cell for the basal and medial electrodes and 12 comparisons for the apical electrode. Values that are significantly different from chance (50%) for $\alpha = 0.05$ are shown in bold. For subject DC's basal and medial electrodes, the pattern of higher rate stimulation yielding a higher-pitch is very strong. For subject DC's basal electrode, when rates between 2400 Hz and 12.8 kHz are compared, higher rates yield lower pitches. For other comparisons on DC's basal electrode, changes in rate do not seem to yield a change in pitch.

detected as being different at least 60% of the time in experiment 2. Except for one point, subject AB consistently rated the higher rate stimulus as having the higher pitch across all three electrodes. Subject JM found higher rate to yield lower pitch on his basal electrode and no pitch differences on his other electrodes. Subject BK reported higher rate stimulation as having a higher pitch on her basal electrode. On the other two electrodes any pitch differences appear to be a pitch reversal. From these data, the conclusion can be drawn that there are indeed pitch differences for changes in high-rate stimulation for some subjects and some electrodes. However, stimuli that are rated as different in experiment 2 are not always perceived as having a different pitch. Therefore, some of the detectable differences between rates of stimulation must not be in the pitch dimension.

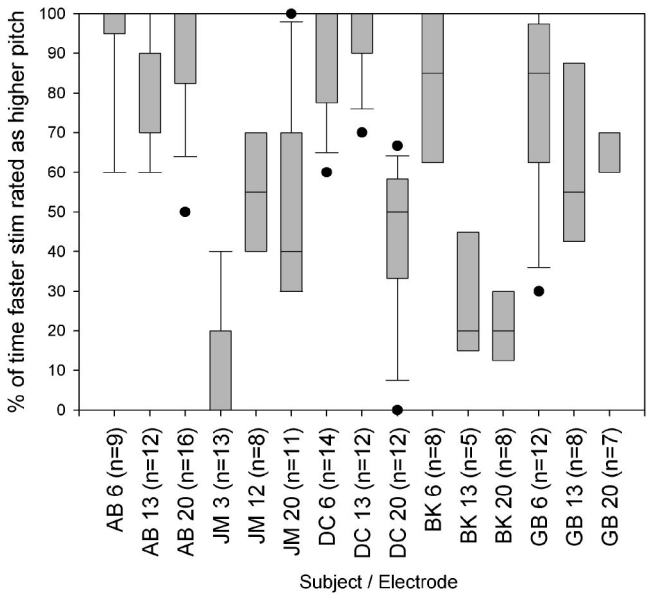


FIG. 5. Boxplot summary data for experiment 3. Each box plot represents the distribution of the data points representing how often the higher-rate stimulus is labeled as having the higher pitch for each electrode and subject. Data points were only included if for the same rate comparison on the same electrode, the subject was able to correctly identify the different stimulus 60% or more of the time. The labels on the *X* axis contain the subject's ID, electrode number, and the number of data points in the corresponding box plot.

V. DISCUSSION

Results for the first experiment are consistent with results found in many other experiments (Blamey *et al.*, 1984; Shannon, 1983; Simmons *et al.*, 1981; Tong *et al.*, 1983; Zeng, 2002). Stimulation on single electrodes for rates up to a saturation point (approximately 300 Hz) were found to produce a perceptual change. All rates above the saturation point were perceptually indistinguishable when loudness was balanced. Unlike the findings of Townshend *et al.* (1987) and Wilson *et al.* (2000), none of the subjects was able to distinguish rates between 400–1000 Hz. With the exception of electrode 20 for subject AB, the 1500-Hz stimulus was not perceptually different from rates between 500 and 1000 Hz. These results suggest that either there are no perceptual differences above approximately 300 Hz or that perceptual changes at high rates generally begin above 1500 Hz.

Experiment 2 provides evidence that there are indeed perceptual changes at high rates above 1500 Hz. Previous experiments had not reported differences at high rates on single electrodes because the highest rates of stimulation examined were not generally over 1000 Hz. However, the results for experiment 2 show that all subjects are able to detect some differences (other than loudness) at rates higher than 1000 Hz. While the high rate required to produce these changes varies across subjects and electrodes, for all subjects except FZ there exists some rate above which changes occur. No relationship was found between rate discrimination performance and the stimulation rate used in the subjects' clinical speech processor.

The data from experiment 3 suggest that some of the detected changes in experiment 2 were pitch changes. However, other detected changes must have been along another perceptual dimension, as they were reliably reported as different but not as higher or lower in pitch. When asked to describe the differences that they heard, some stimuli were described as being different in pitch, while others were reported as changes in clarity or consistency of the sound. No adaptation was reported as there was no evidence of stimuli becoming quieter. However, the 12.8-kHz stimulus was occasionally described as fading in and out. One subject (JM) was able to pick out the 12.8 kHz and occasionally the

7200-Hz stimulus on the apical electrode by recognizing a click at the end of the sound. Electrodograms were made of the 12.8-kHz stimuli to determine if there was anything unusual about the physical stimuli (such as a software artifact). The electrodogram demonstrated that the implant was given the proper instructions for a 12.8-kHz stimuli. We were unable to verify, however, that subject JM's implant was actually delivering the stimulus correctly. To learn more about the click at the end of 12.8-kHz stimuli, we presented them to subject JM for varying durations (from 100 to 2000 ms) and found that regardless of duration, the click always occurred at the offset of stimulation.

On electrode 20, subject DC reported that as rates increased from 2400 Hz to 12.8 kHz the pitch became lower. These results are unlikely to be a result of the subject's confusion of high versus low pitch as the subject reliably rated higher rates as yielding higher pitches for electrodes 6 and 13 and was reliable at labeling pitch changes for low rates (below 300 Hz.) Furthermore, subject DC has had extensive musical training. One possibility is that this percept is created by the refractory periods of auditory neurons. Wilson *et al.* (1997) examined intracochlear evoked potentials for single-electrode pulse trains. He found that for rates of stimulation greater than 400 Hz, the magnitude of the evoked potentials varied with different pulses in the pulse train. Between 400 and 800 Hz, the magnitude of the evoked potentials alternated between two levels. As the rate of stimulation increased, the patterns of the evoked potentials became more complicated with some subjects. Two subjects showed a strong evoked potential every sixth pulse and a weak evoked potential for the seventh pulse at 1016 and 1524 Hz, respectively. Perhaps if the pitch percept we measured is related to the periodicity of the greatest magnitudes of the evoked potentials, then it is possible that higher-rate stimuli produce a lower periodicity. While these results were observed on only one electrode with only one subject, Wilson *et al.* reported differences in responses of neural populations in response to stimulation on different electrodes within a given subject. It is therefore possible that the neural activity in response to the subject DC's apical electrode was different from that for other electrodes and other subjects.

Most current commercial and research speech processing strategies use high per electrode rates of stimulation (Loizou *et al.*, 2003). Higher rates of stimulation allow for greater temporal resolution in stimulation. It is generally assumed that these high rates of stimulation do not cause changes in percepts on single electrodes. However, the data in this article suggest that this assumption is faulty. The perceived changes for different high rates are likely to have an effect on speech perception at different overall rates. If at high rates different rates of stimulation sound different, it is reasonable to assume that speech perception may be affected by the changes in rates. For example, if higher rates of stimulation yield a higher pitch percept, then higher rate-processing strategies may shift the perceived pitches of sounds. While this may or may not cause problems for high-rate strategies, it is important to be aware of these issues when designing or implementing a high-rate strategy.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Health fellowship 1 F32 DC 006774-01 awarded to the first author and by fellowships from the Garnett Passe and Rodney Williams Memorial Foundation and the National Health and Medical Research Council of Australia awarded to the second author. The authors wish to thank Katherine Henshall for her assistance, and the seven cochlear implantees who generously provided their time.

Blamey, P. J., Dowell, R. C., Tong, Y. C., and Clark, G. M. (**1984**). ''An acoustic model of a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **76**, 97–103.

Brill, S. M., Gstottner, W., Helms, J., von Ilberg, C., Baumgartner, W., Muller, J., and Kiefer, J. (**1997**). ''Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+,'' Am. J. Otol. **18**, S104–106.

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. L. (**1978**). ''Auditory prostheses research with multiple channel intracochlear stimulation in man,'' Ann. Otol. Rhinol. Laryngol. **87**, 5–39.

HearWorks Pty. Ltd. (**2003**). ''SPEAR3 Research System,'' http://www.hearworks.com.au/spear.

Holden, L. K., Skinner, M. W., Holden, T. A., and Demorest, M. E. (**2002**). ''Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy,'' Ear Hear. **23**, 463–476.

Kiefer, J., Hohl, S., Sturzebecher, E., Pfennigdorff, T., and Gstoettner, W. (**2001**). ''Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system,'' Audiology **40**, 32–42.

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.

Loizou, P. C., Stickney, G., Mishra, L., and Assmann, P. (**2003**). ''Comparison of speech processing strategies used in the Clarion implant processor,'' Ear Hear. **24**, 12–19.

Seligman, P., and McDermott, H. (**1995**). ''Architecture of the Spectra 22 speech processor,'' Ann. Otol. Rhinol. Laryngol. Suppl. **104 (Suppl. 166)**, 139–141.

Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.

Simmons, F. B., White, R. L., Walker, M. G., and Mathews, R. G. (**1981**). ''Pitch correlates of direct auditory nerve electrical stimulation,'' Ann. Otol. Rhinol. Laryngol. Suppl. **90**, 15–18.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H. J., and Beiter, A. L. (**1994**). ''Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System,'' Am. J. Otol. **15 Suppl 2**, 15–27.

Tong, Y. C., Blamey, P. J., Dowell, R. C., and Clark, G. M. (**1983**). ''Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **74**, 73–80.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). ''Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system,'' Ear Hear. **21**, 608–624.

Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J. *et al.* (**1995**). ''Evaluation of the Nucleus Spectra 22 processor and new speech processing strategy (SPEAK) in postlinguistically deafened adults,'' Acta Oto-Laryngol. **115**, 629–637.

Wilson, B., Wolford, R., and Lawson, D. (**2000**). ''Speech processors for auditory prostheses'' (Center for Auditory Prosthesis Research: Research Triangle Park, NC), 1–61.

Wilson, B. S., Lawson, D. T., Zerbi, M., and Finley, C. C. (**1993**). ''Recent developments with the CIS strategies,'' in *Advances in Cochlear Implants:*

Proceedings of 3rd International Cochlear Implant Conference, Innsbruck, Austria, April 1993, edited by I. J. Hochmair-Desoyer and E. S. Hochmair (Manz, Vienna), 103–112.
Wilson, B. S., Finley, C. C., Lawson, D. T., and Zerbi, M. (**1997**). ''Temporal representations with cochlear implants,'' Am. J. Otol. **18**, S30–34.
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.
Zeng, F. G. (**2002**). ''Temporal pitch in electric hearing,'' Hear. Res. **174**, 101–106.

Estimation of vocal dysperiodicities in disordered connected speech by means of distant-sample bidirectional linear predictive analysis

Frédéric Bettens and Francis Grenez
Department Signals and Waves, Université Libre de Bruxelles, CP 165/51, 50, Av. F.-D. Roosevelt, B-1050 Brussels, Belgium

Jean Schoentgen[a)]
Laboratory of Experimental Phonetics, Université Libre de Bruxelles, CP 110, 50, Av. F.-D. Roosevelt, B-1050 Brussels, Belgium; National Fund for Scientific Research, Belgium

(Received 6 February 2004; revised 4 September 2004; accepted 25 October 2004)

The article presents an analysis of vocal dysperiodicities in connected speech produced by dysphonic speakers. The processing is based on a comparison of the present speech fragment with future and past fragments. The size of the dysperiodicity estimate is zero for periodic speech signals. A feeble increase of the vocal dysperiodicity is guaranteed to produce a feeble increase of the estimate. No spurious noise boosting occurs owing to cycle insertion and omission errors, or phonetic segment boundary artifacts. Additional objectives of the study have been investigating whether deviations from periodicity are larger or more commonplace in connected speech than in sustained vowels, and whether sentences that comprise frequent voice onsets and offsets are noisier than sentences that comprise few. The corpora contain sustained vowels as well as grammatically- and phonetically matched sentences. An acoustic marker that correlates with the perceived degree of hoarseness summarizes the size of the dysperiodicities. The marker values for sustained vowels have been highly correlated with those for connected speech, and the marker values for sentences that comprise few voiced/unvoiced transients have been highly correlated with the marker values for sentences that comprise many. [DOI: 10.1121/1.1835511]

PACS numbers: 43.70.Gr, 43.70.Jt [AL] Pages: 328–337

I. INTRODUCTION

This article concerns the analysis of vocal dysperiodicities in connected speech produced by dysphonic speakers. The method is based on a distant-sample bidirectional linear predictive processing of speech. The analysis has the following properties. The size of the dysperiodicity estimate is zero for periodic speech signals. A feeble increase of the vocal dysperiodicity is guaranteed to produce a feeble increase of the estimate. No spurious noise boosting occurs owing to cycle insertion and omission errors, or phonetic segment boundary artifacts. The method is compared to an analysis of vocal dysperiodicities in running speech, which has been published by Qi *et al.* (1999).

Voice disorders, which may result from disease or accidents affecting the larynx, refer to abnormal conditions of the glottal excitation signal that is generated by the vibrating vocal folds and pulsatile glottal airflow. One frequent symptom of voice disorders is an increase of the dysperiodicity of voiced speech sounds. This observation has led to the development of speech processing tools for clinical and research-related voice assessment.

Most published studies rely on the isolation of individual speech cycles or of individual harmonics in the speech spectrum. Therefore, such analyses have mainly been applied to sustained pseudoperiodic vowel fragments, for which the detection of quantities such as cycle lengths and amplitudes as well as spectral pseudoharmonics is likely to succeed. Common examples of acoustic markers that are so obtained include the average perturbation of the length or amplitude of the speech cycles, as well as harmonics-to-noise ratios.

Studies devoted to vocal dysperiodicities in connected speech or sustained vowels including onsets and offsets are comparatively rare. Table I summarizes published studies that have been devoted to the acoustic characterization of connected speech produced by dysphonic speakers. Apart from one analysis that is discussed hereafter, the studies listed in Table I have in common that they involve the detection of individual speech cycles, or individual pseudoharmonics or rahmonics of the speech spectrum. But, the reliable detection of these is not always possible in running speech produced by hoarse speakers. Cycle or harmonic omission and insertion errors are therefore bound to occur, which bias numerically the values of acoustic markers of vocal noise.

Qi *et al.* (1999) have recently proposed an analysis method of disordered speech, which is not based on the detection of glottal events. The proposal involves a clinical application of a speech signal model that has been developed earlier for speech coding. The processing comprises a proximal (near-sample) and distal (distant-sample) linear prediction of the speech signal. The proximal linear prediction stage is a conventional linear predictive analysis. During this stage, near-sample correlations owing to the vocal tract

a)Electronic mail: jschoent@ulb.ac.be

TABLE I. Summary of studies devoted to the acoustic characterization of connected speech produced by dysphonic speakers. The symbols designate the following: mi=microphone signal, gl=glottogram, tr=throat microphone signal; se=sentence, pa=paragraph, V=vowel, C=consonant; ma=manual, aut=automatic; te =temporal, sp=spectral.

Year	Author(s)	Signal(s)	Material	Speakers	Processing	Feature(s)
1963	Lieberman	mi	se	31	ma	te
1965	Shipp *et al.*	mi	pa	15	ma	te
1968	Dolansky *et al.*	mi, gl, tr	V, VCV, se	10	ma	te
1971	Hecker *et al.*	mi	se	28	ma	te
1971	Lebrun *et al.*	mi	se		ma	te
1980	Gubrynowicz *et al.*	mi	pa	143	aut	te
1984	Abberton *et al.*	gl	pa		aut	te
1986	Laver *et al.*	mi	pa	230	aut	te
1988	Muta *et al.*	mi	V	6	ma, aut	sp
1990	Klingholz	mi	V, pa	124	aut	sp
1993	Valencia *et al.*	mi	CV	12	ma	sp
1996	Hillenbrand *et al.*	mi	V, se	25	ma, aut	te, sp
1999	Qi *et al.*	mi	se	87	aut	te
2001	Parsa *et al.*	mi	V, pa	228	aut	te, sp
2003	Kruse	mi	pa		aut	te, sp

transfer function are removed from the speech signal.

The outcome of this initial stage is the residue signal, which is the modeling error that is made while attempting to model the present speech sample by a weighted sum of recent samples [Markel and Gray, 1974]. The purpose is to exploit near-sample correlations to represent speech signals economically. For voiced speech, the residue signal comprises pseudoperiodic spikes that are due to the inability of a linear model to represent, by recent speech samples, the position and shape of the following glottal excitation pulse. Ramachandran and Kabal (1989) have therefore suggested performing a second, distant-sample, predictive processing of the residue signal. The distal predictive model is fitted to the residue signal so as to represent future by past residue spikes. The outcome of this second processing stage is a modeling error that is ideally spike-free.

Qi *et al.* have proposed to use this second-stage, distal, modeling error as a cue of vocal noise. Distal linear prediction indeed relies on the assumption that the present cycle can be modeled by means of the previous one if the signal, which is so analyzed, is cyclic and if the cycle amplitude evolves smoothly. This modeling is perfect (that is, the distal prediction error is zero) when the signal is periodic. The error is small if the signal is pseudoperiodic and it increases if the original cycle lengths or shapes evolve randomly from one cycle to the next, or if the original signal is heavily contaminated by additive noise. Finally, the variance of the distal prediction error is at a maximum when the original signal is decorrelated noise.

In a clinical framework, desirable properties of double-predictive analysis (i.e., proximal followed by distal predictive analysis) are that the size of the distal prediction error evolves smoothly with the degree of vocal dysperiodicity. Indeed, owing to the linearity of the processing, a small increase in the vocal noise is guaranteed to produce a small increase in the distal prediction error. Actually, a spurious boosting of the measured perturbations owing to erroneous omissions or detections of cycle markers cannot occur because distal linear prediction is not based on the assumption that speech periods or harmonics exist and that they can be detected. It instead relies on the knowledge of the distal prediction error, which can be computed for voiced, unvoiced, and transient speech sounds. The distal modeling is indeed performed over a distance that minimizes the prediction error. This distance agrees with a multiple of the length of a glottal cycle when the speech sound is voiced and pseudoperiodic. When the sound is not voiced, the distance that minimizes the distal prediction error remains meaningful for computational purposes, but is not interpreted in terms of the glottal cycle length.

Double-linear predictive analysis may, however, misrepresent genuine dysperiodicities or insert spurious ones for the following reasons. First, the conventional, proximal, prediction error has been proposed in the past as a cue of vocal noise. It appears that it is not optimally suited for clinical applications because its size and shape also depend on conditions that are clinically not relevant [Schoentgen, 1982]. It is expected, for instance, to increase for nasal or nasalized speech segments, as well as for high-pitched or female voices [Kahn and Garst, 1983].

Second, the proximal linear prediction model represents the current speech sample by a weighted sum of past samples and an error signal. A consequence is that any vocal dysperiodicity is also distributed between two terms, i.e., weighted sum and residue signal. Within the framework of double-linear prediction, the weighted sum is discarded and the residue signal is analyzed further by distal predictive analysis. The dysperiodicities of the residue signal are therefore different from the genuine dysperiodicities that are present in the speech signal.

Third, voiced and unvoiced, as well as silent and noisy segments exist in running speech. The proximal, conventional, prediction error may therefore be spiky or spike-free. This means that the distant-sample predictive analysis must occasionally represent spiky fragments by spike-free ones, and vice versa. Because such a representation is not optimal,

spurious transients may be observed in the distal prediction error. These transients, which are erroneously assigned to vocal noise, are expected to occur at the beginning of the recording interval, as well as next to the boundaries between voiced and unvoiced phonetic segments.

One objective of this article is to show that the solution to these problems consists of replacing Qi *et al.*'s double (proximal and distal) forward predictive analysis by a distal forward and backward analysis. The proximal prediction stage is mandatory in the framework of Qi *et al.*'s proposal, but optional in the framework of bidirectional analysis for reasons that are explained hereafter.

The reasons for replacing double-forward analysis by a distal bidirectional analysis are the following. First, bidirectional processing removes the spurious transients of the distal prediction error that are expected to occur near the boundaries between spiky and spikeless fragments of the residue signal, or near the phonetic boundaries in the speech signal. Bidirectional analysis indeed consists of performing a forward as well as backward distal predictive analysis, recording the backward and forward distal prediction errors, and keeping whichever is smallest as a cue of the vocal dysperiodicities. Unwanted error transients are so removed because the prediction across segment boundaries is omitted. It is replaced by the distal retrodiction of the onsets from the middle fragment of a segment. Retrodiction or prediction is selected automatically by comparing both types of prediction errors and keeping the smallest one.

Second, bidirectional processing therefore enables omitting the conventional proximal prediction error (i.e., residue signal) in favor of the speech signal, or any other signal that is clinically apposite. This turns proximal prediction into an option, which can be dropped by those users who do not wish to assign any clinical relevance to the residue signal, or who would like to analyze speech dysperiodicities in their entirety and not only those that are left after the speech signal has been filtered by means of a whitening filter.

Other objectives of the article are the following. Table I suggests that only few studies have compared connected speech to sustained vowels, or different connected linguistic units to one another. Additional objectives therefore have been investigating whether vocal dysperiodicities are indeed larger and more commonplace in connected speech than in sustained vowels, and whether sentences that comprise frequent voice onsets and offsets are noisier than sentences that comprise few. Finally, the proposal of Qi *et al.* and the one that is made here are compared via their association with a perceptual classification of steady vowel fragments.

II. METHODS

A. Corpora

A first corpus comprises 1-s-long steady fragments of the vowel [a] sustained by 89 normophonic or dysphonic speakers. The recording and perceptual evaluation of this corpus have been presented elsewhere [Schoentgen, Bensaid, and Bucella, 2000]. The rationale of the perceptual experiment, the selection of the experts, as well as the intraexpert and interexpert agreement, have been reported in that article. The degree of hoarseness has been established on a 5-point scale by a jury of five experts on the base of spectrograms [Yanagihara, 1967]. The purpose of this corpus is to enable establishing that signal-to-dysperiodicity ratios calculated via the method of Qi *et al.* and the method proposed here are correlated with perceived hoarseness.

A second corpus comprises sustained vowels [a], including onsets and offsets, as well as four French sentences produced by 22 normophonic or dysphonic speakers. Ten speakers are female and ten are male. Their chronological age has been in the range from 20 to 79 years. The corpus comprises, in addition, one girl aged 10 and one boy aged 14. Five speakers have normal voices, the other speakers have been diagnosed as follows: dysfunctional dysphonia (2), bilateral nodules (3), polyp on the left vocal fold (1), edema of the vocal folds (3), mutational disorder (1), dysphonia plicea ventricularis (1), unilateral vocal fold paralysis (6).

The sentences, spoken affirmatively, have identical grammatical structure (subject, verb, object). Sentences from S1 to S4 are the following: [ta tɑ̃t a apate yn kaʁp], [yn pul a pikoʁe tɔ̃ kəɪk], [bɔb mavɛ gide vɛʁ lɛ dig], [lœ gaʁd a ɑ̃dige labe]. The blanks designate word boundaries. In French orthography, the sentences are spelled as follows: ''*Ta tante a appâté une carpe*;'' ''*Une poule a picoré ton cake*;'' ''*Bob m'avait guidé vers les digues*;'' ''*Le garde a endigué l'abbe*.''

Phonologically speaking, the sentences have the same number of syllables (8) and roughly the same number of resonants (10 or 11) and plosives (between 5 and 7). Sentence S3 comprises in addition two voiced fricatives [v], which are lenis. Sentences S1 and S2 comprise phonologically voiced and unvoiced obstruents, whereas sentences S3 and S4 are voiced throughout. Hereafter, the former are designated as ''voiced and unvoiced'' and the latter as ''voiced'' sentences.

Vowels and sentences have been recorded with a sampling frequency of 48 kHz in an isolated booth by means of a digital audio tape recorder (Sony TCD D8) at the ENT department of a university hospital in Brussels, Belgium. Each token has been repeated three times. A microphone mounted on a headset (AKG C410WL) has been fixed at a distance of approximately 4 cm from the left corner of the mouth of the speaker. The recordings have been transferred later via a digital interface to a personal computer for further processing. The speech signals have been downsampled to 24 kHz, after low-pass finite-impulse-response filtering with a cutoff frequency equal to 10 kHz.

B. Analysis

1. Proximal linear predictive analysis

The purpose of this stage is to obtain the conventional near-sample residue signal, $e_P(n)$, which is a modeling error. Quantity $x(n)$ is the sampled speech signal. Symbol N is the order of the linear model and weights a_i are the coefficients of the linear prediction filter. These weights are computed recursively from reflection coefficients (or partial correlation coefficients) that are estimated via the covariance-lattice method using the harmonic-mean formulation

[Makhoul, 1977]. The filter order N is 24. The processing is carried out without overlap by means of a sliding rectangular window of 10 ms [Ramachandran and Kabal 1989]

$$e_P(n)=x(n)+\sum_{i=1}^{N} a_i x(n-i). \quad (1)$$

2. Forward distal linear predictive analysis

The purpose of this stage is to obtain the distal linear prediction error $e_D(n)$ that is ideally free of any cyclic correlation. When the processing is forward, symbol $e_D(n)$ in Eq. (2) is the error that is observed while modeling distally the present signal fragment by means of a past fragment. In the framework of Qi *et al.*'s proposal, symbol $y(n)$ is the residue signal, $e_P(n)$. In the framework of this study, symbol $y(n)$ is assigned to the speech signal, $x(n)$, or the residue signal, $e_P(n)$.

Weights b_i are the distal linear prediction coefficients. The calculation is carried out without overlap by means of a rectangular window, the length of which is equal to 2.5 ms. Parameter M is equal to 2, and symbol P is the prediction distance

$$e_D(n)=y(n)+\sum_{i=0}^{M} b_i y(n-P-i). \quad (2)$$

In practice, this model has been reformulated in the framework of a lattice structure according to Ramachandran and Kabal (1989). As in the proximal prediction case, the reflection coefficients are computed via the covariance-lattice method using the harmonic-mean formulation, referred to as Burg's formula by Ramachandran and Kabal (1989). The lattice coefficients are then transformed into direct form coefficients, enabling the transversal implementation of the filter.

Distance P is fixed as follows for each analysis window position. The correlation between the content of the analysis window and the content of a shifted window is calculated for the interwindow lag varying between 2.5 and 20 ms. For a given lag, the correlation coefficient is the scalar product of the lagged and unlagged windowed signals, divided by the geometric average of the windowed signal energies to compensate for drifts in the signal amplitude. Distance P is assigned to the lag for which the interwindow correlation is the highest.

In the framework of Qi *et al.*'s method, prediction distance P is found by an exhaustive search in the same interval. The proximal and distal linear predictive analyses are otherwise similar to the proposals of Qi *et al.* (1999) or Ramachandran and Kabal (1989).

3. Bidirectional distal linear predictive analysis

In the context of distal bidirectional linear predictive analysis of speech, the dysperiodicity cue is the error that is observed while modeling distally the present speech fragment by a past or future fragment, depending on which prediction error is the smallest. In practice, distal linear predictive analysis (2) is therefore carried out twice, once forward and once backward.

First, distal linear predictive analysis is applied to the input signal $y(n)$, as described earlier, and the outcome is the forward prediction error, which is stored for further processing.

Second, input signal $y(n)$ is time reversed, that is, the last sample of the recording interval becomes the first, and vice versa. Following that, a second distal linear predictive analysis is carried out, which is a backward prediction because of the time reversal. Care is taken to place the analysis window at the same positions as during the forward prediction stage. A second time reversal obtains the backward prediction error, which is stored for further processing.

Third, the last step consists of comparing the forward and backward prediction errors, analysis window by analysis window, and keeping the prediction error with the least energy. The least-energy prediction error is free of segment boundary artifacts when the phonetic segment is at least two cycles long. The least-energy distal prediction error is therefore a vocal dysperiodicity cue.

In this article, bidirectional distal linear predictive analysis is carried out on the near-sample residue signal, $e_P(n)$, as well as on the speech signal $x(n)$ directly. The purpose is to compare forward to bidirectional, and double to distal linear predictive processing on a same set of speech signals.

C. Signal-to-dysperiodicity ratio

A dysperiodicity marker is computed on the base of the distal prediction error. An increase of the vocal dysperiodicities causes a decrease of a signal-to-dysperiodicity ratio, S/D, that is defined as follows [Qi *et al.*, 1999]. Symbol N_p is the number of samples in the analysis interval. Symbol $e_D(n)$ is the distal prediction error, and $x(n)$ is the speech signal. The subtraction of constant 1 is a correction that takes into account that speech signal $x(n)$ is the noisy signal

$$\mathrm{S/D}=10\log\left[\frac{\Sigma_{n=1}^{N_p}x^2(n)}{\Sigma_{n=1}^{N_p}e_D^2(n)}-1\right]. \quad (3)$$

III. RESULTS

The objective of the Results section is to present the following. First, S/D ratios have been correlated with the degree of hoarseness for a corpus of sustained vowel fragments. The correlation is expected to be statistically significant. Second, bidirectional has been compared to forward analysis, and double to distal analysis. The S/D ratios are expected to be higher for the bidirectional and double-predictive analyses, respectively. Third, the S/D ratios have been compared for different utterances. Sentences have been compared to vowels, and sentences comprising frequent voice onsets and offsets have been compared to sentences comprising few.

To recap, the terminology is the following. A double analysis designates a near-sample, forward, linear predictive analysis of the speech signal that is followed by a distant-sample, linear predictive analysis of the residue signal. A distal analysis refers to a distal linear predictive analysis of the speech signal *per se*. The distal linear predictive analysis is performed unidirectionally or bidirectionally. A forward

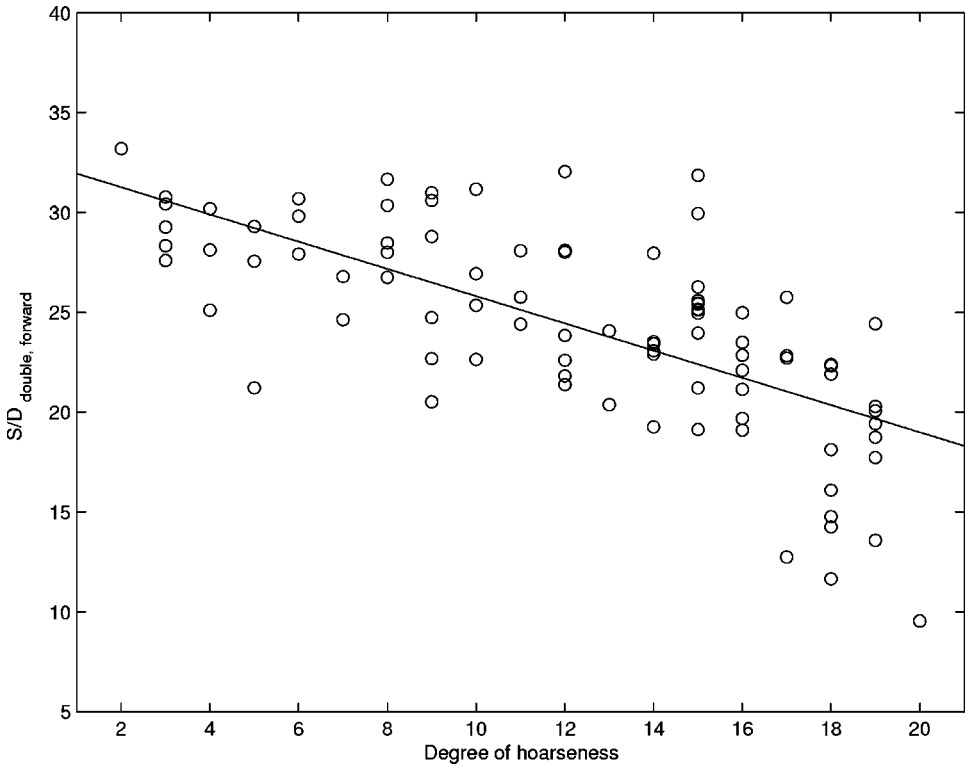


FIG. 1. $S/D_{double,forward}$ ratio, in dB, of stationary fragments of vowel [a] as a function of the perceived degree of hoarseness. The corpus comprises 89 male and female dysphonic and normophonic speakers. The Spearman rank correlation is equal to −0.71.

analysis designates a unidirectional left–right analysis. A bidirectional analysis refers to a left–right as well as right–left analysis, followed by a selection of the smallest prediction error.

Figures 1 and 2 show the S/D ratio of stationary fragments of vowel [a] (first corpus) as a function of the degree of hoarseness. The S/D ratios have been calculated by double-forward and distal bidirectional analyses of speech. The Spearman rank correlation, ρ_s, of S/D ratios and perceptual classification are −0.71 and −0.80, respectively. The null hypotheses ($\rho_S=0$) have been rejected (one-tailed test, $\rho_S>0.18$, $p<0.05$). This observation confirms results published by others [Qi *et al.*, 1999].

In Table II (top) are shown the quartiles of the average S/D ratios for the second corpus. The S/D ratios are averages because each speaker repeated each utterance several times. The corpus comprises sustained vowels [a] and four sentences uttered by 22 male and female normophonic and dys-

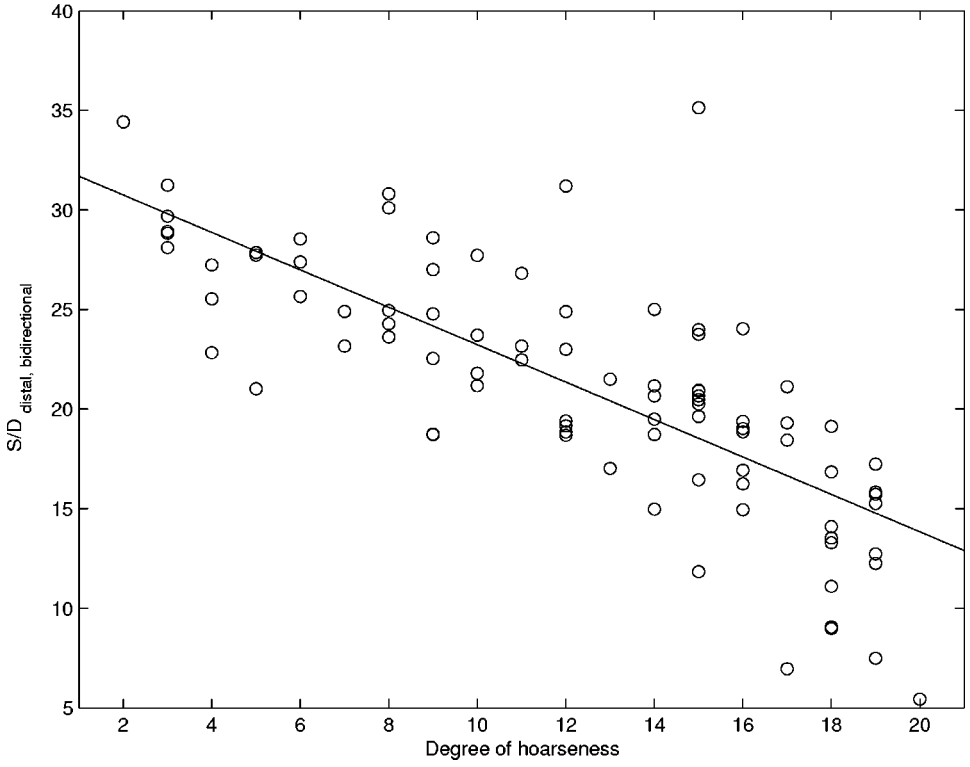


FIG. 2. $S/D_{distal,bidirectional}$ ratio, in dB, of stationary fragments of vowel [a] as a function of the perceived degree of hoarseness. The corpus comprises 89 male and female dysphonic and normophonic speakers. The Spearman rank correlation is equal to −0.80.

TABLE II. Top: Quartiles of S/D ratios, in dB, obtained by double and distal, forward and bidirectional predictive analyses of sustained vowel [a] and four French sentences S1–S4 uttered by 22 talkers. Bottom: Quartiles of arithmetic differences $\Delta = S/D_{bidirectional} - S/D_{forward}$.

	[a]				S1				S2				S3				S4			
	Double		Distal		Double		Distal		Double		Distal		Double		Distal		Double		Distal	
	fwd	bi	fwd	bi	fwd	bi	fwd	bi	fwd	bi	fwd	bi	fwd	bi	fwd	bi	fwd	bi	fwd	bi
Min	14.5	14.7	6.0	7.4	14.1	14.2	3.5	9.2	15.3	15.5	6.1	10.5	13.4	13.6	6.6	8.6	15.1	15.2	7.4	10.3
First quartile	25.7	26.0	18.7	20.9	24.5	24.8	10.4	15.4	26.3	26.6	11.6	19.1	26.7	27.0	9.5	18.0	25.5	25.9	13.9	19.1
Median	28.4	28.7	21.9	25.2	27.4	27.7	12.0	17.7	29.5	29.8	14.4	21.1	29.3	29.6	14.6	21.1	27.9	28.2	15.3	21.5
Third quartile	30.7	30.9	23.8	26.4	28.2	28.6	15.2	20.7	30.9	31.3	17.1	22.2	31.7	32.1	17.8	23.4	29.0	29.3	17.4	22.8
Max	39.2	39.6	26.5	31.1	34.5	35.1	18.7	23.2	36.3	36.7	22.2	25.8	36.9	37.4	22.1	26.5	34.9	35.3	21.4	25.1
Δ																				
Min	0.14		0.99		0.15		1.88		0.15		1.61		0.14		1.14		0.14		1.18	
First quartile	0.20		1.60		0.27		4.57		0.24		4.62		0.25		3.55		0.27		3.36	
Median	0.28		2.37		0.35		5.13		0.32		6.00		0.32		5.19		0.35		3.92	
Third quartile	0.38		3.27		0.44		6.31		0.39		6.84		0.40		7.19		0.44		6.30	
Max	0.62		9.19		0.60		8.99		0.47		10.18		0.56		9.23		0.53		8.36	

phonic speakers. The S/D ratios have been obtained by forward and bidirectional, as well as double and distal analyses of speech.

Discrepancies between processing methods are best understood by examining the raw data, which show that for each speaker and utterance the S/D ratios are ranked as follows:

$$S/D_{double,bidirectional} > S/D_{double,forward} > S/D_{distal,bidirectional} > S/D_{distal,forward}.$$

The same hierarchy has been obtained for a corpus of vowel–consonant–vowel fragments [Bettens, Grenez, and Schoentgen, 2003]. The bounding of $S/D_{forward}$ by $S/D_{bidirectional}$ is deterministic and a planned outcome of the signal processing.

A two-way analysis of variance has been carried out on the data summarized in Table II (top). The factors are the analysis (four methods) and utterance type (one vowel and four sentences). The number of subjects has been equal to 22. The null hypothesis (i.e., the mean S/D ratios are identical) has been rejected for the methods as well as utterances ($F = 166.2, p < 0.001; F = 6.6, p < 0.001$). The interaction between methods and utterances is not significant ($F = 1.6, p > 0.08$).

In Table II (bottom) are displayed the quartiles of the arithmetic differences $\Delta = S/D_{bidirectional} - S/D_{forward}$. For the distal analysis, these distances are typically in the range from 2.4 to 6 dB, depending on the utterance. That is, the squared distal errors typically differ by 75% to 300%. This suggests that the distal forward analysis of speech is an unacceptable option. The boosting of the distal predictive error by the forward analysis is indeed a consequence of the inability to predict distally the speech signal across phonetic boundaries, or at the beginning of the recording interval. Figure 3 illustrates this phenomenon by means of the fragment [bɔb̥] in sentence S3. The transients in the bottom trace near the onsets of [b] and [ɔ] are a consequence of the inability to predict cycle shapes across boundaries.

For the double analyses, the typical distances between $S/D_{double,bidirectional}$ and $S/D_{double,forward}$ are equal to 0.3 dB. That is, the squared distal predictive errors typically differ by 7%. With regard to the results, the double-forward analysis therefore appears to be a valid approximation of the double-bidirectional analysis that is methodologically superior. The rest of the Results section is therefore devoted to a comparison of the double-forward analysis that has been proposed by Qi *et al.* with the distal bidirectional analysis that is proposed in this article.

The ratios $S/D_{double,forward}$ and $S/D_{distal,bidirectional}$ have subsequently been reanalyzed by a two-way analysis of variance. Again, the factors are the analysis (two methods) and utterance type (one vowel and four sentences). The number of subjects has been equal to 22. The null hypothesis (i.e., the mean S/D ratios are equal) has been rejected for the methods ($F = 120.0, p < 0.001$) as well as the utterances, al-

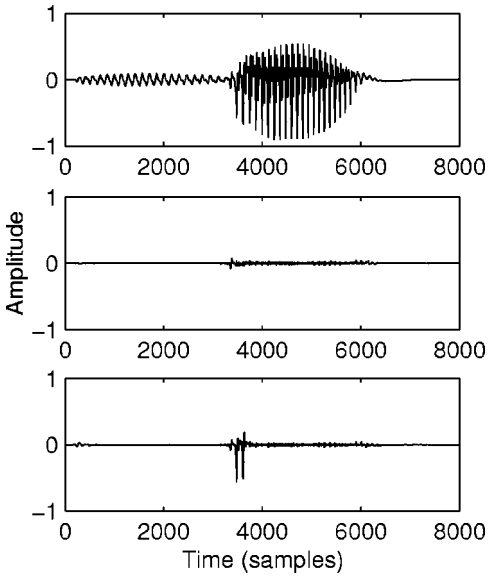


FIG. 3. Distal forward and bidirectional predictive analyses of the fragment [bɔb̥] in sentence S3 of a normophonic male speaker. From top to bottom: speech signal, distal bidirectional prediction error, and distal forward prediction error. The distal forward error displays spurious transients at the onsets of [b] and [ɔ]. The horizontal axis is time in number of samples and the vertical axis is the amplitude of the speech signal or predictive errors in arbitrary units.

TABLE III. Pearson product moment correlation of ratios $S/D_{double,forward}$ and $S/D_{distal,bidirectional}$ (on the main diagonal); Pearson product moment correlation of the utterances for ratio $S/D_{double,forward}$ (below the main diagonal) and of the utterances for ratio $S/D_{distal,bidirectional}$ (above the main diagonal).

	[a]	S1	S2	S3	S4
[a]	**0.95**	0.77	0.80	0.73	0.88
S1	0.96	**0.63**	0.90	0.87	0.89
S2	0.91	0.95	**0.81**	0.94	0.95
S3	0.94	0.96	0.95	**0.64**	0.89
S4	0.96	0.96	0.96	0.97	**0.85**

though barely so ($F=2.42, p<0.049$). The interaction between methods and utterances is not significant ($F=1.1, p>0.36$).

Subsequently, the interutterance differences have been examined separately for each analysis method. The number of subjects has been equal to 22. A one-way analysis of variance (one vowel and four sentences) confirms that the S/D ratios of the utterances do not differ statistically significantly for the double-forward analysis ($F=0.78, p>0.5$), but the null hypothesis is rejected for the distal bidirectional analysis ($F=2.8, p<0.032$). The rejection is due to vowel [a], because a one-way analysis of variance of the sentences alone shows that these do not differ statistically significantly ($F=1.4, p>0.2$).

In the framework of clinical applications, two analyses may be equivalent when one is able to predict the outcome of the other. The degree of association between the utterances as well as between the two methods has therefore been examined numerically by means of Pearson's product moment correlation, ρ_P, of the S/D ratios. Table III displays the intermethod correlation for the five utterances on the main diagonal, the interutterance correlation for the double-forward analysis below the main diagonal, and the interutterance correlation for the distal bidirectional analysis above the main diagonal. Data in Table III show the following.

(i) The null hypothesis ($\rho_P=0$) is rejected for all table entries (one-tailed test, $\rho_P>0.36$, $p<0.05$).
(ii) For a given utterance pair, the interutterance correlation is greater for double-forward analysis than for distal bidirectional analysis.
(iii) The correlation between the two analysis methods is not perfect, but statistically significant. The correlation is high for vowel [a] as well as sentences S2 and S4, and moderate for sentences S1 and S3.

A multiple regression analysis has been carried out to estimate the degree of association between sentential and vocalic S/D ratios. The multiple regression analysis has been carried out for each analysis method separately. The independent and dependent variables have been the sentential and vocalic S/D ratios, respectively. The correlation coefficients R are equal to 0.97 for the double-forward and 0.89 for the distal bidirectional processing. The null hypotheses ($R=0$) can be rejected ($F=89.5, p<0.01; F=17.1, p<0.01$).

IV. DISCUSSION

A. Forward versus bidirectional analysis

To recap, during distal forward processing of voiced speech, present speech cycles are predicted by past cycles. Because the cycle shapes depend on the identity of the phonetic segments, an exact prediction across phonetic boundaries is not possible. Inflated distal prediction errors therefore occur at the phonetic boundaries. Table II (bottom) indeed shows that the typical distances $S/D_{distal,bidirectional} - S/D_{distal,forward}$ are larger than 2.4 dB. The conclusion has been that distal forward processing of speech is not acceptable and that bidirectional processing is mandatory.

Bidirectional processing removes spurious prediction errors, because present cycles may be either predicted by past cycles or retrodicted by future cycles, whichever causes the smallest error. In practice, this means that the cycles within a phonetic segment are likely to be compared to other cycles within the same segment, and the prediction error can indeed be interpreted as a measure of the deviation from periodicity. Intuitively speaking, bidirectional distal analysis mimics numerically the behavior of a human observer who visually scans back and forth speech signal traces to estimate the amount of noise.

Predicting signal fragments across phonetic boundaries is also problematic for the double-forward analysis. In practice, the distances $S/D_{double,bidirectional} - S/D_{double,forward}$ are expected to be smaller, however. The reason is the following. The near-sample decorrelation owing to the proximal predictive analysis causes the shape of the residue signal to be similar to the left and right of the boundaries between phonetic segments when both are voiced or unvoiced. Double-forward cross-boundary prediction errors are therefore smaller and less frequent. Table II (bottom) confirms that distances $S/D_{double,bidirectional} - S/D_{double,forward}$ are typically equal to 0.3 dB, which is one order of magnitude smaller than for the distal analysis.

B. Double versus distal analysis

The proximal predictive processing stage is mandatory in the framework of a forward analysis because it decreases the effect on the S/D ratios of cross-boundary prediction errors, which are clinically irrelevant. But, in the framework of bidirectional analysis the proximal predictive processing can be omitted, because bidirectional processing removes the cross-boundary prediction errors from any signal. The following is therefore a discussion of the conventional proximal predictive processing in the framework of clinical processing of speech *per se*.

An analysis of variance of ratios $S/D_{double,forward}$ and $S/D_{distal,bidirectional}$ has shown that they differ statistically significantly. Also, they are imperfectly, but statistically significantly, correlated. Possible explanations of the observed discrepancies are the following. First, the residue signal is an artificial signal that reports the adequacy of a linear all-pole model for representing stationary speech signals. Earlier studies have shown that the size of the residue signal depends on properties of speech and speaker timbre that are not or not exclusively clinically relevant [Kahn and Garst, 1983].

As a consequence, markers that depend on the shape of the residue signal may mask differences between normophonic and dysphonic speakers when the markers are used clinically [Schoentgen, 1982].

Second, even if the modeling of the speech signal were equally efficient for all speakers and segments, perturbations of the residue signal would still agree neither with the perturbations of the speech signal nor with the perturbations of the glottal excitation signal. The reason is that the residue signal is obtained by decorrelating the speech signal, perturbations included. But, the whitening filter is not the inverse of the vocal tract transfer function. Filtering therefore does not recover the perturbations of the glottal excitation exactly.

The decorrelation *per se* as well as the speaker-dependent and segment-dependent efficiency of the decorrelation would therefore explain why the outcome of the distal processing of the residue signal cannot be expected to agree exactly with the outcome of a distal processing of the speech signal. The main diagonal of Table III shows that the association between speech and residue is highest for vowel [a]. This is expected because oral vowels are the segments for which conventional, proximal, predictive analysis works best.

The previous discussion would therefore suggest dropping the residue signal in favor of the speech signal, or any another signal whose clinical interpretation is less subject to debate. For completeness sake, one argument may be given, however, in favor of basing clinical analyses on the residue signal. At present, this argument is speculative. The ability of clinically analyzing connected speech as easily as stationary fragments of vowels raises questions regarding the speech material. One question relates to the validity of the pooling of multilingual utterances. Pooling multilingual items may be problematic as far as connected speech is concerned, because the grammatical and phonetic matching is expected to be imperfect under most circumstances. Prewhitening of the speech segments might ease the phonetic matching between productions that differ lingually, because decorrelating strips the segments of part of their phonetic identity.

C. Sustained vowels versus connected speech

It has been argued repeatedly that analyzing connected speech is potentially clinically more valid than analyzing sustained vowels [Qi *et al.*, 1999]. One reason is that the frequent voicing onsets and offsets in connected speech are thought to present a greater challenge to a dysphonic speaker. Other tasks that might cause voiced connected speech to be more difficult to produce are the voicing of obstruents, the maintaining of voicing while the larynx continually ascends and descends in the neck, as well as nonflat intonation and accentuation.

For the corpus that has been examined here, vocalic and sentential $S/D_{distal,bidirectional}$ ratios differ statistically significantly, whereas vocalic and sentential $S/D_{double,forward}$ ratios do not. Multiple correlation analyses show that vocalic and sentential S/D ratios are correlated, whatever the analysis method. The vocalic S/D variance that is explained by the sentential S/D variance indeed equals 80% for the distal bidirectional and 94% for the double-forward analysis.

A possible explanation for the observed correlation between vowels and sentences is related to the marker that has been used to sum up vocal dysperiodicities. The S/D ratio is a global marker, which means that each speech sample contributes to its numerical value. Therefore, phonetic segments that are long and intense (e.g., resonants, sibilants) contribute more than segments that are short or weak (e.g., nonstrident obstruents). Equally, the effects of transients that are isolated and few are numerically masked by fragments that are long and pseudostationary. The observed correlation between vocalic and sentential S/D ratios follows from there, because the vocalic segments in connected speech determine the numerical value of the S/D ratio, provided that long strident fricatives [s] and [ʃ] are omitted from the corpus.

This explanation may generalize to other acoustic markers, or even to perceived degrees of hoarseness. Previous studies have indeed shown that connected speech does not necessarily enable discriminating between normophonic and dysphonic speakers any better than sustained vowels that include onsets and offsets [de Krom, 1994; Parsa *et al.*, 2001].

D. Voiced versus voiced and unvoiced speech

The S/D ratios obtained for voiced sentences S3 and S4 as well as voiced and unvoiced sentences S1 and S2 do not differ statistically significantly. The off-diagonal entries in Table III show that, in addition, the corresponding S/D ratios are correlated. Possible explanations are the following. First, voiced sentences S3 and S4 are phonologically voiced. Devoicing may have occurred for phonetic, behavioral, as well as pathological reasons. Even so, voiced–unvoiced transients are expected to be more frequent in voiced and unvoiced sentences than in voiced ones.

Second, the most likely explanation for the lack of statistically significant differences is therefore the one that has been discussed in the previous section. Vocalic segments that are long, intense, and pseudosteady determine the values of the S/D ratio in sibilant-free sentences. Vocalic segments are likely to contribute equally to the S/D ratios in voiced sentences S3 and S4 as well as voiced and unvoiced sentences S1 and S2 because all have been matched for grammatical and prosodic structure, number of syllables, and resonant/obstruent ratio. Abnormally noisy voice onsets or offsets, assuming that they exist, would be too short or infrequent to decrease noticeably the values of the S/D ratios of the voiced and unvoiced sentences. The observed lack of significant differences between sentential S/D ratios follows from there. This suggests that clinically exploiting what is specific to connected speech would imply using markers that enable focusing on transients as well as pseudostationary fragments.

E. Estimation of the distant-sample linear predictive filter coefficients

Here, the values of the distant-sample linear predictive filter coefficients [Eq. (2)] have been obtained via a method recommended by Ramachandran and Kabal (1989) that would guarantee the stability of the filter if it was used for synthesis, that is, for the recovery of the speech signal from the residue signal. But, when the filter is used for analysis,

that is, for the recovery of the residue signal from the speech signal, stability is of minor importance and can be dropped. Then, other estimation methods that are simpler or more flexible may be employed. Indeed, the covariance-lattice method that has been used here leads to a formulation of the distal predictive filter that mixes distant and lagged near samples, obscuring thus the conceptual simplicity of relation (2).

In a subsequent study, we have therefore implemented the estimation method that has been used here as well as the conventional covariance method that does not suffer from the same drawback, and compared the performance of both methods on the first corpus that has been presented above. The Spearman rank correlation coefficients have been equal to -0.80 and -0.83, respectively. The correlation coefficients are not statistically significantly different. The lack of statistically significant differences owing to different coefficient estimation methods has been confirmed for other corpora [Dessalle, 2004].

V. CONCLUSION

First, distal bidirectional linear predictive analysis of disordered speech mimics the behavior of a human expert who visually scans speech signal traces to gauge the lack of periodicity. The method compares speech fragments segment internally and avoids the boosting, owing to cycle omission or insertion errors, of the values of the vocal noise markers. This ease of performance allows connected speech to be analyzed that is produced by speakers who may be severely hoarse.

Second, all types of vocal noise (e.g., additive noise, vocal jitter, vocal shimmy, etc.) contribute to the dysperiodicity trace. At present, the separation and independent analysis of additive and modulation noise require techniques that are based on an explicit knowledge of the positions of the vocal cycles or spectral harmonics [Kruse, 2003].

Third, double-forward analysis enables estimating the degree of aperiodicity of an auxiliary signal, which is the conventional linear predictive error [Qi *et al.*, 1999]. The clinical relevance of this modeling error has been subject to debate. Its deviations from periodicity do not agree with the deviations from periodicity of the speech or glottal excitation signal. Also, the analysis that is exclusively forward inflates the size of the actual perturbations. Results show that this boosting is moderate and that, with regard to the results, the double-forward analysis is an approximation of the double-bidirectional analysis, which is methodologically superior.

Fourth, the signal-to-dysperiodicity ratios computed for vowels and connected speech are highly correlated. The explanation is that vocalic segments determine the values of the S/D ratio of connected speech. For the same speaker, perturbations of vocalic segments in connected speech are expected to vary in accordance with the perturbations of sustained vowels.

Fifth, the signal-to-dysperiodicity ratios computed for speech with frequent or infrequent voice onsets and offsets are highly correlated and do not differ statistically significantly. The explanation is that vocalic segments rather than the transients between obstruents and resonants determine the values of the S/D ratio. The above suggests that demonstrating that analyzing connected speech is clinically more valid than analyzing sustained vowels would require investigating markers that enable focusing on events that are short, exceptional, or transient.

ACKNOWLEDGMENTS

The authors would like to thank Mrs. C. Hansotte and M. Y. Debéthune of the ENT Department of the Erasmus University Hospital, Brussels, as well as Dr. R. De Gucht-eneere for the audio recordings. Part of this work has been supported by a grant from the *Université Paris III*, Paris, France, awarded to Dr L. Crevier-Buchman of the *Laboratoire de Recherche en Phonétique et Phonologie* [UMR 7018] (Professor J. Vaissière). This research was completed as part of the first author's doctoral dissertation.

Abberton, E., and Fourcin, A. (**1984**). ''Electrolaryngography,'' in *Experimental Clinical Phonetics*, edited by C. Code and M. Ball (Croom Helm, London and Canberra), pp. 62–78.

Bettens, F., Grenez, F., and Schoentgen, J. (**2003**). ''Isolating vocal noise in running speech via bidirectional double linear prediction analysis,'' in Proceedings of the 6th International Seminar on Speech Production, Sydney, Australia, 13–18.

de Krom, G. (**1994**). ''Consistency and reliability of voice quality ratings for different types of speech fragments,'' J. Speech Hear. Res. **37**, 985–1000.

Dessalle, E. (**2004**). ''Estimation en ligne des dispériodicités vocales dans la parole connectée'' (''Real-time estimation of vocal dysperiodicities in connected speech''), unpublished Master thesis, Université Libre de Bruxelles, Brussels, Belgium.

Dolansky, L., and Tjernlund, P. (**1968**). ''On certain irregularities of voiced-speech waveforms,'' IEEE Trans. Audio Electroacoust. **16**(1), 51–56.

Gubrynowicz, R., Mikiel, W., and Zarnecki, P. (**1980**). ''An acoustic method for the evaluation of the state of the larynx source in cases involving pathological changes in the vocal folds,'' Arch. Acoust. **5**(1), 3–30.

Hecker, M. H. L., and Kreul, E. J. (**1971**). ''Description of the speech of patients with cancer of the vocal folds. I. Measures of fundamental frequency,'' J. Acoust. Soc. Am. **49**(4), 1275–1282.

Hillenbrand, J., and Houde, R. A. (**1996**). ''Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech,'' J. Speech Hear. Res. **39**, 311–321.

Kahn, M., and Garst, P. (**1983**). ''The effects of five voice characteristics on LPC quality,'' in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Boston, 531–534.

Klingholz, F. (**1990**). ''Acoustic recognition of voice disorders: A comparative study of running speech versus sustained vowels,'' J. Acoust. Soc. Am. **87**(5), 2218–2224.

Kruse, E. (**2003**). ''The Göttinger hoarseness diagram,'' Workshop WK1/18, Pan European Voice Conference, Graz, Austria.

Laver, J., Hiller, S., Mackenzie, J., and Rooney, E. (**1986**). ''An acoustic screening system for the detection of laryngeal pathology,'' J. Phonetics **14**, 517–524.

Lebrun, Y., and Hasquin, J. (**1971**). ''Variations in vocal wave duration,'' J. Laryngol. Otol. **LXXXV**, 1, 43–56.

Lieberman, P. (**1963**). ''Some acoustic measures of the fundamental periodicity of normal and pathologic larynges,'' J. Acoust. Soc. Am. **35**(3), 344–353.

Makhoul, J. (**1977**). ''Stable and efficient lattice methods for linear prediction,'' IEEE Trans. Acoust., Speech, Signal Process. **25**(5), 423–428.

Markel, J. D., and Gray, A. H. (**1974**). *Linear Prediction of Speech* (Springer, New York).

Muta, H., Baer, T., Wagatsuma, K., Muraoka, T., and Fukuda, H. (**1988**). ''A pitch-synchronous analysis of hoarseness in running speech,'' J. Acoust. Soc. Am. **84**(4), 1292–1301.

Parsa, V., and Jamieson, D. G. (**2001**). ''Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech,'' J. Speech Hear. Res. **44**, 327–339.

Qi, Y., Hillman, R. E., and Milstein, C. (**1999**). ''The estimation of signal-to-noise ratio in continuous speech for disordered voices,'' J. Acoust. Soc. Am. **105**(4), 2532–2535.

Ramachandran, R., and Kabal, P. (**1989**). ''Pitch prediction filters in speech coding,'' IEEE Trans. Acoust., Speech, Signal Process. **37**(4), 467–478.
Schoentgen, J. (**1982**). ''Quantitative evaluation of the discrimination performance of acoustic features in detecting laryngeal pathology,'' Speech Commun. **1**, 269–282.
Schoentgen, J., Bensaid, M., and Bucella, F. (**2000**). ''Multivariate statistical analysis of flat vowel spectra with a view to characterizing dysphonic voices,'' J. Speech Lang. Hear. Res. **43**, 1493–1508.
Shipp, T., and Huntington, D. A. (**1965**). ''Some acoustic and perceptual factors in acute-laryngitic hoarseness,'' J. Speech Hear Disord. **30**(4), 350–359.
Valencia Naranjo, N., Mendoza Lara, E., Mateo Rodriguez, I., and Carballo Garcia, G. (**1994**). ''High-frequency components of normal and dysphonic voices,'' J. Voice **8**(2), 157–162.
Yanagihara, N. (**1967**). ''Significance of harmonic changes and noise components in hoarseness,'' J. Speech Hear. Res. **10**, 531–541.

Development of vocal tract length during early childhood: A magnetic resonance imaging study[a)]

Houri K. Vorperian[b)]
Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue #481, Madison, Wisconsin 53705

Ray D. Kent
Department of Communicative Disorders and Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue #435, Madison, Wisconsin 53705

Mary J. Lindstrom
Waisman Center and Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, 600 Highland Avenue, Room K6/446, Madison, Wisconsin 53792

Cliff M. Kalina
Department of Communicative Disorders, University of Wisconsin-Madison, 1975 Willow Drive, Madison, Wisconsin 53706

Lindell R. Gentry
Department of Radiology, University of Wisconsin Hospital and Clinics, 600 Highland Avenue, E1-311 Clinical Science Center, Madison, Wisconsin 53792

Brian S. Yandell
Departments of Statistics and Horticulture, University of Wisconsin-Madison, 1300 University Avenue, Medical Sciences Center—Room 1239, Madison, Wisconsin 53706

(Received 9 July 2002; revised 28 September 2004; accepted 25 October 2004)

Speech development in children is predicated partly on the growth and anatomic restructuring of the vocal tract. This study examines the growth pattern of the various hard and soft tissue vocal tract structures as visualized by magnetic resonance imaging (MRI), and assesses their relational growth with vocal tract length (VTL). Measurements on lip thickness, hard- and soft-palate length, tongue length, naso-oro-pharyngeal length, mandibular length and depth, and distance of the hyoid bone and larynx from the posterior nasal spine were used from 63 pediatric cases (ages birth to 6 years and 9 months) and 12 adults. Results indicate (a) ongoing growth of all oral and pharyngeal vocal tract structures with no sexual dimorphism, and a period of accelerated growth between birth and 18 months; (b) vocal tract structure's region (oral/anterior versus pharyngeal/posterior) and orientation (horizontal versus vertical) determine its growth pattern; and (c) the relational growth of the different structures with VTL changes with development—while the increase in VTL throughout development is predominantly due to growth of pharyngeal/posterior structures, VTL is also substantially affected by the growth of oral/anterior structures during the first 18 months of life. Findings provide normative data that can be used for modeling the development of the vocal tract.
 [DOI: 10.1121/1.1835958]

PACS numbers: 43.70.Aj, 43.70.Ep [AL] Pages: 338–350

I. INTRODUCTION

As the infant vocal tract increases more than twofold in length from infancy to adulthood, its geometric proportions also change, so the term ''anatomic restructuring'' has been used to denote the physical changes that the vocal tract structures undergo during development. Reported changes include the bending of the vocal tract to form a right angle in the nasopharyngeal region; the disengagement of the velic–epiglottic contact; the descent of the larynx, the hyoid, and the epiglottis; as well as the descent of the posterior part of the tongue to form the anterior wall of the pharyx. See Fig. 1. (Bosma, 1975a, b, 1976, 1985; Crelin, 1976, 1973; Fried, Kelly, and Strome, 1982; Kent, 1981; Kent and Vorperian, 1995; Laitman and Crelin, 1976; Lieberman, 1977, 1984; Sasaki *et al.*, 1977; Westhorpe, 1987). The implication of such anatomic reorganization is that the various bony and soft-tissue structures in the oral and pharyngeal regions have different growth rates or growth patterns such that adult sizes are reached anywhere from age 7 to 18 years (Kent and Vorperian, 1995; Fitch and Giedd, 1999; Vorperian *et al.*, 1999; Lieberman *et al.*, 2001; Vorperian, 2000). Using advances in imaging technology, such as magnetic resonance imaging (MRI), the goal of our research is to quantitatively characterize the macroanatomic developmental changes in the bony and soft-tissue structures of the vocal tract during the first 2 decades of life. Our ultimate goal is to describe the

[a)]Portions of this paper were presented in 2001 at the 141st meeting of the Acoustical Society of America in Chicago, also at the 1999 annual convention of the American Speech-Language Hearing Association in San Francisco, CA.

[b)]Electronic mail: vorperian@waisman.wisc.edu

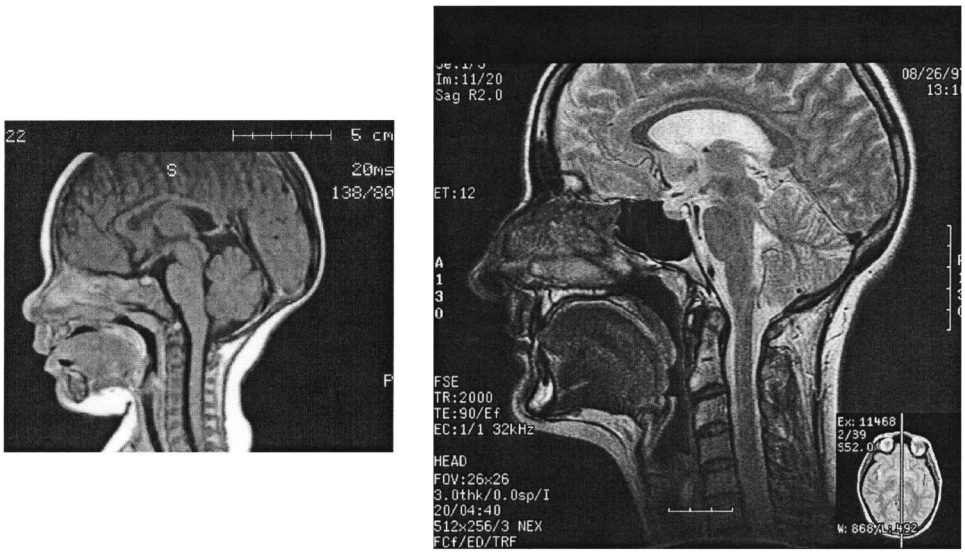


FIG. 1. Midsagittal magnetic resonance images of a 7-month-old female (left) and an adult female (right).

relational growth of those structures for the purpose of understanding the biological foundation of speech. Speech emergence and development is presumed to be dependent, at least in part, on the physical changes that the vocal tract structures undergo during development (Bosma, 1975a; Kent, 1976, 1981, 1992; Thelen, 1991). Also, the nature of the speech sounds that can be produced is determined in part by the physical constraints of the supralaryngeal system or the articulatory–resonatory system (e.g., Mowrer, 1980; Smith and Oller, 1981). This paper contains findings on the development of the various hard-and soft-tissue vocal tract structures and vocal tract length during the first 6 years of life as compared to the mature adult vocal tract. It also includes findings on developmental changes in the relational growth of those various vocal tract structures with vocal tract length. Relational growth is defined as the percent variation in vocal tract length explained by each of those various hard- and soft-tissue structures that affect vocal tract length.

Vocal tract length, defined as the curvilinear distance along the midline of the tract starting at the glottis to the intersection with a line drawn tangentially to the lips, has been estimated to increase from approximately 6 to 8 cm in infants to about 15 to 18 cm in adult females and males, respectively. Such estimates of vocal tract length have been calculated from acoustic studies (Fant, 1960) and from direct measurements of the adult vocal tract from radiographic images (Fant, 1960). More recently, Fitch and Giedd (1999), and Vorperian *et al.* (1999) reported vocal tract length measurements using MRI. Fitch and Giedd's data from subjects between the ages 2 and 25 years indicate significant length increases in all the different portions in the vocal tract. They report growth in the pharyngeal region to be prominent between early childhood and puberty, and even more pronounced between puberty and adulthood. Such findings support the hypothesis that an increase in vocal tract length is predominantly due to growth in the pharyngeal region (Kent and Vorperian, 1995; Vorperian, 2000). In this paper, rather than segmenting vocal tract length into different regions—such as oral and pharyngeal—the development of vocal tract length is examined in children as compared to the mature vocal tract in the adult by first independently examining all the oral and pharyngeal vocal tract structures in the sagittal plane, and then assessing the relational growth of each of those structures with vocal tract length. This latter assessment was done while taking into consideration the assigned region (oral versus pharyngeal versus combined) for each vocal tract structure. Structures examined, in the sagittal plane, include maxillary and mandibular lip thickness, hard and soft palate length, tongue length, mandibular length and depth, and the distance of the larynx, the epiglottis, and the hyoid bone from the posterior nasal spine (PNS). Such an approach was used to address the following objectives: (i) Determine the growth pattern that the various vocal tract structures follow. (ii) Determine the age or age range of accelerated growth periods, if such periods exist. (iii) Determine if there are gender differences in growth rate during the first 6 years of life. (iv) Examine differences in growth rate of the vocal tract structures based on region (anterior versus posterior) and orientation of plane of growth (horizontal versus vertical). Since the vocal tract is housed in the head and neck region, it is expected that the various vocal tract structures will follow the growth pattern and the plane of growth that the head and face follow (Farkas *et al.*, 1992a, b). (v) Assess the developmental changes in the relational growth of each vocal tract structure with vocal tract length.

II. METHODS

A. Subjects

MR images were secured from 37 white patients/subjects who received MRI for medical reasons known not to affect growth and development. The 37 subjects include 12 adult subjects (6 male, 6 female), and 25 pediatric subjects (16 male and 9 female). All adults and 9 pediatric subjects (6 male, 3 female) were imaged once, and the remaining 16 pediatric subjects (10 male, 6 female) were imaged two or more times. In this paper each MR imaging date is referred to as a case. The data reported are from 79 cases that include the 12 adult MRI cases, 9 pediatric MRI cases (6 male, 3 female) from the subjects imaged once, and 58 MRI cases

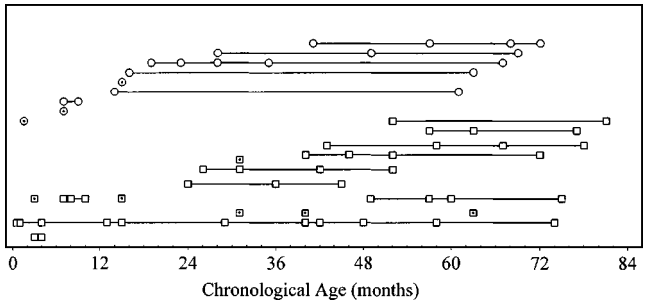


FIG. 2. Distribution of pediatric cases across chronological age in months. Square symbols denote male cases; round symbols, female cases. Repeat MRI cases are connected with a line. Nonrepeat cases have a dot in the symbol. Note that the frequency of repeat MRI cases does not exceed 2 weeks during the first month of life, and 1 month during the first year of life.

(40 male, 18 female) from the 16 pediatric subjects who received serial/repeat MRI. The age range of the pediatric cases is 2 weeks (0;.5) to 6 years 9 months (6;9). Figure 2 displays the distribution of the pediatric cases, and indicates that the frequency of repeat MRI cases does not exceed 2 weeks during the first month of life, and 1 month during the first year of life. The weights of the majority of the cases were at the 50th percentile reference growth curves for boys and girls, with all cases falling between the 25th to 95th percentile growth curves as per the National Center for Health Statistics growth charts (1976). No participants were significantly under- or overweight. The mean age of the 12 adult subjects was 23 years with mean height 166 and 171 cm; and mean weight 56 and 68 kg for females and males, respectively.

B. Procedures

Image acquisition methods have been described previously (Vorperian *et al.*, 1999). Basically, image acquisition involves two phases. The first phase is the actual MRI study, where virtually all of the pediatric patients were sedated using either chloral hydrate 50 mg/kg administered orally, or DPT (demerol, phenergan, and thorazine) 1 mg/kg administered intramuscularly, prior to entering the scanner. Once in the scanner, the facial structures of all subjects were placed centrally in the head coil using the laser lights of the MR imager (GE scanner or Resonex). All images were acquired during quiet respiration. MRI image acquisition parameters were as follows: The imaging matrix was 256×256 or 256 ×192 or 512×256. All images were obtained using a spin–echo pulse sequence or a fast spin–echo pulse sequence. Sagittal slices were obtained with T1-weighted sequences (repetition time [TR]=350 to 700 ms, echo delay time [TE]=14 to 30 ms) as well as T2-weighted sequences. The second phase of image acquisition involved digitizing the MR images by scanning all sagittal slices in the oral region into the computer using the UMAX Powerlook 2100XL oversized scanner with a transparency adapter, and storing the images for subsequent measurements. The DPI (dots per inch) setting was set to 600. When available, images with contrast injection, i.e., images acquired after intravenous administration of a contrast medium (Gadolinium-DTPA) to increase visualization of body tissues, were selected over images with no contrast injection.

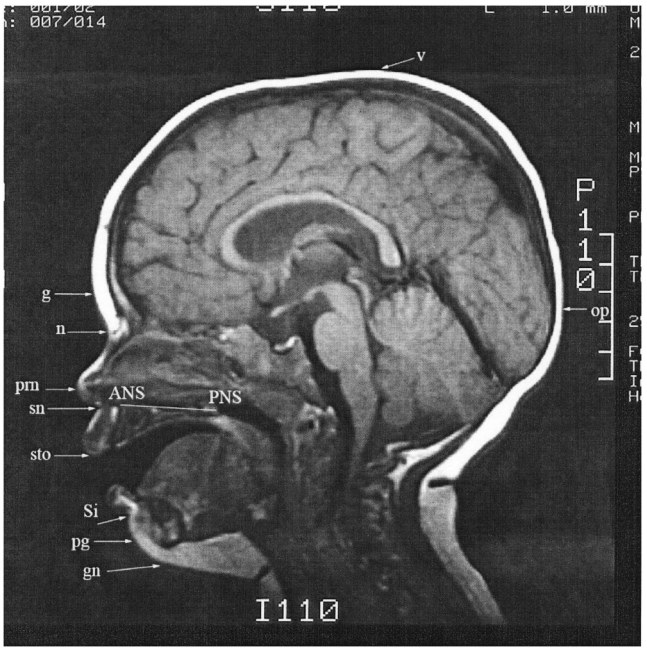


FIG. 3. Midsagittal magnetic resonance image of a 1-year, 3-month-old male subject with anatomic landmarks used for making measurements. ANS-PNS=anterior nasal spine–posterior nasal spine; g=glabella; gn =gnathion; n=nasion; op=opisthocranion; pg=pogonion; prn=pronasale; sto=stomion; sn=subnasale; Si=sulcus inferior or supramentale; v=vertex.

Data were acquired by measuring the following: *Maxillary and mandibular lip thickness, hard and soft palate length, naso-oro-pharyngeal length, laryngeal level, hyoid level, tongue length, mandibular length, mandibular depth*, and *vocal tract length*. Most measurements were made from the midsagittal plane, where distinct cerebral sulci extending to the corpus callosum are visible; also visible is the fourth ventricle, the full length of the cerebral aqueduct of Sylvius, the pituitary gland, part of the optic chiasm, the brainstem, and the cervical cord (Shorten *et al.*, 1994). In addition, the most distal parasagittal slices, where the condylar process of the mandible can be visualized, were used to calculate mandibular length and depth measurements. The image measurement software SIGMASCAN PRO by SPSS (formerly Jandel Scientific) was calibrated for each case/slice using the hash-scale mark on the MR image/slice. The measurement set for this study is described below. Refer to Fig. 3 for anatomic landmarks (Farkas, Posnick, and Hreczko, 1992a). Although most anatomic landmarks are evident in the midsagittal slice, sagittal slices immediately to the right and left of the midsagittal slice were used to ensure the accurate placement of anatomic landmarks to make measurements, particularly for structures that could deviate from midline such as the epiglottis and the uvula. The definition for each measurement follows: *Vocal tract Length*: The curvilinear distance along the midline of the tract starting at the thyroid notch to the intersection with a line drawn tangentially to the lips. The thyroid notch, which is slightly superior (about 5 mm in adults) to the anterior commissure—the junction of the vocal folds anteriorly in the larynx—was used as one of the two

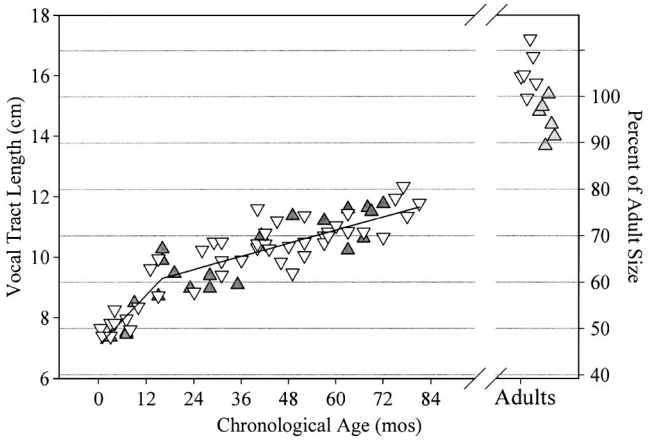


FIG. 4. Vocal tract length of the pediatric and the adult cases (open triangle down for males, and shaded triangle up for females). The second *Y* axis reflects the percent of adult size. Vocal tract length is defined as the curvilinear distance along the midline of the tract starting at the thyroid notch to the intersection with a line drawn tangentially to the lips.

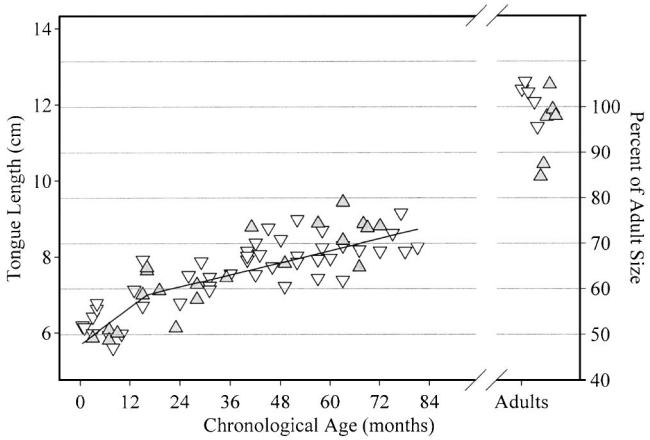


FIG. 5. Tongue length of the pediatric and the adult cases (open triangle down for males, and shaded triangle up for females). Tongue length is defined as the curvilinear distance along the dorsal superior contour of the tongue from the tongue tip to the valleculae.

end points of the vocal tract length measure instead of the conventional use of the glottis to ensure the visibility of this end point on midsagittal MRI. *Hard palate length*: The curvilinear distance along the hard palate contour from the anterior point of the incisor or tooth bud to the beginning of the soft palate, which is marked by the presence of increased fat and the beginning of curvature. *Soft palate length*: The curvilinear distance from the posterior edge of the hard palate to the inferior edge of the uvula. *Mandibular length and depth*: The horizontal and vertical distances in the midsagittal plane from the mental protuberance to the orthogonal projection of the condylar process on the midsagittal plane. *Tongue length*: The curvilinear distance along the dorsal superior contour of the tongue from the tongue tip to the valleculae. *Hyoid bone level or tongue level*: The vertical distance from the PNS (posterior nasal spine) to the level of the anterior inferior point of the hyoid bone. *Laryngeal level*: The vertical distance of a line drawn from the thyroid notch cartilage to the PNS (posterior nasal spine). *Naso-oro-pharyngeal length*: The curvilinear distance along the posterior pharyngeal wall above the soft palate extending from the posterior naris to the end of the upper airway. A line drawn horizontally from the superior border of the hyoid bone to the posterior pharyngeal wall was taken as the dividing line between the nasopharynx and the oropharynx. *Maxillary lip thickness*: The anteroposterior distance from the subnasale (sn) to the anterior nasal spine (ANS). *Mandibular lip thickness*: The horizontal anteroposterior distance from the sulcus inferior (Si) or supramentale to the hard tissue line.

C. Statistical analysis

Measurement error includes a small calibration error and an error due to measurement. The calibration error, which occurs during software calibration while assigning the number of pixels per 1-cm hash marks, is in the range of 0.1 to 0.4 mm. This error was unavoidable since the different MR images had different specification with respect to matrix size and FOV (field of view). To compute the standard deviation (σ) of error due to measurement, duplicate measurements were made on 22 different hard- and soft-tissue structures, from nine cases at three different ages (15, 40, and 63 months). Duplicate measurements were made at two independent times with an interval of 1 month. The standard deviation of measurement error was independent of age ($\sigma = 0.14$ cm; $n = 53$ at age 15 mos; $\sigma = 0.18$ cm; $n = 52$ at age 40 mos; $\sigma = 0.16$ cm; $n = 48$ at age 63 mos). Thus, the accuracy of visualizing the various soft and particularly bony structures on MRI does not change during the course of development since the measurement error is approximately the same at the different ages examined. For all subjects, the standard deviation of measurement error for hard-tissue structures ($\sigma = 0.22$ cm; $n = 79$) was larger than for soft-tissue structure ($\sigma = 0.12$ cm; $n = 74$). This was expected since it is difficult to determine the exact borders of bony structures in the oral-pharyngeal region on MRI (Katzberg and Westersson, 1991). Also, since the absolute size of the measurements varies (e.g., large values for vocal tract length and small values for lip width), the percent measurement difference (time 1 versus time 2) of the length of the struc-

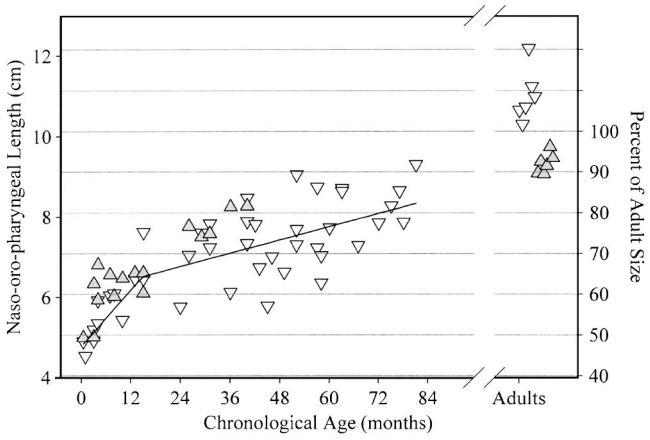


FIG. 6. Pharyngeal length of the pediatric and the adult cases (open triangle down for males, and shaded triangle up for females). Naso-oro-pharyngeal length is defined as the curvilinear distance along the posterior pharyngeal wall above the soft palate extending from the posterior naris to the level of the thyroid cartilage or the end of the upper airway.

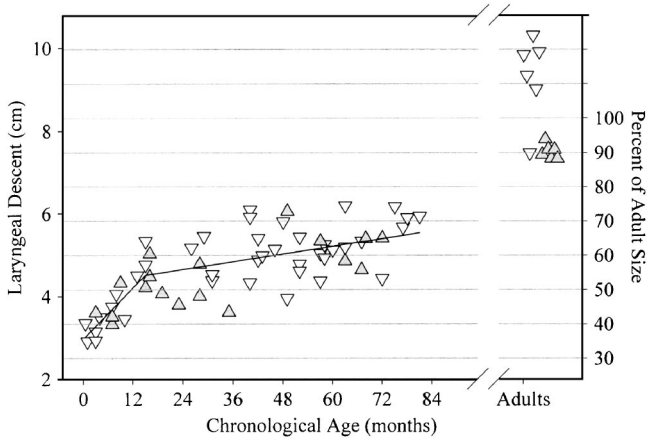


FIG. 7. Laryngeal descent of the pediatric and the adult cases (open triangle down for males, and shaded up triangle for females). Laryngeal descent is defined as the vertical distance of a line drawn from the thyroid notch to the PNS (posterior nasal spine).

tures remeasured, was examined individually for the nine cases. The maximum percent deviation between measurements was as follows: No more than 3% for vocal tract length, less than 5% for pharyngeal length and tongue length, less than 6.5% for total length for hard and soft palates, and less than 8% for mandibular length and depth.

1. Anatomic growth of the vocal tract structures

The absolute growth (male versus female measurements) of the different hard- and soft-tissue vocal tract structures during the first 6 years of life and in adults is graphically presented in Figs. 4 to 11. The broken line growth curve model was used to fit the data because it allows a test for a change in the rate of growth, and provides an estimate of the time of the change or the ''breakpoint'' (Bacon and Watts, 1971). The broken line consists of two straight lines joined at a point. It has the form $a+b$ (ca.mos) for ca.mos $\leqslant p$ and $d+c$ (ca.mos) for ca.mos $>p$, where $d=(a-b)/p+c$ is determined by the restriction that the two lines must meet at ca.mos$=p$ (ca.mos stands for chronological age in months). The intercept parameter (a) is assumed random in the population of subjects and accounts for within-subject correlation. The S-plus NLME software was used for estimating the parameters in this nonlinear mixed effects model. These data were fit with a broken line growth curve model both with and without the inclusion of terms for gender effects. The Wald F-test was used to test for the effect of gender (Pinheiro and Bates, 2000).

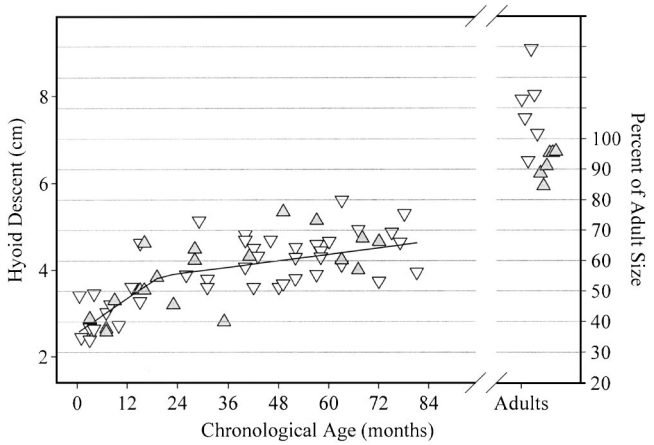


FIG. 8. Hyoid descent of the pediatric and the adult cases (open triangle down for males, and shaded triangle up for females). Hyoid bone level or tongue level is defined as the vertical distance from the PNS to the level of the antero-inferior point of the hyoid bone.

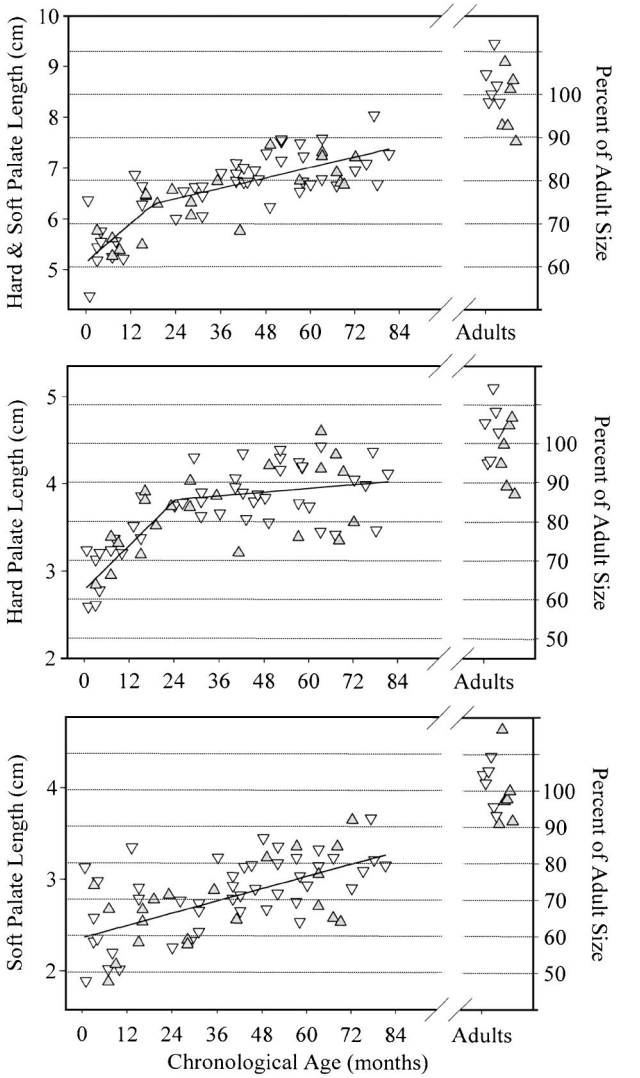


FIG. 9. Hard and soft palate length (top figure), hard palate length (middle figure), and soft palate length (bottom figure) development for pediatric and adult cases (open triangle down for males, and shaded triangle up for adult females. Hard palate length is defined as the curvilinear distance along the hard palate contour from the anterior point of the incisor or tooth bud to the beginning of the soft palate, which is marked by the presence of increased fat and the beginning of curvature. Soft palate length is defined as the curvilinear distance from the posterior edge of the hard palate to the inferior edge of the uvula.

2. Growth of vocal tract structures based on region/orientation

Using the angular bend of the vocal tract, the measurements of the various hard- and soft-tissue vocal tract structures were assigned to one of the following three regions/orientations: (i) Anterior or oral structures that grow in the horizontal plane [i.e., run parallel to head length—the distance from the glabella (g) to the opisthocarnion (op)]. Mea-

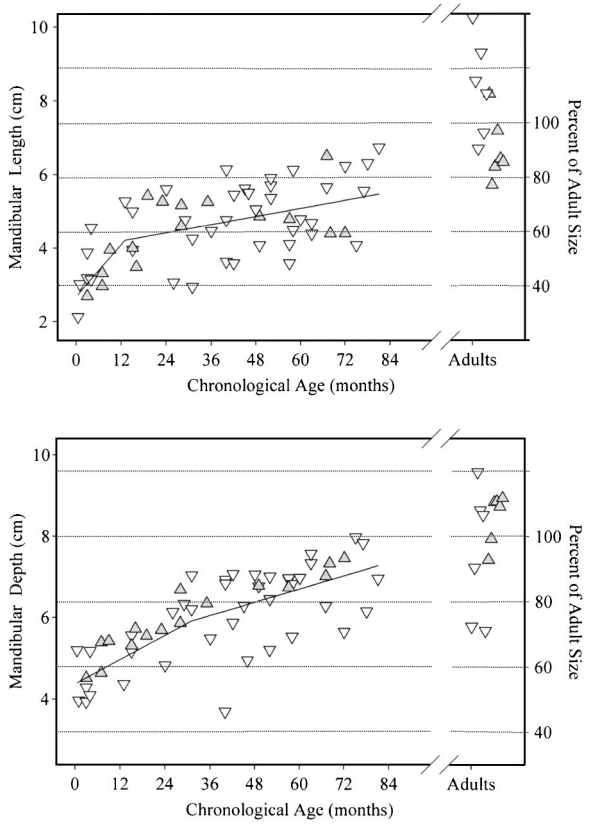


FIG. 10. Mandibular length (top figure) and mandibular depth (bottom figure) development for pediatric cases and adult cases (open triangle down for males, and shaded triangle up for females). Mandibular length and depth are defined as the horizontal and vertical distances in the midsagittal plane from the mental protuberance to the orthogonal projection of the condylar process on the midsagittal plane.

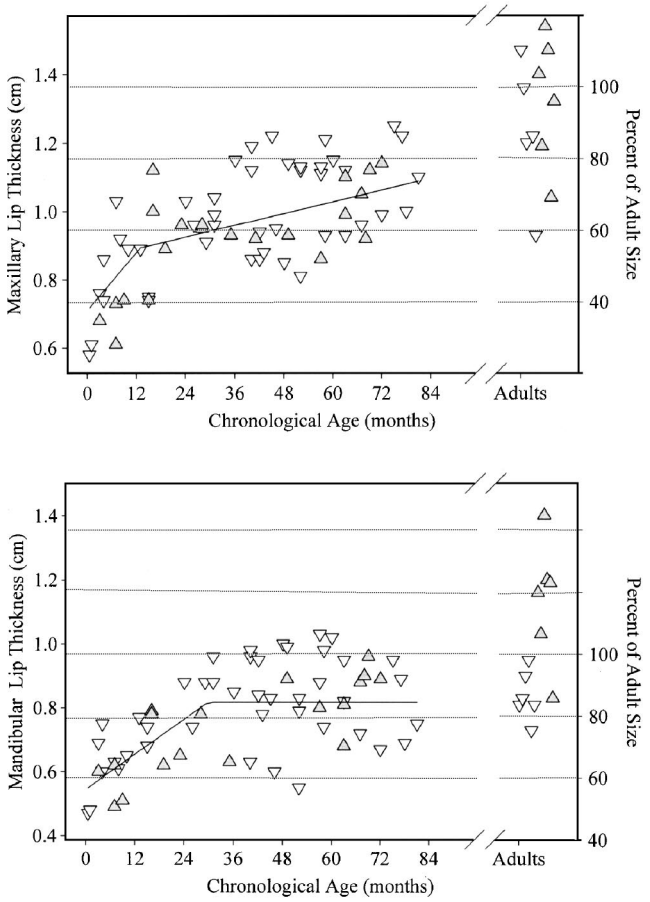


FIG. 11. Maxillary lip thickness (top figure) and mandibular lip thickness (bottom figure) for pediatric cases and adult cases (open triangle down for males, and shaded triangle up for females. Maxillary lip thickness is defined as the antero-posterior distance from the subnasale (sn) to the anterior nasal spine (ANS). Mandibular lip thickness is defined as the horizontal antero-posterior distance from the supramentale to the hard-tissue line.

surements included were the maxillary and mandibular lip thickness, hard palate length, and mandibular length. (ii) Posterior or pharyngeal structures that grow mostly in the vertical plane (i.e., run parallel to facial height [the distance from the nasion (n) to gnathion (gn)]. Measurements included were soft palate length, pharyngeal length, hyoid, and laryngeal descent. (iii) Combined (anterior and posterior) structures, that grow in both the horizontal and vertical planes. Measurements included were vocal tract length and tongue length. Using the estimated breakpoints that varied over the structures in the above-described broken line model (see Table I), the effect of pre- versus postbreakpoint segment (i.e., initial versus final slope) and structure region/orientation (i.e., anterior/horizontal versus posterior/vertical versus combined) on slope was investigated using a nested analysis of variance with adult structure size included as a covariate. This was followed by pairwise comparisons of the three regions/orientations using the least significant difference method.

3. Relational growth of the different vocal tract structures with vocal tract length

The estimated breakpoints varied over the structures in the above described broken line model (Sec. II C 1). A typical breakpoint value of 18 months was chosen in order to divide pediatric subjects into two groups. These groups were then used to assess the developmental changes in the relational growth of the different vocal tract structures with vocal tract length. This was achieved by calculating the between-subject *r*-squared statistic (between-subject correlation coefficient of each structure with vocal tract length squared) and then comparing these values across the different structures and the different age groups. The *r*-squared statistic is the percent of subject-to-subject variation in vocal tract length explained by other structures (a growth measurement).

III. RESULTS

A. Anatomic growth of the vocal tract structures

All available measurements from the 69 pediatric cases, between the ages 2 weeks to 6 years 9 months, and the 12 adults are plotted in Figs. 4 to 11, with growth curve fits for the pediatric data and a second *Y* axis referencing the pediatric data to the percent of the average adult size. The figures depict the development of *vocal tract length* (Fig. 4), *tongue length* (Fig. 5), *naso-oro-pharyngeal length* (Fig. 6), *laryngeal descent* (Fig. 7), *hyoid decent* (Fig. 8), *hard and soft palate length* (Fig. 9), *mandibular length*, and *mandibular depth* (Fig. 10), and *maxillary and mandibular lip thickness*

TABLE I. Parameter values with standard error (SE) of the broken line growth curve model described. (a) intercept; (b) slope term for ages less than the breakpoint (brk); (c) slope term for ages greater than the breakpoint; and (brk) the estimated breakpoint in chronological age (ca) in months.

	a	SE (*a*)	*b*	SE (*b*)	*c*	SE (*c*)	brk	SE (brk)
Vocal tract length	7.079	0.028	0.139	0.021	0.036	0.131	15.935	2.803
Tongue length	5.681	0.033	0.082	0.019	0.027	0.166	16.007	4.857
Pharyngeal length	4.772	0.054	0.117	0.037	0.027	0.196	14.964	5.178
Laryngeal descent	2.966	0.067	0.103	0.030	0.015	0.317	14.990	4.476
Hyoid descent	2.545	0.068	0.066	0.018	0.012	0.476	19.884	5.801
Palate (hard and soft)	5.155	0.034	0.064	0.018	0.017	0.258	17.830	5.645
Hard palate	2.787	0.039	0.043	0.008	0.004	0.804	24.212	3.739
Soft palate	2.362	0.036	0.011	0.002	NA	NA	NA	NA
Mandibular length	2.657	0.133	0.120	0.078	0.018	0.382	12.936	8.446
Mandibular depth	4.374	0.057	0.049	0.013	0.027	0.329	30.976	14.562
Maxillary lip thick	0.711	0.064	0.014	0.007	0.003	0.313	13.017	7.151
Mandibular lip thick	0.545	0.070	0.009	0.002	NA	NA	29.767	5.512

(Fig. 11). Figure 4 shows that the development of vocal tract length increases from birth to age 6 years 9 months, with a somewhat more rapid growth during approximately the first 16 months of life. The vocal tract length measurements for children between 2 to 6 years as well as the adults are in congruence with data reported by Fitch and Giedd (1999) using MRI procedures. The average vocal tract length by age 18 months is 8.3 cm, which, as plotted on the second *Y* axis of Fig. 4 labeled ''percent of adult size'', is a measure that is about 55% the adult vocal tract length; and at age 6 years, the average length is 11.4 cm, which is about 75% the adult vocal tract length. Figures 5 to 11 depict the development of the various hard- and soft-tissue structures that contribute towards vocal tract lengthening. Most structures, despite differences in growth rate, appear to follow a growth pattern or a growth curve that is similar to that of the vocal tract length. That is, most structures appear to have an ongoing growth from age 2 weeks to age 6 years 9 months with a somewhat more rapid growth during approximately the first 18 months of life. At about age 18 months, the various vocal tract structures achieve between 55% to 80% of the adult size, and at about age 6 years, they are between 65% to 85% of the adult size. Such percentage ranges indicate that some structures get closer to their adult mature size sooner than others. For example, the hard palate length and maxillary lip thickness (both anterior or oral structures) are at 80% of their adult mature size by age 18 months; however, other structures, such as mandibular depth and pharyngeal length (predominantly posterior or pharyngeal structures), reach 80% of their adult mature size by about age 6 years. Other structures such as laryngeal descent (65%), hyoid descent (65%), and tongue length (70%), continue to undergo considerable growth after age 6 until they reach their adult mature size.

The observation of a more rapid rate of growth during early childhood (depicted in Figs. 4 to 11) was statistically supported by modeling the relationships between age and each of the vocal tract structures of the pediatric data with the broken line growth curve model. As noted above (Sec. II C 1), this model was used because it allows testing for a change in the rate of growth, and provides an estimate of the time of the change or the ''breakpoint.'' Table I lists the estimated parameters and their standard errors. The mean value for the intercept term ''*a*'' is used in the fitted model curves plotted in Figs. 4 to 11. As seen in Table I, with the exception of the soft palate that does not have a breakpoint (brk) (i.e., one line fits all the data), the growth rate for most structures is faster prebreakpoint than postbreakpoint. That is, the slope terms *b* (prebreakpoint) are generally greater than slope terms *c* (postbreakpoint; $p<0.0001$). Also, growth continues postbreakpoint ($c>0$) for most structures except for mandibular lip thickness, which has a zero slope (horizontal) for the second segment (postbreakpoint). We also fit models with random slope terms (*b* and *c*) and random breakpoints, but likelihood ratio tests indicated that these models did not improve the fit significantly. All vocal tract measurements were transformed to the log scale to stabilize the variance. Furthermore, we tested for differences due to gender by comparing a general model which allows the (mean) parameter values to vary by gender to a simpler model which ignores gender. In all cases the effect of gender was not significant ($p>0.05$). The need for a more complex within-subject correlation structure was also explored by comparing a within-subject independence to autoregressive serial correlation. Once again, the fit of the more complex model was not a significant improvement.

B. Growth of vocal tract structures based on region/orientation

The different vocal tract structures were grouped into three region/orientation classes (i.e., anterior/horizontal versus posterior/vertical versus combined (see Sec. II C 2). The effect of these classes on the slopes was significant ($p=0.019$). The interaction term, however, was not significant ($p=0.654$), indicating that the pattern of differences in slopes between the three regions/orientations was consistent across segments (pre- and postbreakpoints). Pairwise comparisons of these three regions/orientation classes found significant differences among all pairs: vertical versus horizontal $p=0.013$; combined versus vertical $p=0.026$; combined versus horizontal $p=0.0006$. Table II lists the means for region/orientation of pre- and postbreakpoint segments. Note that the means are larger prebreakpoint than postbreakpoint, and that the ordering of the means (combined>vertical

TABLE II. Means of initial slope (prebreakpoint segment) and final slope (postbreakpoint segment) separated by region/orientation. SE indicates standard error. See the text for the vocal tract structures' region/orientation assignment.

Region/orientation	Mean initial slope prebreakpoint segment	SE	Mean final slope postbreakpoint segment	SE	*n*
Anterior/horizontal	0.046	0.022	0.006	0.004	12
Posterior/vertical	0.069	0.020	0.019	0.004	12
Combined	0.095	0.026	0.027	0.005	12

>horizontal, which reflects the order of growth) is the same for the initial slope (prebreakpoint segment) and final slope (postbreakpoint segment).

C. Relational growth of the different vocal tract structures with vocal tract length

As noted in Table I, the different structures have different breakpoints (brk) ranging from 13 months (e.g., mandibular length) to 31 months (e.g., mandibular depth). It is therefore difficult to compare growth rates (i.e., slopes) across structures. In order to assess the amount of variability in vocal tract length which can be explained by the variation in each of the other structures (i.e., to assess the relational growth of the different vocal tract structures with vocal tract length), we first divided the subjects into three groups (two pediatric groups and one adult group), and then computed the between-subjects *r*-squared statistic for all the structures in each group. The pediatric subjects were divided into two groups using the age of 18 month as the dividing value. This decision was based on the 95% confidence intervals for the estimated breakpoints, and age 18 month was chosen as a compromise between the median estimated breakpoint for the vocal tract measurements (16 months) and the transition point indicated by the acoustic data (24 months) (Robb *et al.*, 1997; Gilbert *et al.*, 1997). In all but one case (mandibular lip thickness) the 95% confidence intervals for the estimated breakpoints included 18 months. Thus, the first pediatric group (peds I; $n=19$) consisted of cases between the ages of 2 weeks and 18 months. The second pediatric group (peds II; $n=48$) consisted of cases ages 19 months to 6 years and 9 months. The third group (adults; $n=12$) consisted of adults with vocal tract structures that have reached their mature size. The adult group was included to assess the pediatric data in terms of percent of adult size as discussed above (right-hand second *Y* axis in Figs. 4 to 11), and to compare the relational growth of the different vocal tract structures with vocal tract length during the course of development. For each group, the between-subject *r*-squared was calculated (between-subject correlation coefficient of each structure with vocal tract length squared). The between-subject *r*-squared is the percent of variability in vocal tract length among subjects explained by that measurement. The *r*-squared calculations were not sensitive to small changes in the breakpoint, i.e., the results and conclusions outlined below were not altered when the data were analyzed using different cutoff ages, such as 16 or 24 months, for the pediatric data.

Figure 12 is a squares plot that represents the percent relational growth by the size (volume) of the squares. Larger squares indicate more relational growth. For all three groups (with the exception of the hard palate in the peds II group and of mandibular depth in adults) all hard- and soft-tissue vocal tract structures demonstrate concurrent or relational growth with vocal tract length. However, the extent of relational growth changes for each structure during the course of development. The first column of Fig. 12 lists in order the percent relational growth of the various structures with overall vocal tract length for the first pediatric group, peds I. The second and third columns reflect the percent relational growth for the second pediatric group, peds II, and adults. For all three age groups, naso-oro-pharyngeal length, tongue length, and laryngeal descent have the highest relational growth with vocal tract length. However, comparison of the percentages for each of those structures, as well as of the remaining structures, across the three groups indicates that the extent of relational growth changes (decreases or increases) during the course of development. For example, pharyngeal length is the largest for peds I (82%) and adults (66%), but the third in line for peds II (57%). In contrast, laryngeal descent is the largest for peds II (66%) but the third largest for peds I (67%) and adults (45%). Such differences in order and percent variation explained appear to stem from differences in growth in the naso-pharyngeal region (horizontal plane) versus the oro-pharyngeal region (vertical plane). This inference is based on examining change in structures in the naso-oro-pharyngeal region, namely, palatal length and laryngeal descent. The former structure is below the posterior naris where the naso-pharynx measure starts, and the latter structure is where the oro-pharynx measure ends. In peds I and in adults, the percent variation explained by the hard palate (58% and 22%) is more than the soft palate (2% and 16%). However, for peds II, the hard palate (0.01%) has barely any relational growth with vocal tract length. Thus, it appears that in peds I and in adults there are developmental changes, in the horizontal plane, at the beginning of where the measure of the naso-pharynx is initiated. Thus, during the course of development, there are not only differences in the extent of relational growth of the different vocal tract structures with vocal tract length (decreases or increases), but differences in relational growth based on structure region/orientation (anterior/horizontal versus posterior/vertical). Indeed, comparison of the various structures listed in Fig. 12 based on region (anterior versus posterior) and orientation or plane of growth (horizontal versus vertical) indicate that during approximately the first 18 months of life (peds I group) there is about as much growth and subsequently relational growth of the anterior structures in the horizontal plane—such as hard palate length (58%)

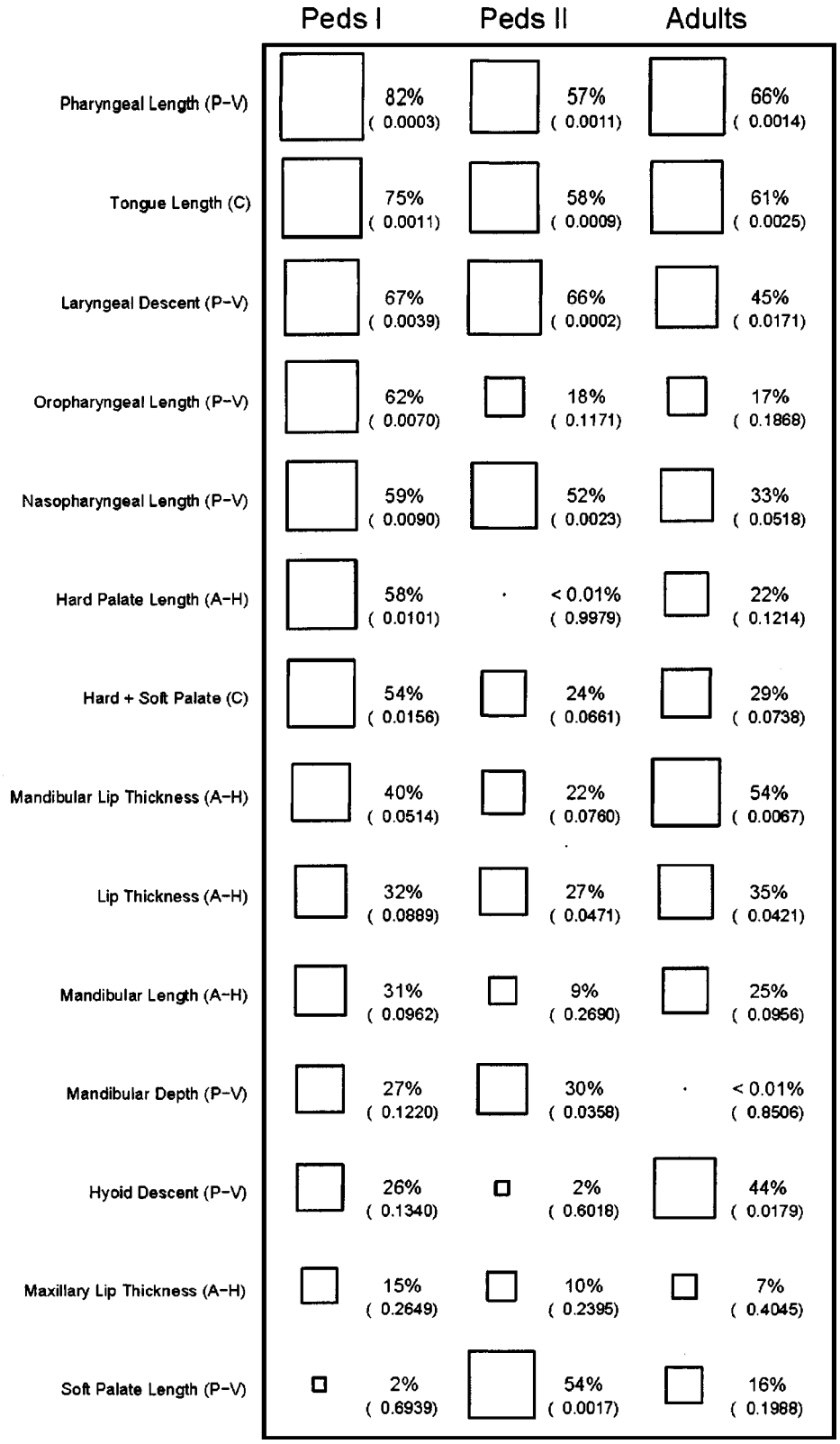


FIG. 12. Squares plot showing the *r*-squared or percent variability in vocal tract length explained by the various vocal tract structures. The area of the squares is proportional to *r*-squared. The numbers in parentheses represents the *p* value for the test where *r*-squared is different from zero. Structures include pharyngeal length (naso+oro pharyngeal lengths), tongue length, laryngeal descent, palate length (hard+soft palate lengths), lip thickness (average maxillary +mandibular lip thickness), mandibular length, mandibular depth, and hyoid descent. Each structure is followed by its region/orientation: P–V (posterior/vertical); A–H (anterior/horizontal); and C (combined anterior/horizontal and posterior/vertical). Note that the listed order of structures is based on the percentages for peds I only. The percent vocal tract variability explained changes (increases or decreases) for each group (peds I, peds II, and adults).

and mandibular length (31%)—as there is growth of the posterior structures in the vertical plane—such as laryngeal descent (67%), mandibular depth (27%), and hyoid descent (26%). In contrast, from the ages of 19 months to 6 years and 9 months (peds II group), the growth, and subsequently the relational growth, of most of the posterior structures in the vertical dimension, such as laryngeal descent (66%), soft palate length (54%), and mandibular depth (30%), exceeds the growth of the anterior structures in the horizontal plane, such as mandibular length (9%) and hard palate length (0.01%). Similarly, for the adults' group, the growth and consequently the relational growth of the posterior structures with vocal tract length, such as laryngeal descent (45%), and hyoid descent (44%), exceeds that of the anterior structures, such as mandibular length (25%), hard palate length (22%), and lip thickness (35%). Thus, although there is notable

growth for most structures in both the vertical and horizontal planes at all age groups, growth in the horizontal plane slows down for peds II, but then increases at some later point to reach the mature adult size.

IV. DISCUSSION

A. Current findings

The current data from sagittal MRI slices provide detailed developmental information on the changes that the various hard- and soft-tissue structures undergo, and their relational growth with vocal tract length. These data are without precedent in that they were obtained from children from birth to age 6 as well as adults, and thus they capture an age range—specifically birth to age 4—where most of the important anatomic restructuring occurs. Current findings, as discussed below, indicate that (a) there is ongoing growth, with no sexual dimorphism, between birth and age 6 years 9 months and a period of accelerated growth or growth spurt for most vocal tract structures between birth and approximately age 18 months; (b) the growth pattern of the various vocal tract structures varies according to its region (anterior, posterior, or combined) and orientation of growth (horizontal, vertical, or combined), wherein anterior structures appear to have a neural growth curve, posterior structures have a somatic growth curve, and structures/measurements encompassing both anterior and posterior regions—such as vocal tract length—have a combined or intermediate neural and somatic growth pattern as defined by Scammon (1930); and (c) the relational growth of the different vocal tract structures with vocal tract length changes during the course of development as a function of structure orientation (horizontal/vertical). These findings, followed by implications for speech acoustics and methodological issues, are discussed below.

The findings described above support the documented fact that growth or development does not advance in a linear fashion, but rather involves one or more periods of rapid growth or growth spurts (Gollin, 1981). As seen in Figs. 4 to 11, all the vocal tract structures examined exhibit a consistent trend of growth with no sexual dimorphism during the first 6 years of life, and a period of significant rapid or accelerated growth during early childhood, typically between birth and approximately age 18 months. See Table I for the breakpoints (brk.—age where growth rate changes) of specific structures as calculated by the broken line growth curve model. In general, the various vocal tract structures appear to follow the neural and/or the somatic growth curves as defined by Scammon (1930). The neural growth curve is characterized by a period of extremely rapid growth following birth until some variable time in early childhood where the structure is 2/3 of its adult size, followed by a period of very slow growth until maturity. The somatic growth curve is similar to the neural growth curve in having a phase of very rapid growth following birth and during infancy, but at the end of this rapid growth period, the structure is barely over a quarter of its adult size. The rapid growth period is then followed by an interval of regular but slow growth in early and middle childhood, then a period of rapid growth during puberty, and finally a period of slow but steady growth until early maturity. Figures 4 to 11 indicate that generally the growth of structures that are in the horizontal plane, such as the hard palate, follows a neural growth curve. For example, as seen in Fig. 9, the hard palate has an accelerated growth period from age 2 weeks to age 24 months, at which time it has reached 84% of its adult size. In contrast, the growth of structures that are in the vertical plane, such as the soft palate and laryngeal descent, follows a somatic growth curve. Furthermore, structures that are in both the horizontal and vertical planes, such as vocal tract length and tongue length, follow a combined or intermediate, neural and somatic growth curve. As noted in the Results section, the differences in growth between these three regions/orientations are significant. Although the available data in the current study are limited up to age 6 years 9 months and adults, these statements on type of growth curve are substantiated by examining the measures of the various structures at around age 18 months, and determining their percentage of the mature adult size. Differences in growth trajectories of vocal tract structures where the growth rates of the vertical and the horizontal portions of the vocal tract are different are also reported by Lieberman *et al.* (1999, 2001). The finding of accelerated growth during early childhood, particularly for vocal tract length, has implications for speech acoustics that are discussed below.

The ordering of the means reported in Table II—for both pre- and postbreakpoint segments (initial and final slopes)—reflects that combined (anterior and posterior) structures have the largest growth followed by posterior structures, and last by anterior structures for both the prebreakpoint and postbreakpoint segments. This order of growth was expected since the combined structures have larger measurements than the anterior or posterior structures alone because they include both anterior/horizontal and posterior/vertical regions/orientations. The posterior/vertical mean was also expected to exceed the anterior/horizontal mean since most anatomic restructuring—specifically laryngeal and hyoid descent—is reported to occur in the posterior or pharyngeal region where growth is predominantly in the vertical dimension. Thus, there is significantly more growth in the vertical dimension that in the horizontal dimension for both prebreakpoint and postbreakpoint segments.

Comparison of slope terms *b* (prebreakpoint segment) and slope terms *c* (postbreakpoint segment) in Table I, shows that the various vocal tract structures, or measures within a structure, have different growth rates that vary as a function of age but not gender. Consequently, it is to be expected that the relational growth of each of those structures with vocal tract length changes during the course of development (see Fig. 12). Based on knowledge of anatomic restructuring of the vocal tract (descent of the larynx and tongue), and the above finding on significant main effect for region/orientation in the nested ANOVA of pediatric pre- and postbreakpoint slopes, where comparison of the means in Table II indicates the growth of posterior structures to exceed the growth of anterior structures for both prebreakpoint and postbreakpoint segment (all pairwise differences significant), we expected the relational growth of the posterior structures (such as hyoid descent, laryngeal descent, and mandibular

depth—structures with growth in the vertical dimension) with vocal tract length to exceed the relational growth of the anterior structures (such as palate length and mandibular length—structures with growth in the horizontal plane). Indeed, comparisons of the structures ordered according to percent variability (*r*-squared) in decreasing order (as is the case for peds I in Fig. 12), while taking into account the magnitude of percent variability, for each of the three age groups (peds I, peds II, and adults) indicate that although the relational growth of the different vocal tract structures with vocal tract length changes as a function of age, the relational growth of the posterior structures with vocal tract length is ongoing for all age groups. However, during the first 18 months of life (peds I group), there is almost as much associated growth of the anterior structures in the horizontal plane with vocal tract length as there is of posterior structures in the vertical plane with vocal tract length. For example, Fig. 12 shows that for peds I the first largest (82%) percent variability is from pharyngeal length—a posterior structure in the vertical dimension—and the sixth largest (58%) percent variability is from hard-palate length—an anterior structure in the horizontal dimension. After age 18 months, in general, the relational growth of posterior structures exceeds that of the anterior structures (i.e., growth is predominantly in the vertical dimension). For example, the third largest (57%) percent variability is from pharyngeal length, and the last (0.01%) percent variability is from hard-palate length. Such growth is also more pronounced for the adults' group. Thus, between the age of 6 years 9 months and adulthood, where the various vocal tract structures reach their mature size, the growth of all the different vocal tract structures persist, particularly for structures in the posterior region of the vocal tract. Fitch and Giedd's (1999) findings on vocal tract length using MRI indicate that growth in the posterior region of the vocal tract occurs predominantly during adolescence, particularly in males.

B. Acoustic implications

The current data on anatomic growth of the vocal tract structures provide an important first step towards exploring anatomic–acoustic interdependencies, that is, relating anatomic growth patterns to developmental changes in speech acoustics. The relevance of the anatomic findings in this report of an early period of accelerated growth of all vocal tract structures, including vocal tract length, and the absence of sexual dimorphism are discussed below with special consideration given to current knowledge of developmental changes in speech acoustics. The three major areas in speech acoustics discussed are (a) anticipated decrease in formant frequency as vocal tract length increases (Fant, 1960); (b) discriminable sex differences in speech acoustics by age 4; and (c) variability of the acoustic signal during development.

As the vocal tract develops, its acoustic properties change (Fant, 1960). Specifically, as age increases and the vocal tract lengthens, formant frequencies decrease. This decrease is typically attributed to increases in vocal tract length. However, during the first 2 years of life, there is evidence that the average formant frequencies remain unchanged, i.e., do not decrease, though there is an increase in the range of formant values (dispersion) (Buhr, 1980; Gilbert, Robb and Chen, 1997; Kent and Murray, 1982; Robb, Chen, and Gilbert, 1997; Stathopoulos, 1995). The anatomic implication of such acoustic findings is that there would be little or no expected change in vocal tract length during the first 2 years of life. However, as seen in Fig. 4, vocal tract length increases by 1.5 to 2 cm during this time period. Robb *et al.* (1997) attributed the lack of decrease in formant values during the first 2 years of life to developmental changes in anatomy that involve a more complex reconfiguration than a simple lengthening process. They also attributed the dispersion of formant frequencies to a larger vocal tract that allows for greater variability of tongue movement. Thus, there is an apparent need to supplement findings from the sagittal plane with data from multiple planes to gain a thorough understanding of the anatomic changes that the developing vocal tract and its component structures undergo so as to better understand acoustic–anatomic interdependencies. In other words, anatomic measurements should include changes in size (such as length and width) as well as changes in shape and configuration (such as area and volume). Such knowledge, in addition to assessing the role that anatomic growth plays in developmental changes in speech acoustics, would also lay the necessary normative foundation for the study of atypical growth patterns to determine how structural differences may contribute toward early neuromotor problems.

A detailed multidimensional understanding of the anatomic changes in supralaryngeal region that occur during development should, furthermore, lay the biological foundation for the perceived sex differences in speech acoustics. Despite the absence of sexual dimorphism in vocal tract structures and vocal tract length—measured in the sagittal plane—there are differences in formant frequencies between preadolescent boys and girls where boys as young as 4 years old have vowel formant frequencies that are lower than those of girls (Perry, Ohde, and Ashmead, 2001; Whiteside and Hodgson, 2000). This has been attributed to boys having a larger vocal tract than girls. Such differences in formant frequencies between boys and girls are present even when physical measurements such as body height and weight are taken into account since vocal tract length correlates with body dimensions (Fitch and Giedd, 1999). Thus male/female differences in formant frequencies before age 10 may be due to anatomical and/or articulatory sources other than vocal tract length. However, such conclusions based on acoustic changes are inferential and nonspecific, and again point to the need to have detailed understanding of the anatomic changes in the oral and pharyngeal regions that occur during development. Of particular interest are changes in the pharyngeal region, since most of the sex differences in adult speech acoustics have been attributed to differences in the pharyngeal region (Fant, 1966). Though not replicated here, King (1952), using x-ray cephalometry, has reported sex differences in pharyngeal length by the age of 1 where males have a longer pharyngeal length than females. King, however, defined pharyngeal length as a vertical measurement from the hard palate to the hyoid bone, which is only a portion of the actual naso-oro-pharyngeal length. Thus, a thorough investigation of developmental changes in the naso-oro-pharyngeal region is

warranted using advanced imaging technology to assess changes in size (such as length and breadth) and volume of the pharynx. Such anatomic examination of sex-related differences in the growth of the front and back cavities may also help address the problem of the scaling of formant patterns as discussed by Kent (1976) and Goldstein (1980) where the uniform axial growth of the vocal tract is in question (Fant, 1966).

Having a detailed account of the biological changes in vocal tract anatomy, specifically identifying the growth pattern across development—including all periods of accelerated growth—would also promote our understanding of the anatomic bases of motor adjustments in speech development. A common interpretation made regarding motor development from acoustic and physiologic studies is that the high degree of variability that is observed in children's performance reflects the developing neuromuscular control or speech motor control whereby variability decreases as age increases (Eguchi and Hirsh, 1969; Tingley and Allen, 1975; Smith, Sugarman, and Long, 1983; Smith, 1994; Kent, 1976, 1992; Kent and Forner, 1980; Sharkey and Folkins, 1985; Lee *et al.*, 1999; Walsh and Smith, 2002; Wohlert and Smith, 2002). However, the assumption that variability is a direct measure of the maturation of the neural processes in speech motor control can be questioned. The variability measures are taken while there is ongoing anatomic growth including increases in size and changes in shape that alter the relative position of the various structures. Interestingly, as anatomic growth begins to decelerate after the first few years of life, so does variability. Thus, in addition to neural maturation, variability may reflect other developmental aspects, such as anatomic growth. At the least, it is necessary to consider several sources of variability in speech movements, including physical growth and remodeling of individual structures.

C. Other ramifications

Aside from using anatomic growth data to relate anatomic development to changes in speech acoustics, such data can also be used toward: gaining an appreciation of the simultaneous changes that the various vocal tract structures undergo while relating it to changes of the composite head and face structures (Vorperian *et al.*, 1999; Vorperian, 2000); developing infant and young child specific models of the developing vocal tract such as neural network models (Callan *et al.*, 2000); advancing statistical models of the developing vocal tract; and programming dynamic articulatory models of the developing vocal tract that have computer codes to make predictions regarding the acoustic formant space (Milenkovic, 1998; Milenkovic and Milenkovic, 1998). Such efforts would be instrumental in understanding how differential rates of growth of structures during development effect vocal tract geometry and in turn determine the articulatory space, i.e., production models specific to infants and young children. Such efforts will also help delineate anatomic versus motor contributions towards changes in acoustic output of the vocal tract.

ACKNOWLEDGMENTS

This work was supported by NIH/NIDCD Grant No. R03-DC4362, titled ''Anatomic Development of the Vocal Tract: MRI Procedures,'' and Grant No. R01-DC006282, titled ''MRI & CT Studies of the Developing Vocal Tract.'' Portions of this paper were presented in 2001 at the 141st meeting of the Acoustical Society of America in Chicago, also at the 1999 annual convention of the American Speech-Language Hearing Association in San Francisco, CA. Special thanks to the UW/VA Swallowing Laboratory for assistance in securing the adult MRIs; and Reid Durtschi and Celia Choih for assistance with figure preparation. The authors also thank Anders Lofqvist, John Tiedemann, and two anonymous reviewers for comments on earlier versions of this manuscript.

Bacon, D. W., and Watts, D. G. (**1971**). ''Estimating the transition between two intersecting straight lines,'' Biometrika **58**, 525–534.

Bosma, J. F. (**1975a**). ''Anatomic and physiologic development of the speech apparatus,'' in *The Nervous System: Human Communication and its Disorders*, edited by D. B. Tower (Raven, New York), Vol. 3, pp. 469–481.

Bosma, J. F. (**1975b**). ''Introduction,'' in *Symposium on Development of Upper Respiratory Anatomy and Function: Implications Regarding Sudden and Unexpected Infant Death*, edited by J. F. Bosma and J. Showacre (Government Printing Office, Washington, D.C.), pp. 5–44.

Bosma, J. F. (**1976**). Discussion of the paper, ''Postnatal development of the basicranium and vocal tract in man,'' by J. T. Laitman and E. S. Crelin, in *Symposium on the Development of the Basicranium*, edited by J. Bosma. DHEW Publication No. 76-989 (PHS-NIH, Bethesda, MD).

Bosma, J. F. (**1985**). ''Postnatal ontogeny of performance of the pharynx, larynx, and mouth,'' Am. Rev. Respir. Dis. **131**, S10–S15.

Buhr, R. D. (**1980**). ''The emeregence of vowels in an infant,'' J. Speech Hear. Res. **23**, 73–94.

Callan, D. E., Kent, R. D., Gunther, R. W., and Vorperian, H. K. (**2000**). ''An auditory-feedback-based neural network model of speech production in the developing child,'' J. Speech Hear. Res. **43**(3), 721–736.

Crelin, E. S. (**1973**). *Functional Anatomy of the Newborn* (Yale University Press, New Haven, CT).

Crelin, E. S. (**1976**). ''Development of the upper respiratory system,'' Clin. Symp. **28**, 1–30.

Eguchi, S., and Hirsh, I. J. (**1969**). ''Development of speech sounds in children,'' Acta Oto-Laryngol., Suppl. **257**, 5–51.

Fant, G. (**1960**). *Acoustic Theory of Speech Production: With Calculations based on X-ray Studies of Russian Articulations* (Mouton, The Hague).

Fant, G. (**1966**). ''A note on vocal tract size factors and non-uniform F-pattern scalings,'' Speech Trans. Lab. Quart. Prog. Stat. Rep. **4**, 22–30.

Farkas, L. G., Posnick, J. C., and Hreczko, T. M. (**1992a**). ''Growth and development of regional units in the head and face based on anthropometric measurements,'' Cleft Palate Craniofac J. **29**, 301–302.

Farkas, L. G., Posnick, J. C., and Hreczko, T. M. (**1992b**). ''Anthropometric growth study of the head,'' Cleft Palate Craniofac J. **29**, 303–308.

Fitch, T., and Giedd, J. (**1999**). ''Morphology and development of the human vocal tract: A study using magnetic resonance imaging,'' J. Acoust. Soc. Am. **106**(3), 1511–1522.

Fried, M. P., Kelly, J. H., and Strome, M. (**1982**). ''Comparison of the adult and infant larynx,'' J. Fam. Pract. **15**(3), 557–561.

Gilbert, H. R., Robb, M. P., and Chen, Y. (**1997**). ''Formant frequency development: 15 to 36 months,'' J. Voice **11**(3), 260–266.

Goldstein, U. G. (**1980**). ''An articulatory model for the vocal tracts of growing children,'' Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Gollin, E. (**1981**). ''Development and plasticity,'' in *Developmental Plasticity*, edited by E. Gollin (Academic, New York), pp. 231–252.

Katzberg, R. W., and Westersson, P. L. (**1991**). ''Magnetic resonance imaging,'' Cranio Clin. Int. **1**(1), 93–116.

Kent, R. D. (**1976**). ''Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies,'' J. Speech Lang. Hear. Res. **19**, 421–447.

Kent, R. D. (**1981**). ''Articulatory-acoustic perspectives on speech development,'' in *Language Behavior in Infancy and Early Childhood*, edited by R. Stark (North-Holland, Elsevier, Amsterdam), pp. 105–239.
Kent, R. D. (**1992**). ''Phonological development as biology and behavior,'' in *Processes in Language Acquisition and Disorders*, edited by R. S. Chapman (Mosby-Year Book, St. Louis, MO), pp. 67–85.
Kent, R. D., and Forner, L. L. (**1980**). ''Speech segment durations in sentence recitations by children and adults,'' J. Phonetics **12**, 157–168.
Kent, R. D., and Murray, A. (**1982**). ''Acoustic features of infant vocalic utterances at 3, 6, and 9 months,'' J. Acoust. Soc. Am. **72**, 353–365.
Kent, R. D., and Vorperian, H. K. (**1995**). ''Anatomic development of the craniofacial-oral-laryngeal systems: A review,'' J. Med. Speech-Lang. Pathol. **3**, 145–190.
King, E. W. (**1952**). ''A roetgenograohic study of pharyngeal growth,'' Angle Orthod. **22**, 23–37.
Laitman, J. T., and Crelin, E. S. (**1976**). ''Postnatal development of the basicranium and vocal tract region in man,'' in *Symposium on the Development of the Basicranium*, edited by J. Bosma. DHEW Publication No. 76-989 (PHS-NIH, Bethesda, MD), pp. 206–220.
Lee, S., Potamianos, A., and Narayanan, S. (**1999**). ''Acoustics of children's speech: Developmental changes of temporal and spectral parameters,'' J. Acoust. Soc. Am. **105**(3), 1455–1468.
Lieberman, P. (**1977**). *Speech Physiology and Acoustic Phonetics* (Macmillan, New York).
Lieberman, P. (**1984**). *The Biology and Evolution of Language* (Harvard University Press, Cambridge, MA).
Lieberman, D. E., and McCarthy, R. C. (**1999**). ''The ontogeny of cranial base angulation in humans and chimpanzees and its implications for reconstructing pharyngeal dimensions,'' J. Hum. Evol. **36**, 487–517.
Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., and Palmer, J. B. (**2001**). ''Ontogeny of postnatal hyoid and larynx descent in humans,'' Arch. Oral Biol. **46**, 117–128.
Milenkovic, P. H. (**1998**). ''Reconstructing the vocal tract area function of vowels and liquids from microbeam data with an articulatory model,'' in *Proceedings of Linguistic Phonetics*, edited by O. Fujimura and B. Palek, pp. 425–437.
Milenkovic, P. H., and Milenkovic, V. (**1998**). ''Coordination of redundant articulatory parameters near a singularity in the vocal mechanism,'' in *Advances in Robot Kinematics and Control*, edited by J. Lenarcic and M. Husty (Kluwer, Dordrecht), pp. 493–500.
Mowrer, D. E. (**1980**). ''Phonological development during the first year of life,'' in *Speech and Language: Advances in Basic Research and Practice*, edited by N. J. Lass (Academic, New York), Vol. 4, pp. 99–137.
National Center for Health Statistics: NCHS Growth Charts (1976). Monthly Vital Statistics Report, 25 (3) Supp HRA 76-1120, Rockville, MD.
Perry, T. L., Ohde, R. N., and Ashmead, D. H. (**2001**). ''The acoustic bases for gender identification from children's voices,'' J. Acoust. Soc. Am. **109**(6), 2988–2998.
Pinheiro, J. C., and Bates, D. M. (**2000**). *Mixed Effects Models in {S} and {S-PLUS}* (Springer, New York).
Robb, M. P., Chen, Y., and Gilbert, H. R. (**1997**). ''Developmental aspects of formant frequency and bandwiths in infants and toddlers,'' Folia Phoniatr Logop **49**, 88–95.
Sasaki, C. T., Levine, P. A., Laitman, J. T., and Crelin, E. S. (**1977**). ''Postnatal descent of the epiglottis in man,'' Arch. Otolaryngol. **103**, 169–171.
Scammon, R. E. (**1930**). ''The measurement of the body in childhood,'' in *The Measurement of Man*, edited by J. A. Harris, C. M. Jackson, D. G. Patterson, and R. E. Scammon (University of Minnesota Press, Minneapolis), pp. 173–215.
Sharkey, S. G., and Folkins, J. W. (**1985**). ''Variability of lip and jaw movements in children and adults: Implications for the development of speech motor control,'' J. Speech Hear. Res. **28**, 8–15.
Shorten, G. D., Opie, N. J., Graziotti, P., Morris, I., and Khangure, M. (**1994**). ''Assessment of upper airway anatomy in awake, sedated, and anaesthetised patients using magnetic resonance imaging,'' Anaesth. Intensive Care **22**(2), 165–169.
Smith, B. L. (**1994**). ''Effects of experimental manipulations and intrinsic contrasts on relationships between duration and temporal variability in children's and adult's speech,'' J. Phonetics **22**, 155–175.
Smith, B. L., and Oller, D. K. (**1981**). ''A comparative study of the development of stop consonant production in normal and Down's syndrome children,'' J. Speech Hear Disord. **48**, 114–118.
Smith, B. L., Sugarman, M. D., and Long, S. H. (**1983**). ''Experimental manipulations of speaking rate for studying temporal variability in children's speech,'' J. Acoust. Soc. Am. **74**, 744–749.
Stathopoulos, E. T. (**1995**). ''Variability revisited: An acoustic, aerodynamic, and respiratory kinematic comparison of children and adults during speech,'' J. Phonetics **23**, 67–80.
Thelen, E. (**1991**). ''Motor aspects of emergent speech: A dynamic approach,'' in *Biological and Behavioral Determinants of Language Development*, edited by N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch, and M. Studdert-Kennedy (Erlbaum, Hilldale, NJ), pp. 339–362.
Tingley, B. M., and Allen, G. D. (**1975**). ''Development of speech timing control in children,'' Child Dev. **46**, 186–194.
Vorperian, H. K. (**2000**). *Anatomic Development of the Vocal Tract Structures as Visualized by MRI*, Published doctoral dissertation (University of Wisconsin-Madison, Madison, WI).
Vorperian, H. K., Kent, R. D., Gentry, L. R., and Yandell, B. S. (**1999**). ''Magnetic resonance imaging procedures to study the concurrent anatomic development of the vocal tract structures: Preliminary results,'' Int. J. Pediatr. Otorhinolaryngol. **49**, 197–206.
Walsh, B., and Smith, A. (**2002**). ''Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes,'' J. Speech Lang. Hear. Res. **45**(6), 1119–1133.
Westhrope, R. N. (**1987**). ''The position of the larynx in children and its relationship to the ease of intubation,'' Anaesth. Intensive Care **15**, 384–388.
Whiteside, S. P., and Hodgson, C. (**2000**). ''Some acoustic characteristics in the voices of 6- to 10-year-old children and adults: A comparative sex and developmental perspective,'' Logoped. Phoniatr. Vocol. **25**(3), 122–132.
Wohlert, A., and Smith, A. (**2002**). ''Developmental change in variability of lip muscle activity during speech,'' J. Speech Lang. Hear. Res. **45**(6), 1077–1087.

The emergence of mature gestural patterns in the production of voiceless and voiced word-final stops[a)]

Susan Nittrouer,[b)] Sandy Estee,[c)] Joanna H. Lowenstein, and Jennifer Smith
Utah State University

(Received 25 June 2003; revised 28 September 2004; accepted 12 October 2004)

The organization of gestures was examined in children's and adults' samples of consonant–vowel–stop words differing in stop voicing. Children (5 and 7 years old) and adults produced words from five voiceless/voiced pairs, five times each in isolation and in sentences. Acoustic measurements were made of vocalic duration, and of the first and second formants at syllable center and voicing offset. The predicted acoustic correlates of syllable-final voicing were observed across speakers: vocalic segments were shorter and first formants were higher in words with voiceless, rather than voiced, final stops. In addition, the second formant was found to differ depending on the voicing of the final stop for all speakers. It was concluded that by 5 years of age children produce words ending in stops with the same overall gestural organization as adults. However, some age-related differences were observed for jaw gestures, and variability for all measures was greater for children than for adults. These results suggest that children are still refining their organization of articulatory gestures past the age of 7 years. Finally, context effects (isolation or sentence) showed that the acoustic correlates of syllable-final voicing are attenuated when words are produced in sentences, rather than in isolation. [DOI: 10.1121/1.1828474]

PACS numbers: 43.70.Bk; 43.70.Ep; 43.70.Fq [AL] Pages: 351–364

I. INTRODUCTION

That young children do not produce speech as adults do is not news, but still, many questions remain concerning how children do eventually attain mature patterns of gestural organization. Concerning initial states, one clear trend is that children's earliest utterances are produced with greater constraints of the surrounding phonetic environment imposed on individual places of constriction for both vocalic and consonantal gestures. For example, this enhanced constraint was found in 1-year-olds' speech for segments involving lingual gestures such that close, front vowels and consonants with alveolar constrictions tended to co-occur, while close, back vowels and consonants with velar constrictions tended to co-occur (Davis and MacNeilage, 1990). In fact, this enhanced constraint of the phonetic environment on lingual gestures has been reported for children as old as 7 years (Nittrouer, 1993). Another trend observed in the earliest productions of children is that they avoid multisyllabic productions involving more than one constriction location for consonants, preferring instead to duplicate a single consonantal constriction within these utterances (e.g., Donahue, 1986; Oller, 1980). Finally, MacNeilage and Davis (1991) described children's earliest productions as deriving almost completely from jaw movements, with other articulators being tightly coupled to these movements. So, for example, variation in vowel quality derives from variation in the degree of jaw lowering, rather than from variation in lingual shape. All these examples suggest that one important developmental change that must be accomplished in order for children to acquire mature gestural patterns for speech production is independent control over articulators. Along with this developmental change, appropriate and precise ways of coordinating separate articulatory gestures must be learned to provide the kinds of stable inter-articulator relations observed for skilled adult speech production (e.g., Tuller and Kelso, 1984).

Tracking the emergence of mature gestural organization is not easy because of methodological obstacles. Some investigators have described the gestural patterns of children's early speech production using narrow phonetic transcription (e.g., Ferguson and Farwell, 1975; Menn, 1978; Piske, 1997; Waterson, 1971), but these analyses are so intensive that it is difficult to do more than diary studies with them. Recent technological advances have made it possible to use acoustic analysis (e.g., Goodell, 1991; Katz, Kripke, and Tallal, 1991; Nittrouer, 1993; Sussman *et al.*, 1999) and direct kinematic measures (e.g., Green *et al.*, 2000; Smith and Goffman, 1998) with children's speech, but both methods have limitations. With acoustic analysis care must be taken to select stimuli for which there is minimal chance of multiple articulatory patterns producing the same acoustic form (Hoole, Nguyen-Trong, and Hardcastle, 1993). Direct kinematic measures with children are best suited for examination of lip and jaw movements because these articulators are the only ones visible without invasive techniques, which are both less likely to succeed with young children and more likely to spawn concern about risk to individual research participants.

The current study investigated the organization of articulatory gestures for the production of words with final stops that were either voiced or voiceless, and was an extension of Nittrouer (1993), in which acoustic analyses of children's and adults' stop-vowel sequences were reported. That study found that children's jaw movements had attained mature

[a)]Portions of this work presented at the 143rd meeting of the Acoustical Society of America, Pittsburgh, June, 2002.
[b)]email: nittrouer@cpd2.usu.edu
[c)]Affiliation: Boys Town National Research Hospital.

gestural form by 3 years of age, but lingual gestures remained highly constrained by the phonetic environment until at least the age of 7 years. The overarching conclusion of that study was that the path to mature patterns of gestural organization is not uniform. Rather, the rate of learning varies depending on the articulator to be used and the shape of the utterance to be produced.

The question of how children learn to make and coordinate the gestures required to produce words differing in the voicing of final stops is interesting partly because of what is known about children's *perception* of syllable-final stop voicing. The two most widely studied properties that convey information about syllable-final stop voicing are (1) the duration of the vocalic portion of the syllable and (2) the first-formant ($F1$) offset transition. Three studies found that children as old as 6 years of age weighted the $F1$ offset transition more and vocalic duration less than adults in making voicing decisions about syllable-final stops (Greenlee, 1980; Krause, 1982b; Wardrip-Fruin and Peach, 1984), although interpretation differed slightly across studies. In particular, Wardrip-Fruin and Peach suggested that the problem exhibited by children might best be described as difficulty in *integrating* the vocalic duration with the $F1$ cue, rather than as an age-related difference in the *weighting* of these cues. Minor differences in interpretation aside, these developmental perception results are interesting in light of cross-linguistic studies demonstrating that the integration and/or weighting of the cues to syllable-final voicing depends on native language experience. Speakers of languages without syllable-final stops (such as Mandarin and Japanese) or of languages that fail to show a vocalic-length difference depending on the voicing of the final stop (such as Arabic) fail to use vocalic duration as a cue in their phonetic decisions for English words ending with stops (Crowther and Mann, 1992; 1994), unless they receive intensive training (Flege, 1989). Furthermore, several studies have demonstrated that native speakers of languages without final stops, or without a vocalic-duration difference associated with the voicing of final stops, do not differentiate vocalic duration for voiceless and voiced final stops in their own productions of English words as much as native English speakers do (Crowther and Mann, 1994; Flege and Port, 1981). Such results support the hypothesis that native language experience with the various acoustic properties distinguishing voicing in word-final stops is required for speaker/listeners to be able to use and reproduce these properties in their own listening and speaking. Thus, the question arises of how children's productions of words containing syllable-final stops differing in voicing are organized, given that they either do not integrate or do not weight the primary cues responsible for distinguishing this voicing dimension as adults do.

One study investigated vocalic duration in children's speech production and reported that 3- and 6-year-olds demonstrated more of a difference in vocalic duration than adults depending on syllable final voicing (Krause, 1982a). Upon first consideration, this finding seems counterintuitive in light of the cross-linguistic studies: If speakers must learn how to vary vocalic duration according to syllable-final voicing, then why would children show a greater effect? Buder and Stoel-Gammon (2002) answered this question by suggesting that vocalic length may inherently vary with syllable final voicing, and so be present from the onset of speech. According to this position, speakers of languages that lack a vowel-length distinction as a function of syllable final voicing actually must learn not to make the distinction. To support this position, they present data showing that children learning Swedish, which has only an attenuated vocalic-length distinction associated with syllable-final voicing, show the effect in their productions at age 24 months, but not at 30 months. Although this result supports their contention, there is at least one reason for skepticism: Vocalic length of syllables ending with voiced consonants produced by Swedish-learning 24-month-olds was generally greater than 400 ms, which is much longer than typical syllables. Be that as it may, Buder and Stoel-Gammon did not report on other aspects of production, and so it is impossible to obtain a picture of gestural *organization* for these syllables. Neither did they provide estimates of variability in vocalic duration for individual children, and so it is impossible to determine how consistent children were in their productions. Krause did provide estimates of variability, and reported greater variability in vocalic duration for children's than for adults' samples. However, Krause did not examine spectral properties in children's speech samples, and so we cannot evaluate gestural organization for these word tokens. The current study was conducted in order to investigate this larger question of how the several articulatory gestures involved in producing words with final stops differing in voicing are organized.

Most acoustic analyses done on adults' samples of words with voiced and voiceless final stops have focused only on the duration of the preceding vocalic portion, generally defined as the sum of the voiced initial transitions, the steady-state region, and any voiced final transitions. These studies have found that this portion is shorter before voiceless than before voiced final stops in adults' speech (e.g., Chen, 1970; Crowther and Mann, 1992; 1994; Flege and Port, 1981; House and Fairbanks, 1953; Peterson and Lehiste, 1960). Other studies have investigated $F1$ frequency at voicing offset, showing that it is higher before voiceless rather than voiced stops (e.g., Crowther and Mann, 1992; 1994; Summers, 1987). Taken together, the vocalic duration and $F1$ differences indicate that speakers abduct the vocal folds before completing the closing gesture when the final stop is voiceless, but close the vocal tract before voicing ends when the final stop is voiced.

Methods other than acoustic analysis have been used to study adults' productions of words differing in the voicing of final stops, and have uncovered other differences in gestural organization for these words. Most kinematic studies have been done using words with final bilabial stops only, likely due to the visibility of these gestures. Using that approach, Gracco (1994) found that both peak jaw lowering and peak jaw velocity occur sooner in the production of /æp/ than of /æb/. Summers (1987) showed similar results for the jaw lowering gesture, and also demonstrated that the jaw achieves a more open position before voiceless, rather than voiced, stops. However, the jaw remained in a relatively

stable position for longer periods of time preceding voiced stops. At the same time, Summers found no consistent difference for jaw raising gestures as a function of the voicing of the final stop. Of particular interest to the current acoustic analysis, Summers was able to correlate $F1$ change with jaw position and movement, but strongly so for medial syllabic portions only. For syllable margins, jaw movement and $F1$ frequency did not correlate as well. Specifically, Summers found that $F1$ at voicing offset was higher for voiceless than for voiced final stops, as expected, even though the jaw was in roughly the same position at voicing offset for both voicing conditions in samples from two of his three speakers. Nonetheless, $F1$ will be considered here as the best acoustic estimate of mandibular movement because of the well-documented fact that $F1$ is a function of the cross-sectional area of the front portion of the vocal tract (Schroeder, 1967; Stevens and House, 1955). The jaw and lips are the biggest determinants of this area, which probably explains Summers' results: The lips may have affected $F1$ frequency at voicing offset for two of his speakers, but he did not investigate lip movements. The lips (particularly the lower lip) contribute significantly to adults' closing gestures for bilabial stops (e.g., Green *et al.*, 2000), as studied by Summers, but they would not be expected to contribute much to closing for stops with lingual placements, as primarily studied here.

Raphael (1975) used physiological measures to examine adults' gestural organization for words with syllable-final voiced and voiceless consonants. He purposely avoided words in which the tongue was used for both vowel and final consonant constrictions. Instead, he examined EMG activity for the tongue (genioglossus) during vowel constrictions and for the lips (orbicularis oris and depressor anguli oris) during consonant constrictions in the production of words such as *lap* and *lab*. Results showed that EMG activity for the genioglossus muscle reached its peak at the same time, relative to the onset of voicing, for words with both voiced and voiceless final stops, but that the decay of that EMG activity was more rapid for words with voiceless final stops than for those with voiced final stops. EMG activity for muscles responsible for lip closure reached its peak at the same time relative to the decay of genioglossus EMG activity for words with both kinds of final stops. According to this gestural account, we might predict that there would be no difference in frequency of formants higher than $F1$ at voicing offset for words with voiced and voiceless final stops. However, this conclusion must be tempered mainly because Raphael's study did not involve words requiring tongue gestures for both vowel and final consonant constrictions. The present study examined words with alveolar and velar constrictions for final stops, so that the tongue was required for both vowel and consonant constrictions. The frequency of $F2$ at syllable center and at voicing offset was measured and was used to speculate as to whether there were differences in the coordination of lingual and vocal-fold gestures as a function of voicing or place of the final stop. In particular, we wanted to examine whether children's lingual gestures are similar to those of adults in the production of these words. Both MacNeilage and Davis (1991) and Nittrouer (1993) suggested that young children acquire mature patterns of vocal-tract opening and closing (accomplished primarily with the jaw) earlier than they acquire mature gestural patterns for other articulators. That suggestion was tested in this study by examining jaw gestures (through $F1$ measures) and tongue gestures (through $F2$ measures) in the production of words with voiced and voiceless final stops with several places of vowel and consonant constriction.

Another concern addressed by this study has to do with the organization of articulatory gestures for words when they are embedded in continuous speech. By far, most studies examining either the acoustic or the articulatory attributes of words with voiced and voiceless final stops have used tokens produced in isolation or in short, consistent carrier phrases. For adults, it may be that the articulatory score (and so, the acoustic consequences) differ when speakers produce words in real, meaningful contexts. For children, it may be that greater challenges are encountered in trying to organize gestures over longer, more involved utterances. For these reasons, speech samples were obtained in isolation and in sentences.

In summary, the current study sought to examine the gestural organization of monosyllabic words ending in voiceless or voiced final stops (particularly those with lingual constrictions) spoken by children and adults, both in isolation and in sentences, using acoustic measures. Measures made were the duration of the vocalic portion, and $F1$ and $F2$ frequencies at both voicing offset and syllabic center.

II. METHOD

A. Speakers

Eight speakers (four male and four female) in each of three age groups (adults, 7-year-olds, and 5-year-olds) participated. None of the speakers had ever been treated for a speech, language, or hearing problem. All speakers passed hearing screenings of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB HL. In addition, children were administered the Goldman–Fristoe Test of Articulation Sounds in Words Subtest (Goldman and Fristoe, 1986). All children, except one, were judged to produce all items in that subtest correctly. The one exception was a 5-year-old girl judged to be substituting /w/ for word-initial /ɹ/. Because none of the test items involved /ɹ/ production and this substitution is common for 5-year-olds, that was not considered to be a reason to exclude her.

B. Equipment and materials

All speech samples were recorded in a soundproof booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Labs Soundblaster 16-bit analog-to-digital converter. CSPEECHSP software (Milenkovic, 1997) was used both for recording speech samples and for the acoustic analyses. Speech samples were digitized at a sampling rate of 22.05 kHz using low-pass filtering with a high-frequency cutoff of 11.025 kHz.

Five minimal pairs made up the set of test items: *feet/feed*, *boot/booed*, *pick/pig*, *buck/bug*, and *cop/cob*. Two pairs had alveolar stop closures, two had velar closures, and one

TABLE I. Sample sentences for each word.

(1)	His *feet* were too big for his shoes.
(2)	Sam has to *feed* the dogs everyday after school.
(3)	Her new *boot* slipped off her foot.
(4)	The comedian was *booed* off the stage last night.
(5)	Susie likes to *pick* flowers on Sunday afternoons.
(6)	My grandfather has a *pig* on his farm.
(7)	Tim shot a *buck* on the first day of hunting season.
(8)	A lady *bug* is red and has black spots.
(9)	The man gave the *cop* the wallet he found in the street.
(10)	We had corn on the *cob* for dinner last night.

had bilabial closures. All words had initial consonants that permitted clear identification of voicing onset in the waveform. Vowels varied on the dimensions both of front/back and open/close. A set of pictures (5×5 in.) depicting each word was used to elicit word samples in isolation from children. A set of cards with the words printed on them served to elicit word samples in isolation from adults. Five sets of sentences were constructed for use, following the simple rules that each sentence should be about eight words long, the target word should not occur in phrase-final position, but the position of the word should differ across sets. A sample set of these sentences is presented in Table I. An adult female speaker with a midwestern dialect recorded these sentences onto tape.

C. Procedures

Hearing screenings were administered first. For children, the Goldman–Fristoe Test of Articulation was administered next. The order of collection of words in isolation and in sentences was alternated and randomized for each speaker, with the stipulation that a complete set of the ten words was collected in one condition before moving to the next. Five sets of words in each condition were collected. Words in isolation were elicited from children with picture cards and from adults with printed cards. Words in sentences were obtained by imitation of the sentences heard on tape. Two experimenters participated in the collection of these samples: one worked with the person speaking and one controlled the computer. As a result, samples in each condition were checked before moving to the next condition, and any samples judged to be unacceptable (due to extraneous noise or low amplitude) were immediately recorded again. In all, 100 samples were collected from each speaker: 2 conditions×10 words×5 samples.

D. Acoustic analysis

Each word was separated and saved to its own waveform file. Spectrograms of two productions of each word, in each context, were made so that general information about the acoustic form of each word would be available while making more fine-grained measurements. Figure 1 shows sample spectrograms of *buck* and *bug* spoken in isolation by an adult male. Several temporal and spectral measures were made on each speech token, following the procedures of Nittrouer (1993):

Duration of vocalic portion (*Dv*): time from the onset of voicing for the vocalic portion to either the offset of voicing (for words with voiceless final stops) or the point of vocal-tract closure (for words with voiced final stops). Arrows below the *x* axis in Fig. 1 show the ends of vocalic portions. These measures were obtained from the waveform. Cursors marking both the start and end of the vocalic portion were placed at zero crossings.

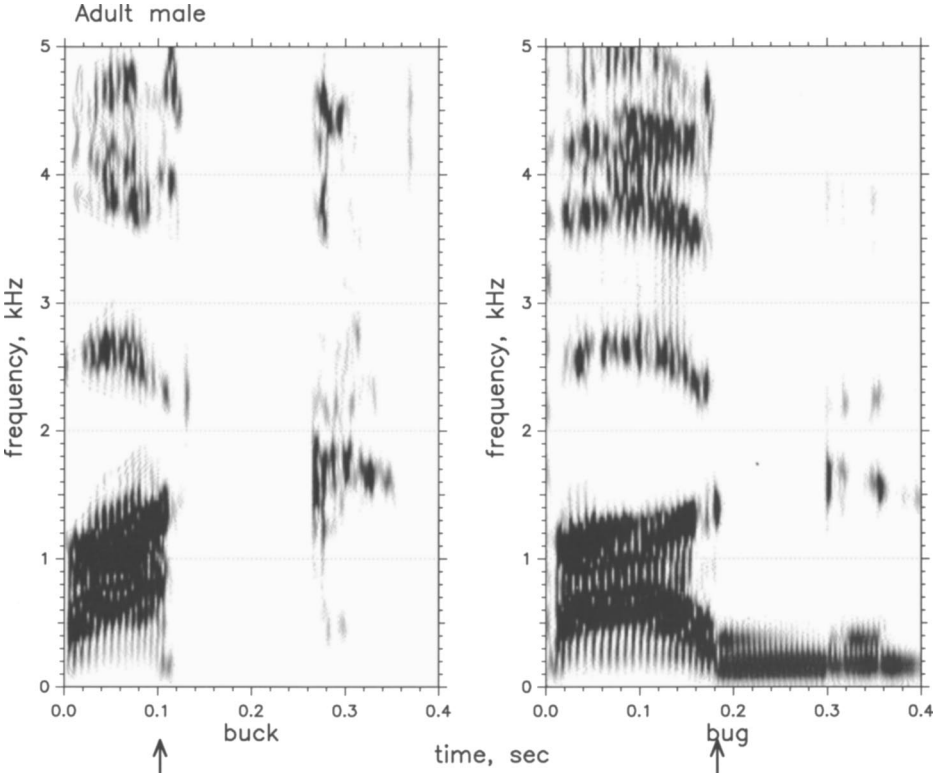


FIG. 1. Sample spectrograms of *buck* and *bug* spoken in isolation by an adult male. The arrows under the *x* axis indicate the points marked as *offset* in the acoustic analyses.

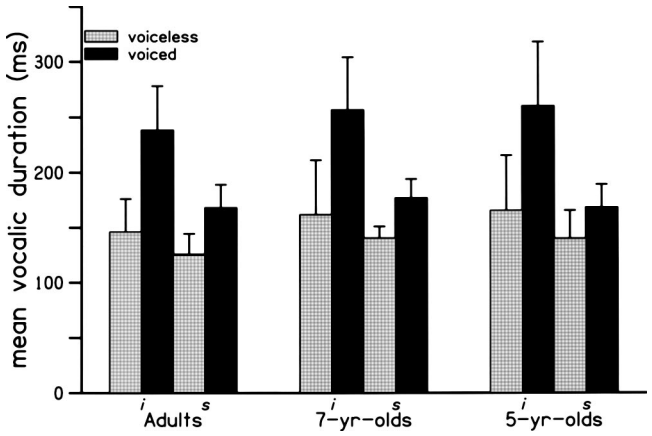


FIG. 2. Mean vocalic duration (*Dv*) (ms) for each age group, across all tokens of words with voiceless (*feet, pick, boot, buck*, and *cop*) and voiced (*feed, pig, booed, bug*, and *cob*) final stops, spoken in isolation and in sentences. Error bars represent standard deviations (SDs).

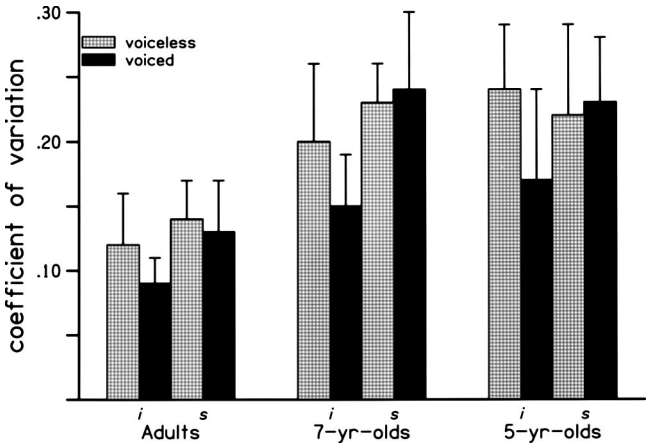


FIG. 3. Mean coefficients of variation (CVs) for *Dv* for each age group, across tokens of words with voiceless (*feet, pick, boot, buck*, and *cop*) and voiced (*feed, pig, booed, bug*, and *cob*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

F1 at voicing offset (F1off): $F1$ in the last pitch period. The onset and offset of the pitch period were marked at zero crossings. This measure, as well as all formant frequencies, were obtained using LPC analysis with 22 coefficients of the selected sample.

F2 at voicing offset (F2off): $F2$ for the last pitch period.

F1 at temporal center of the vocalic portion (F1center): $F1$ for three pitch periods at the temporal center of the vocalic portion.

F2 at temporal center of the vocalic portion (F2center): $F2$ for three pitch periods at the temporal center of the vocalic portion.

For each measure, a coefficient of variation (CV) was computed across tokens of each word spoken by each speaker by dividing the standard deviation (SD) by the mean. This provided an estimate of variability for individual speakers, and so allowed us to test the hypothesis that children are more variable in their productions than adults are.

Before measurements were made, laboratory staff and the first author discussed procedures for making these measurements, and practiced together to ensure that everyone was using the same procedures. Research assistants, including the second author, made all measurements. If there was a question about the locations of the onsets or offsets of voicing, research assistants consulted with each other or with the first author. The first or second author checked roughly 10 percent of all measurements to ensure that procedures were followed correctly. In all cases, they were.[1]

III. RESULTS

A. Vocalic duration

For the analysis *Dv*, results were collapsed across words with voiceless final stops (*feet, pick, cop, buck*, and *boot*) and across words with voiced final stops (*feed, pig, cob, bug*, and *booed*). Figure 2 shows mean *Dv* for words with voiceless and voiced final stops, spoken in isolation and in sentences. Overall, children and adults produced words with similar *Dv*'s. As predicted, it appears that *Dv*'s were shorter for words with voiceless, rather than voiced, final stops. Words with voiced final stops also appear to have had much shorter *Dv*'s when produced in sentences rather than in isolation. Mean *Dv*'s of words with voiceless final stops appear to have been somewhat shorter when produced in sentences instead of in isolation, but the magnitude of this difference is far less than that seen for words with voiced final stops. Across speakers there was a 75-ms difference for words with voiced final stops that were produced in isolation versus in sentences, and a 27-ms difference for words with voiceless final stops produced in the two contexts.

A three-way analysis of variance (ANOVA) with age as the between-subjects factor and context (isolation or sentence) and voicing as the within-subjects factors supported these observations. The main effect of age was not significant, but the main effects of both context and voicing were: for context, $F(1,21)=26.86$, $p<0.001$; for voicing, $F(1,21)=286.32$, $p<0.001$.[2] The only interaction that was significant was context×voicing, $F(1,21)=53.36$, $p<0.001$, supporting the observation that *Dv*'s for words with voiced final stops differed across contexts more than the *Dv*'s of words with voiceless final stops. Overall, it can be concluded that speakers of all ages produced words with similar *Dv*'s, that these *Dv*'s were shorter for words with voiceless rather than voiced final stops, and that when words with voiced final stops were spoken in sentences, *Dv* was substantially shorter than when words were spoken in isolation.

Figure 3 displays mean CVs for *Dv*, and shows a complicated pattern of results. In general, variability appears greater for the children's groups than for adults: that is, CVs are larger for children. For words spoken in isolation, it appears that speakers of all ages showed greater variability in *Dv* for words with voiceless, rather than voiced, final stops. For words spoken in sentences, variability is similar for words with voiceless and voiced final stops, although the relation between the means for the two conditions varies across speaker age. For adults, mean CV for words with voiceless final stops (spoken in sentences) is slightly greater than for words with voiced final stops. For 7- and 5-year-olds, mean CVs for words with voiced final stops (spoken in sentences) are slightly greater than for words with voiceless

final stops. Results of a three-way ANOVA provide help interpreting the patterns seen in Fig. 3. The main effect of age was significant, $F(2,21)=22.33$, $p<0.001$, as well as the main effect of voicing, $F(1,21)=11.15$, $p=0.003$. The first of these results is readily apparent from Fig. 3; the second result is harder to discern. Across speakers and contexts, mean CV was 0.19 for words with voiceless final stops and 0.17 for words with voiced final stops. The main effect of context was also significant, $F(1,21)=17.89$, $p<0.001$, reflecting the fact that across speakers and words ending in voiceless and voiced final stops, mean CV was 0.20 for words spoken in sentences and 0.16 for words spoken in isolation. Finally, the two-way interaction of context $\times$voicing was significant, $F(2,21)=8.23$, $p=0.01$. Judging from Fig. 3, it appears that this interaction can be accounted for by the patterns of variability exhibited by 7- and 5-year-olds and described above: When words were spoken in isolation, variability was greater for words with voiceless final stops; when words were spoken in sentences, variability was greater for words with voiced final stops.

A simple effects analysis, holding age constant, was also performed on these CVs to provide further insight into these results. For adults, no significant main effects or interactions were found, but the main effect of voicing was close to significant, $F(1,21)=3.56$, $p=0.073$, as well as the main effect of context, $F(1,21)=3.24$, $p=0.086$. For 7-year-olds, the effect of context was found to be significant, $F(1,21)=15.96$, $p<0.001$. This result indicates that 7-year-olds were more variable for words spoken in sentences than for those spoken in isolation. In addition, the context$\times$voicing interaction was close to significant for 7-year-olds, $F(1,21)=3.03$, $p=0.096$. For 5-year-olds, the main effect of voicing was significant, $F(1,21)=7.23$, $p=0.014$, as well as the context$\times$voicing interaction, $F(1,21)=6.32$, $p=0.02$. This finding supports the impression that variability was slightly greater for words with voiced, rather than voiceless, final stops when words were spoken in sentences, but when words were spoken in isolation, 5-year-olds showed decreased variability for words with voiced, rather than voiceless, final stops. Consequently, several conclusions may be drawn: (1) Children were more variable than adults; (2) All speakers were more variable in their productions of words with voiceless, rather than voiced, final stops; (3) This pattern of greater variability for words with voiceless, rather than voiced, final stops is mostly restricted to words spoken in isolation; and (4) 7-year-olds were more variable for words spoken in sentences than for those spoken in isolation.

B. F1off

Two word pairs (i.e., *feet/feed* and *boot/booed*) were not included in the analyses of *F1off* because the vocal tract remains in a fairly closed position throughout the production of these words, and so there is very little change in $F1$ across the word. For *cop* and *cob*, it was difficult to separate $F1$ from $F2$ in many of the children's samples. As a result, *F1off* was analyzed for *pick*, *pig*, *buck*, and *bug* only. Figure 4 displays mean *F1off* for words with voiced and voiceless final stops, for male and female speakers separately. As expected, *F1off* was higher for speakers in all three age groups

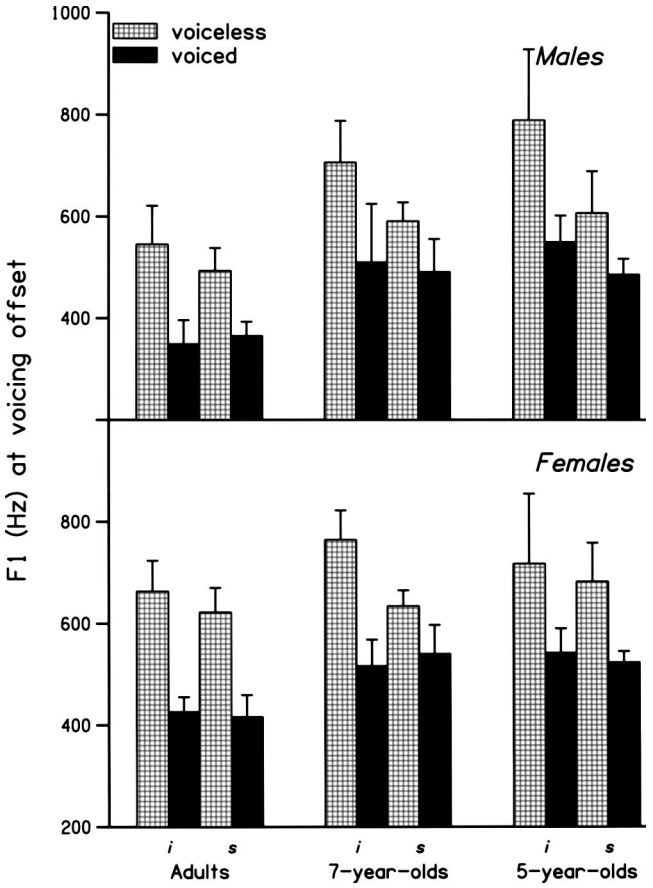


FIG. 4. Mean *F1off* (Hz) for males and females in each age group, for words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

for words ending in voiceless stops than for those ending in voiced stops. In addition, *F1off* specifically for words with voiceless final stops was lower when produced in sentence contexts rather than in isolation. This trend is particularly apparent for children's samples.

A four-way ANOVA, with age and sex as the between-subjects factor and context and voicing as the within-subjects factors, was done on these *F1off* frequencies. Speaker sex was not a variable that was of particular interest in this study, but any time spectral measures are examined it seems prudent to include sex as a factor. All four main effects were significant: age, $F(2,18)=14.92$, $p<0.001$; sex, $F(1,18)=5.21$, $p=0.035$; context, $F(1,18)=32.83$, $p<0.001$; and voicing, $F(1,18)=139.82$, $p<0.001$. The only interaction term that was significant was context$\times$voicing, $F(1,18)=27.26$, $p<0.001$. The main effects of age and sex simply reflect the fact that speakers with longer vocal tracts (i.e., adults and males) have lower $F1$ frequencies, and the main effect of voicing reflects the fact that *F1off* is lower for words with final voiced, rather than for voiceless, stops. The findings of a significant main effect of context and of a significant interaction for context$\times$voicing reflect the same trend: *F1off* is lower for words spoken in sentences than in isolation, and this trend is almost entirely due to *F1off* for words specifically with voiceless final stops being lower when produced in sentences, rather than in isolation. For words with voiced final stops, *F1off* is similar across the two

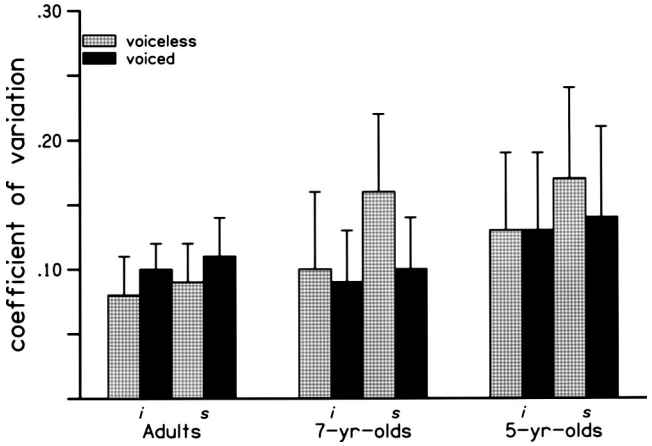


FIG. 5. Mean CVs for *F1off* for each age group, across tokens of words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

contexts. The fact that a significant three-way interaction of context×voicing×age was not found, as might be expected given that this trend seems particularly pronounced for children, is probably due to the fact that all groups showed the effect to some degree. Children simply demonstrated it to a greater extent. To test this suggestion, a simple effects analysis was run on *F1off*, looking at the effects of context and voicing for each age group separately. The context×voicing interaction did not reach statistical significance for adults, but did for both children's groups: 7-year-olds, $F(1,18) = 21.93$, $p < 0.001$; and 5-year-olds, $F(1,18) = 6.37$, $p = 0.021$. Children in both age groups had lower *F1off* frequencies for words with voiceless final stops when those words were spoken in sentences rather than in isolation. In articulatory terms, this result indicates that children's vocal tracts were closed more at voicing offset for these words when they were produced in sentences. This trend must have resulted from one of three possible reasons: (1) children never opened their mouths as much during the production of these words in sentence contexts as in isolation; (2) children began closing their vocal tracts sooner during word production in sentence contexts, compared to isolation; or (3) some combination of these two factors. The analysis of *F1center* will help resolve the issue.

Coefficients of variation were computed for *F1off* of each word used in the analysis above: *pick* and *buck* for the voiceless condition, and *pig* and *bug* for the voiced condition. The amount of variation in *F1off* serves as a metric of how consistently individual speakers coordinated vocal-fold abduction with jaw gestures in the production of these words. Mean CVs are shown in Fig. 5. There is a clear developmental decrease in the amount of variation associated with these measures. This age-related difference appears across voicing and contexts for 5-year-olds versus adults, but for 7-year-olds, CVs for *F1off* do not appear higher than those of adults except for words ending in voiceless stops spoken in sentences. In fact, both 7- and 5-year-olds displayed increased variability for words with voiceless final stops in sentences, compared to the other three conditions.

A three-way ANOVA was performed on CVs for *F1off*, with age as the between-subjects factor and voicing and context as within-subjects factors. Sex was not a factor here because there is no reason to predict that variability would differ for male and female speakers. The main effect of age was significant, $F(2,21) = 5.25$, $p = 0.014$, and the main effect of context approached significance, $F(1,21) = 4.16$, $p = 0.054$. These main effects indicate that children were more variable than adults in how they coordinated vocal-fold abduction with jaw gestures in the production of words with voiceless and voiced final stops, and that overall speakers were more variable when words were produced in sentences. But, this last effect is probably explained in large part by the increased variability found just for children's productions of words with voiceless final stops in sentences. To examine this suggestion, a simple effects analysis was done on these CVs for each age group separately. The term of interest was the context×voicing interaction. This interaction term was not significant for adults, indicating that adults showed similar differences in variability of *F1off* across voicing conditions for words spoken in isolation and in sentences. However, the context×voicing interaction was significant for 7-year-olds, $F(1,18) = 21.93$, $p < 0.001$, as well as for 5-year-olds, $F(1,18) = 6.37$, $p = 0.021$. These results support the suggestion that children were particularly variable in their attainment of *F1off* for words with voiceless final stops spoken in sentences.

C. F1center

This spectral measure effectively examined whether there was an age-related difference in the degree of jaw opening at the middle of the syllable, depending on the voicing of the final stop. Summers' (1987) finding that maximum excursion was greater for words with voiceless, rather than voiced, final stops led to the prediction that, at least for adults, *F1center* would be higher for words with voiceless, rather than voiced, final stops. We wanted to see if the same effect would be found for children. As with the analysis of *F1off*, this examination was performed only on measures from the *pick/pig* and *buck/bug* minimal pairs. Figure 6 shows mean *F1center* for male and female speakers in each group.

A four-way ANOVA was performed on these measures with age and sex as the between-subjects factors and context and voicing as the within-subjects factors. All four main effects were found to be statistically significant: age, $F(2,18) = 8.36$, $p = 0.003$; sex, $F(1,18) = 5.13$, $p = 0.036$; context, $F(1,18) = 11.06$, $p = 0.004$; and voicing, $F(1,18) = 21.12$, $p < 0.001$. In addition, the context×voicing interaction was significant, $F(1,18) = 11.54$, $p = 0.003$, suggesting that the difference in *F1center* across voicing conditions may have been attenuated for words spoken in sentences instead of in isolation. To examine whether children actually showed the difference in *F1center* predicted by Summers (1987), a simple effects analysis was done, holding age constant. All three age groups showed a significant effect of voicing: adults, $F(1,18) = 5.91$, $p = 0.026$; 7-year-olds, $F(1,18) = 5.07$, $p = 0.037$; and 5-year-olds, $F(1,18) = 10.74$, $p = 0.004$. It can thus be concluded that adults and children alike showed the difference in *F1center* predicted by Summers, at least in general. The simple effects analysis was also

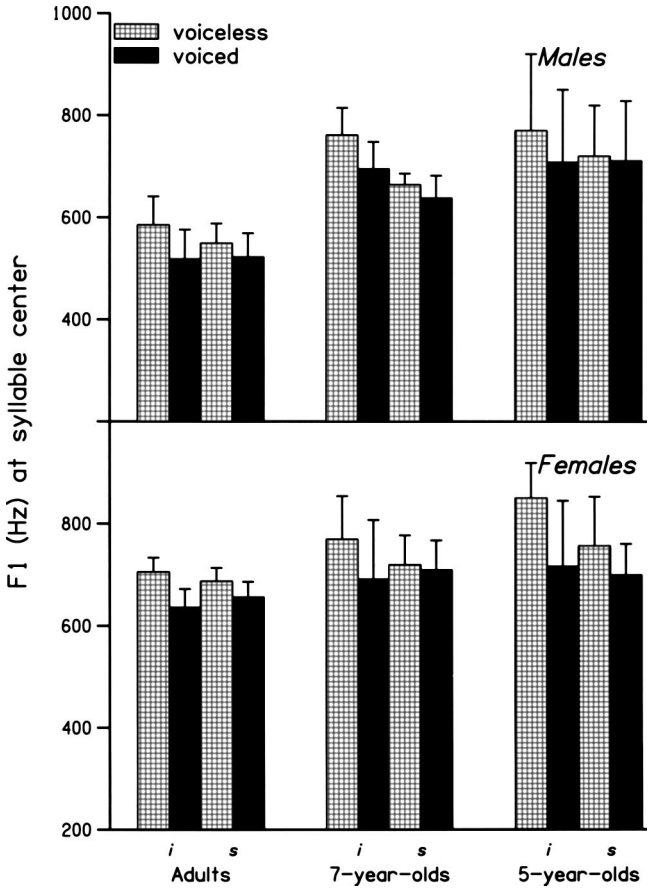


FIG. 6. Mean *F1center* (Hz) for males and females in each age group, for words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

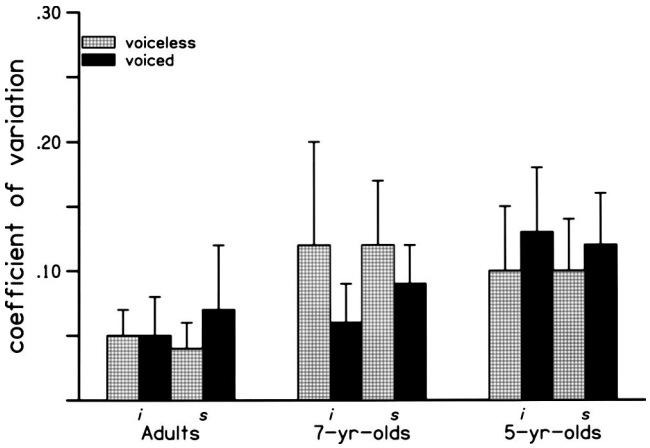


FIG. 7. Mean CVs for *F1center* for each age group, across tokens of words with voiceless (*pick* and *buck*) and voiced (*pig* and *bug*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

used to help determine whether this voicing-related difference was attenuated more for children than for adults when words were spoken in sentences rather than in isolation. To this end, the context×voicing interaction was examined for each age group separately. For adults, this interaction was not significant, and so we may conclude that adults showed the voicing-related difference in *F1center* to the same extent for words spoken in isolation and in sentences. However, the interaction was significant, or at least close to it, for both children's groups: 7-year-olds, $F(1,18)=3.99$, $p=0.061$; 5-year-olds, $F(1,18)=5.87$, $p=0.026$. Thus, there was an age-related difference in speakers' abilities to preserve the increase in *F1center* for words with voiceless final stops when producing words in sentences: The older the speaker, the better this increase was preserved. In articulatory terms this finding indicates that children do not completely maintain the difference in jaw excursions for words with voiced and voiceless final stops described by Summers when words are produced in sentences. This trend may explain, at least in part, the finding that *F1off* is lower when words with voiceless final stops are produced in sentences rather than in isolation for children's samples, but not for adults' samples. It may be that children never open the vocal tract as much as adults do for words with voiceless final stops produced in sentences. This issue is examined more closely in subsection D below.

As was done with *F1off*, CVs were computed for *F1center*. Figure 7 displays mean CVs for *F1center* and reveals several interesting findings when compared to Fig. 5 (which showed CVs for *F1off*). For adults, variability in *F1center* across all conditions is roughly only half as large as what it was for *F1off*. For both children's groups, there is a reduction in variability for *F1center* compared to *F1off*, but this reduction is not as great as for adults. For 7-year-olds, the enhanced variability observed for *F1off* for words with voiceless final stops produced in sentences is seen here for *F1center* as well, but is also observed for these same words produced in isolation. These 7-year-olds clearly demonstrated variability in jaw gestures during the production of words with voiceless stops.

A three-way ANOVA was performed on these CV measures, with age as the between-subjects factor and voicing and context as within-subjects factors. The main effect of age was significant, $F(2,18)=9.05$, $p=0.001$, as well as the age ×voicing interaction, $F(2,21)=6.36$, $p=0.007$. This interaction is likely due to the finding that 7-year-olds showed greater variability when producing words with voiceless, rather than voiced, final stops.

D. Age-related differences in vocal-tract closing gestures for words with voiceless final stops

An analysis was undertaken to examine why *F1off* was lower for children's samples of words with voiceless final stops spoken in sentences, rather than in isolation. At issue was whether this trend could be entirely attributed to the finding that children seemed to be constrained in how much they opened their vocal tracts for words with voiceless final stops when those words were spoken in sentences. To explore this question, the difference between *F*1 frequency at syllable center and at voicing offset was computed on the mean frequencies of words with voiceless final stops (i.e., *pick* and *buck*) for each speaker, for words produced in isolation and in sentences separately. The means for these difference scores are shown in Table II for each age group.

A simple effects analysis was performed on these difference scores to examine the effect of context, holding age constant. The effect of context was not significant for adults,

TABLE II. Mean differences, in Hz, between $F1$ at syllable center and at voicing offset of words with voiceless final stops, spoken in isolation and in sentences. Standard deviations are given in parentheses.

	Isolation	Sentences
Adults	41	60
	(39)	(34)
7-year-olds	29	79
	(96)	(48)
5-year-olds	20	102
	(63)	(40)

was close to significant for 7-year-olds, $F(1,21)=3.48$, $p=0.076$, and was significant for 5-year-olds, $F(1,21)=9.49$, $p=0.006$. Thus, it can be concluded that for children, but not for adults, the change in $F1$ frequency between syllable center and voicing offset was greater for words with voiceless final stops produced in sentences, rather than in isolation. In articulatory terms this means that children began closing the vocal tract before the offset of voicing for words with voiceless final stops produced in sentences. This pattern of articulatory organization can be seen in Fig. 8. Although this figure shows spectrograms of speech samples from just one child, these patterns are typical of what all children did when producing words in sentences. Clearly the $F1$ trajectory for *buck* is different from what is found in Fig. 1, showing spectrograms from an adult's samples. For the most part, this articulatory pattern was not found for children's word samples obtained in isolation, although a few children exhibited the pattern even for words in isolation (note the large SDs in Table II for children's samples in isolation). This articulatory pattern was not observed in adults' samples, regardless of whether samples were obtained in isolation or in sentences.

E. F2off

If the vocal folds are abducted earlier relative to the closing gesture for words with voiceless, rather than voiced, final stops, then $F2$ will differ at voicing offset for these voicing conditions, at least when the tongue forms the consonant closure. Furthermore, $F2$ will either rise or fall going into closure, depending on the locations of vocalic and consonantal constrictions. For these analyses, only words in which $F2$ was presumed to be rising at the end of the syllable (i.e., words with relatively low vocalic $F2$ and relatively high $F2$ near the consonantal constriction) were used. To do otherwise would have severely constrained the likelihood of obtaining a statistically significant difference in *F2off* as a function of final-stop voicing, if one truly exists: In some cases *F2off* would be lower for voiced than voiceless stops; in other cases it would be higher. Accordingly, only the minimal pairs *pick/pig*, *buck/bug*, and *boot/booed* were used in analyses of *F2off*. Figure 9 shows mean *F2off* frequencies, and supports the suggestion that *F2off* is higher for words ending with voiced stops.

A four-way ANOVA was performed on *F2off* with age and sex as the between-subjects factors, and context and voicing as the within-subjects factors. Both between-subjects main effects were found to be statistically significant, as expected: for age, $F(2,18)=29.69$, $p<0.001$; for sex, $F(1,18)=37.01$, $p<0.001$. The main effect of voicing was also significant, $F(1,18)=55.50$, $p<0.001$, indicating that *F2off* was higher preceding voiced rather than voiceless stops. This finding demonstrates that the vocal folds were abducted sooner relative to the lingual closing gesture for words with voiceless, rather than voiced, final stops. The main effect of context was not significant, nor were any of the interactions.

As was done for other measures, CVs were computed for *F2off*, and means are shown in Fig. 10. Both groups of children appear to have been more variable than adults in their attainment of *F2off*. Unlike measures of variability for *Dv* and $F1$ (at both voicing offset and syllable center), 7-year-olds seem to have been as variable as 5-year-olds for *F2off* across voicing conditions and contexts.

A three-way ANOVA with age as the between-subjects

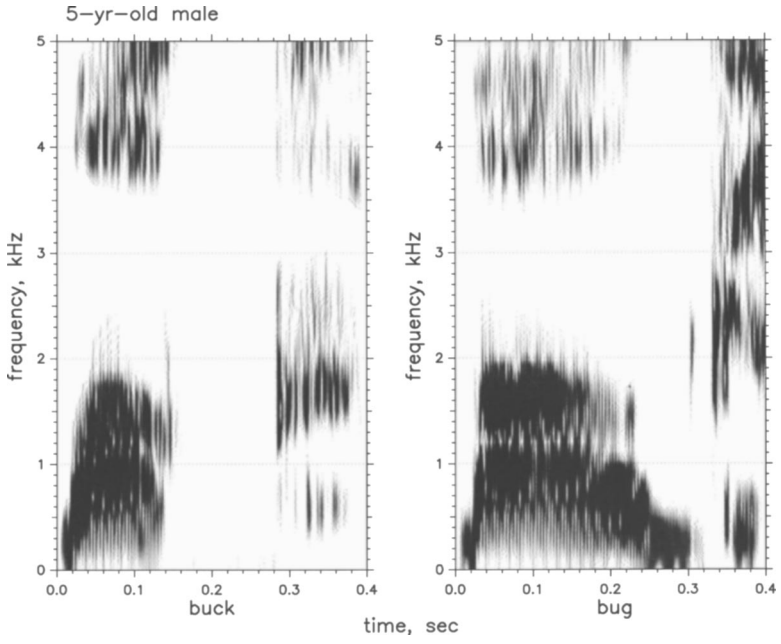


FIG. 8. Sample spectrograms of *buck* and *bug* spoken by a 5-year-old boy, showing that vocal-tract closing can be seen for *buck*.

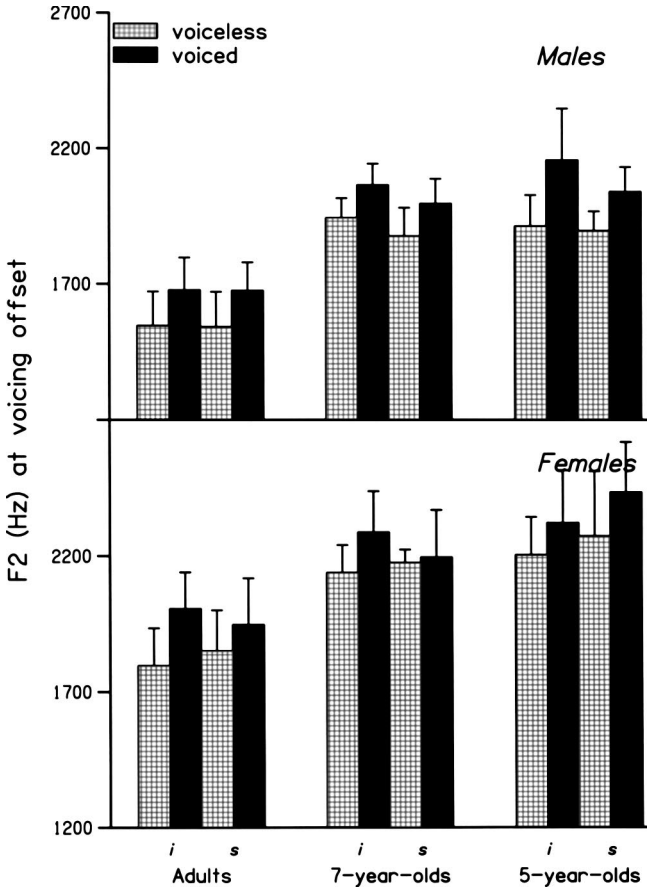


FIG. 9. Mean *F2off* (Hz) for males and females in each age group, for words with voiceless (*pick, buck,* and *boot*) and voiced (*pig, bug,* and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

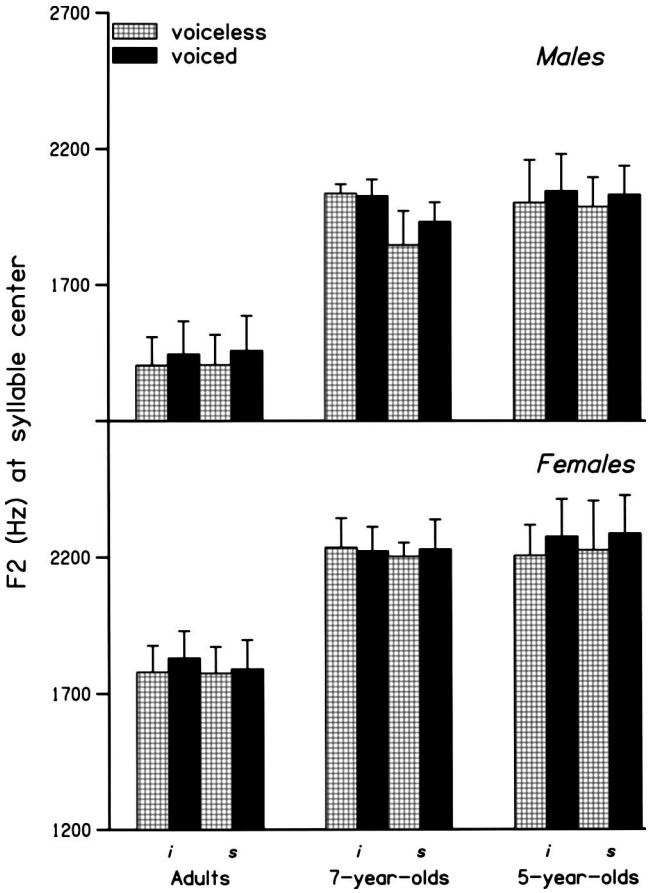


FIG. 11. Mean *F2center* (Hz) for males and females in each age group, for words with voiceless (*pick, buck,* and *boot*) and voiced (*pig, bug,* and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

factor and context and voicing as the within-subjects factors was performed. Indeed, the main effect of age was found to be significant, $F(2,21)=11.50$, $p<0.001$, supporting the suggestion that children were more variable than adults. In addition, the main effect of context was significant, $F(1,21)=8.18$, $p=0.009$. This finding reflects the fact that variability for *F2off* was greater for words spoken in sentences rather than in isolation: Across speakers CV was 0.087 for words in sentences versus 0.075 for words in isolation. This result may not reflect a simple increase in variability for words produced in sentences, but instead may indicate that speakers adjusted the precise place of stop closure depending on what gesture was required for the next word in the sentence. The failure to find an age×context interaction indicates that this cross-word coarticulation was no greater in children's than in adults' samples. Similarly, Nittrouer (1993) reported that coarticulation across word boundaries was no greater in children's than in adults' samples.

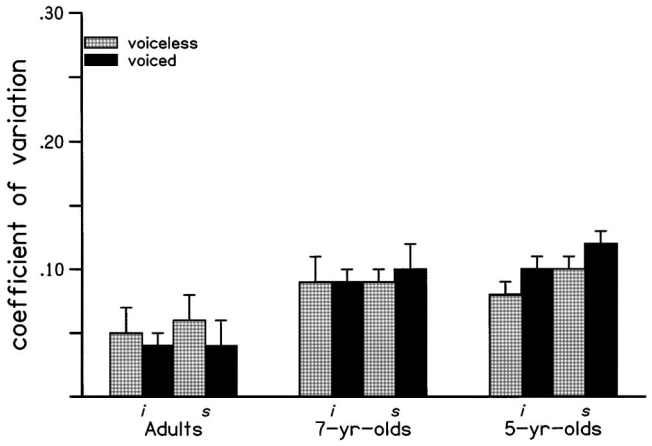


FIG. 10. Mean CVs for *F2off* for each age group, across tokens of words with voiceless (*pick, buck,* and *boot*) and voiced (*pig, bug,* and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

F. F2center

Figure 11 shows group means for *F2center*. It appears from this figure that the only factors that affected *F2center* were age and speaker sex. In particular, frequency of *F2center* seems to be largely unaffected by the voicing of the final stop. To evaluate these observations, a four-way ANOVA was performed. The main effects of age and sex were indeed significant: age, $F(2,18)=69.68$, $p<0.001$; and sex, $F(1,18)=37.01$, $p<0.001$. Contrary to impressions from Fig. 11, the effect of voicing was also significant, $F(1,18)=13.43$, $p=0.002$. However, the magnitude of this difference was quite small: across speakers, contexts, and words, *F2center* was just 40 Hz higher in words with voiced final stops than in those with voiceless final stops. This contrasts greatly with the 160-Hz difference observed for *F2off*.

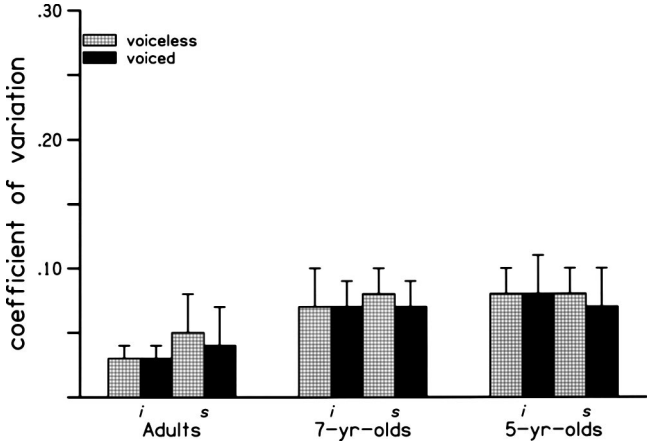


FIG. 12. Mean CVs for *F2center* for each age group, across tokens of words with voiceless (*pick, buck,* and *boot*) and voiced (*pig, bug,* and *booed*) final stops, spoken in isolation and in sentences. Error bars represent SDs.

As done with all measures, CVs were computed for *F2center*. Figure 12 shows these values. Here, the only effect that appears to be significant is age, and a three-way ANOVA (age×context×voicing) confirmed that impression: Only the effect of age was significant, $F(1,21)=12.41$, $p<0.001$. As with *F2off*, 7- and 5-year-olds appear to have been similarly variable, but more variable than adults.

G. Tongue fronting for alveolar stops

One stated objective of this study was to examine whether there was evidence of a greater synergy between tongue gestures required for vowel and consonant production in children's than in adults' samples. Specifically, it was thought that young children might front the tongue more than adults in anticipation of upcoming alveolar stops. Precisely because this hypothesis was to be tested, a word pair that consisted of the back vowel /u/ and an alveolar stop was used: *boot* and *booed*. In the absence of anticipatory tongue fronting, we would expect *F2center* in *boot* and *booed* to be lower than in *buck* and *bug*. For example, Peterson and Barney (1952) found *F*2 in /u/ to be 300–400 Hz lower than *F*2 in /ʌ/ for men, women, and children. Table III shows mean *F2center* for *boot, buck, booed*, and *bug* spoken in isolation

TABLE III. Mean *F*2 frequency (Hz) at syllable center across words spoken in isolation. Standard deviations are given in parentheses.

	Buck	Boot	Bug	Booed
Males				
Adults	1182	1245	1185	1235
	(68)	(186)	(98)	(223)
7-year-olds	1664	1834	1578	1824
	(164)	(213)	(77)	(94)
5-year-olds	1639	1635	1605	1667
	(180)	(296)	(98)	(303)
Females				
Adults	1491	1653	1532	1623
	(98)	(174)	(97)	(164)
7-year-olds	1794	2127	1823	2032
	(80)	(165)	(188)	(109)
5-year-olds	1833	1863	1819	2073
	(136)	(266)	(145)	(409)

TABLE IV. Mean *F*2 frequency (Hz) at syllable center across words spoken in sentences. Standard deviations are given in parentheses.

	Buck	Boot	Bug	Booed
Males				
Adults	1198	1296	1223	1279
	(70)	(197)	(85)	(291)
7-year-olds	1474	1711	1554	1634
	(82)	(219)	(123)	(104)
5-year-olds	1582	1642	1635	1667
	(87)	(278)	(56)	(287)
Females				
Adults	1426	1635	1478	1573
	(107)	(126)	(114)	(104)
7-year-olds	1740	2152	1829	2017
	(91)	(66)	(89)	(174)
5-year-olds	1849	2065	1889	2043
	(131)	(303)	(119)	(364)

for all speaker groups in this study. Table IV shows *F2center* for these words spoken in sentences. These tables indicate that *F2center* was actually higher for *boot* and *booed* than for *buck* and *bug*, but not only for children's groups. Adults showed this effect as well. Presumably this was due to anticipatory tongue fronting, which was predicted, but no evidence of an age-related difference in the amount of this tongue fronting was found. This last result differs from predictions.

IV. DISCUSSION

The goal of the study reported here was to use acoustic measures to examine the organization of articulatory gestures for words with voiceless and voiced final stops, produced by adults and children in isolation and in sentences. In general, it was found that children distinguished between words with voiceless and voiced final stops in the same ways that adults do. But, children were more variable in their productions, and children's organization of jaw and vocal-fold gestures for words with voiceless final stops produced in sentences differed from their organization of these gestures for words produced in isolation and from the organization of adults' gestures for these words (either in isolation or in sentences).

A. Dv

The durations of the vocalic word portions were similar across tokens produced by speakers of all ages, and these portions were shorter for words with voiceless final stops than for words with voiced final stops. For words with voiced final stops, there was a significant difference in *Dv* depending on whether the word was spoken in isolation or in a sentence. The fact that this particular context effect was attenuated for words with voiceless final stops suggests that there may be some limit to how short the vocalic portion of a word can realistically be: it simply takes time to open the vocal tract as needed for the vowel being produced. It may be that speakers operate near this effective limit when they produce monosyllables with voiceless final stops, regardless of whether they are produced in isolation or in sentences.

At the same time, variability in *Dv* was greater for words with voiceless, rather than voiced, final stops. This observa-

tion agrees with the notion that the coordination of gestures involved in producing words with voiceless final stops is likely more difficult than the coordination of gestures involved in producing words with voiced final stops. For words with voiceless final stops, the speaker must abduct the vocal folds at some point in the vocal-tract opening/closing gesture. Accomplishing this goal at precisely the same point across words surely requires great skill. On the other hand, obtaining consistency in *Dv*'s across tokens of words with voiced final stops would seem easier. Voicing offset for words with voiced final stops was defined here as the point at which the vocal tract reaches closure, and so *Dv* effectively measures only the duration of the opening/closing gesture. Nonetheless, 5-year-olds had more difficulty than adults achieving consistency in *Dv*, regardless of whether the goal was simply to produce equally timed opening and closing gestures (as in words with voiced final stops), or to abduct the vocal folds at a precise point in the opening/closing gesture (as in words with voiceless final stops). This enhanced difficulty was observed for 5-year-olds regardless of whether words were spoken in isolation or in sentences. Seven-year-olds, on the other hand, showed more consistency in the timing of these gestures when they were producing words in isolation. When they had the more difficult task of orchestrating the gestural score for the production of an entire sentence, however, they showed increased variability. Apparently children are fine-tuning their speech production skills past the age of 7 years.

It is interesting to compare adults' results for *Dv* across contexts with the cross-linguistic data. In this study adults demonstrated exactly a 100-ms difference in *Dv* for words with voiceless and voiced final stops spoken in isolation (using means across all voiceless and voiced tokens), which is what others have reported for English speakers (e.g., Chen, 1970; Crowther and Mann, 1994). For words spoken in sentences in this study, the voicing effect diminishes to 40 ms, which is similar to differences reported for non-native English speakers producing English words in isolation or in short, consistent carrier phrases (e.g., Crowther and Mann, 1992; 1994; Flege, Munroe, and Skelton, 1992; Flege and Port, 1981).[3] Presumably, in the natural, running speech that non-native speakers customarily hear, the difference in *Dv* for words with voiced and voiceless final stops is closer to the 40 ms measured in this study for words in sentences. Perhaps non-native English speakers imitate exactly what they hear. The question then becomes: Why do native English-speaking adults elongate *Dv* before voiced stops when producing words in isolation? That question is not answerable with these data.

The results reported here do not agree with those of Krause (1982a), who reported that young children showed a *greater* difference than adults in *Dv* for words spoken in isolation depending on the voicing of the final stop. In Krause's study the mean difference between voicing conditions was 97 ms for 6-year-olds and 60 ms for adults. The reason for this discrepancy in results, however, probably has more to do with Krause's findings for adults than with those for children. Adults' mean *Dv* for words with voiceless stops was 209 ms in Krause's study, considerably longer than the 157 ms obtained in the current study (for words spoken in isolation) or the 146 ms reported by Chen (1970). At the same time, Krause reported a mean *Dv* of 269 ms for words with voiced final stops spoken by adults, which is similar to the 251-ms mean found in the current study (for words spoken in isolation) and the 238-ms mean reported by Chen. Therefore, it is not that children's differences in *Dv* for words with voiceless and voiced final stops were unusually large in Krause' study, but rather that adults' differences were somewhat small. In particular, adults in that study produced words with voiceless final stops that had rather long *Dv*s.

B. F1off

The first thing to be said about *F*1 frequency at voicing offset is that its utility in aiding the listener make decisions about the voicing of syllable-final stops would appear to be limited because it does not vary as a function of voicing when the preceding vowel is close. Nonetheless, examination of *F1off* informs us about gestural organization in the production of consonant–vowel–stop sequences when we examine it for words with open vowels.

An age-related difference in gestural organization that emerges from the analysis of *F1off* is that children begin closing the vocal tract before the cessation of voicing for voiceless final stops—especially when trying to organize articulatory gestures over the length of a sentence. Adults, on the other hand, abduct the folds before they begin closing the vocal tract whether they are producing words with voiceless final stops in isolation or in sentences. While children are somewhat restricted in the extent to which they open their vocal tracts (i.e., lower their jaws) when producing words with voiceless final stops in sentences, rather than in isolation, this finding cannot completely account for the age-related difference in patterns of *F1off* across contexts. After reaching maximum jaw opening, adults apparently maintain stable jaw positions until they abduct their vocal folds, as Summers (1987) reported. Children begin to raise their jaws, and this effect is more pronounced for words spoken in sentences rather than in isolation. This finding mandates revision of the conclusion of Nittrouer (1993) that by 3 years of age children have acquired mature patterns of jaw movements. At least for some syllable shapes, it appears that children as old as 7 years have not completely mastered mature jaw patterns. Thus, this finding is one specific example of the suggestion that the emergence of mature gestural patterns is not uniform, that instead children attain mature patterns for some word forms sooner than for others (e.g., Green *et al.*, 2000; Nittrouer, 1993).

Regarding variability, *F1off* was less variable for all speakers than was *Dv*. The reason for this enhanced consistency may have to do with the fact that *Dv* can be influenced by several articulatory parameters, such as how rapidly or slowly the opening gesture is made, how long any steady-state vocalic portion is, and when vocal-fold abduction occurs. However, *F1off* is determined only by the degree of vocal-tract openness at the time the vocal folds are abducted. Apparently speakers tolerate some variability in the several parameters that affect *Dv*, but execute the vocal-fold abduction gesture at a relatively stable point in the opening/closing

gesture across tokens of each word. At the same time, the finding of increased variability in *F1off* for children's productions of words with voiceless final stops indicates that their coordination of vocal-fold abduction and jaw gestures for these words was particularly unstable.

C. F1center

The results of this study replicate those of Summers (1987): speakers open the vocal tract more for words with voiceless, rather than voiced, final stops, and so *F1center* is higher for words with voiceless final stops. Although this gestural pattern was attenuated somewhat for children producing words in sentences, it was nonetheless found. Regarding variability, it was found that adults more consistently achieved the same *F1center* than they achieved the same *F1off*. This finding indicates that adults were more consistent in how they organized and produced vocal-tract gestures over the first part of the word than in how they did so over the latter portion of the word. Children did not always show this consistency in gestural organization for early word portions. In particular, 7-year-olds showed greater variability for *F1center* of words with voiceless, rather than voiced, final stops.

It is interesting that the context-related changes observed for the temporal (*Dv*) and spectral (*F1off* and *F1center*) measures are uncorrelated. That is, *Dv* was shorter for words spoken in sentences rather than in isolation, but primarily for words with voiced final stops. On the other hand, *F1off* and *F1center* differed across contexts for children's samples, but only for words with voiceless final stops. Thus, the changes in *Dv* associated with context did not affect spectral measures. In articulatory terms this means that changes in vocalic duration *per se* did not influence the articulatory gestures themselves, or the organization of these gestures.

D. F2off

Results for *F2off* reveal that *F*2 frequency at voicing offset can provide information regarding the voicing of syllable-final consonants. For the words analyzed here, with rising *F*2 at syllable offset, *F2off* was generally higher for voiced than for voiceless final stops. Although not reported, a similar trend would be expected and was observed in casual inspection of words with falling *F*2 at syllable offset. For example, across all speakers *F2off* was 2687 Hz for *feet* spoken in isolation and 2461 Hz for *feed* spoken in isolation. Variability in children's samples was similar for *F2off* and *F1off* (except that 7-year-olds showed variability comparable to that of 5-year-olds in all conditions for *F2off*), but adults demonstrated decreased variability for *F2off* compared to *F1off*. Particularly for adults' speech, *F2off* appears to provide very reliable information about the voicing of the final stop.

E. F2center

One predicted finding that was not observed had to do with tongue gestures. It had been predicted that children might demonstrate greater synergy between tongue gestures required for vowel and consonant production than adults. This could have shown up as greater tongue fronting in anticipation of the alveolar stops in *boot* and *booed* in children's than in adults' samples. If present, the acoustic consequence of this gestural pattern would have been a greater difference between *F2center* for *buck/bug* and *F2center* for *boot/booed* in children's than in adults' samples. However, the degree of tongue fronting was similar for adults and children. At the same time, children were more variable in their attainment of *F2center* than adults were.

As was found with *F*1 measures, there was no effect of shortened vocalic segments for words with voiced final stops spoken in sentences, rather than in isolation, on F2 measures. Neither *F2off* nor *F2center* showed a significant context effect.

In summary, several conclusions can be drawn from these analyses. First, children as old as 7 years of age still organize their gestures for the productions of words differently from adults for some syllable shapes. Second, children are generally more variable in their execution of linguistic gestures than are adults. Overall, learning to coordinate the various gestures involved in producing speech with appropriately timed events is a difficult task that extends well into childhood. Third, the acoustic correlates of syllable-final voicing are attenuated somewhat when words are produced in sentences, rather than in isolation. In general, this finding serves as a reminder that we must be careful about generalizing results obtained for speech samples produced in isolation to our understanding of speech produced in natural contexts. A final conclusion to emerge from these data is that the acoustic correlates of speech production are spread throughout the word. This finding highlights a fact long understood by speech scientists, and yet frequently overlooked in applications to technology and clinical work: there are no discrete acoustic segments that correspond to linguistic units in the speech stream.

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, to Susan Nittrouer. The contributions of Lesley Larive and Tammy Mertes to the collection and analysis of these stimuli are gratefully acknowledged.

[1]Because the methods of acoustic analysis were so regimented there could be no computation of inter-rater reliability. As long as each experimenter adhered strictly to the outlined procedures the measurements obtained across experimenters were identical. All indications were that experimenters adhered to those procedures. Thus, inter-rater reliability was effectively 1.0.

[2]Although the alpha level of 0.05 is typically set, many investigators recognize the potential interest of ''marginally'' significant statistical tests (i.e., those with *p* values slightly above 0.05). For that reason, all tests with resulting *p* values of less than 0.10 will be reported throughout this paper. If an exact *F*- or *t* ratio is not given, it can be assumed that the value had an associated *p* of greater than 0.10.

[3]One study (Flege and Port, 1981) did report voicing differences for *Dv* in samples from English-speaking adults of just 40 ms. Interestingly, the carrier phrase in that study (''I say_____again to Bob'') was slightly longer than others, which are commonly just three words (e.g., ''Say_____again.'')

Buder, E. H., and Stoel-Gammon, C. (**2002**). ''American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing,'' J. Acoust. Soc. Am. **111**, 1854–1864.
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. A. (**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. A. (**1994**). ''Use of vocalic cues to consonant voicing and native language background: The influence of experimental design,'' Percept. Psychophys. **55**, 513–525.
Davis, B. L., and MacNeilage, P. F. (**1990**). ''Acquisition of correct vowel production: A quantitative case study,'' J. Speech Hear. Res. **33**, 16–27.
Donahue, M. (**1986**). ''Phonological constraints on the emergence of two-word utterances,'' J. Child Lang **13**, 209–218.
Ferguson, C. A., and Farwell, C. B. (**1975**). ''Words and sounds in early language acquisition,'' Language **51**, 419–439.
Flege, J. E. (**1989**). ''Chinese subjects' perception of the word-final English /t/–/d/ contrast: Performance before and after training,'' J. Acoust. Soc. Am. **86**, 1684–1697.
Flege, J., and Port, R. (**1981**). ''Cross-language phonetic interference: Arabic to English,'' Lang Speech **24**, 125–146.
Flege, J. E., Munro, M. J., and Skelton, L. (**1992**). ''Production of the word-final English /t/–/d/ contrast by native speakers of English, Mandarin, and Spanish,'' J. Acoust. Soc. Am. **92**, 128–143.
Goldman, R., and Fristoe, M. (**1986**). *Goldman–Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).
Goodell, E. W. (**1991**). ''Gestural organization in the speech of 22- to 32-month-old children,'' Ph.D. dissertation, University of Connecticut.
Gracco, V. L. (**1994**). ''Some organizational characteristics of speech movement control,'' J. Speech Hear. Res. **37**, 4–27.
Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (**2000**). ''The physiologic development of speech motor control: Lip and jaw coordination,'' J. Speech Lang. Hear. Res. **43**, 239–255.
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.
Hoole, P., Nguyen-Trong, N., and Hardcastle, W. (**1993**). ''A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data,'' Lang Speech **36**, 235–260.
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.
Katz, W. F., Kripke, C., and Tallal, P. (**1991**). ''Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data,'' J. Speech Hear. Res. **34**, 1222–1232.
Krause, S. E. (**1982a**). ''Developmental use of vowel duration as a cue to postvocalic stop consonant voicing,'' J. Speech Hear. Res. **25**, 388–393.
Krause, S. E. (**1982b**). ''Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults,'' J. Acoust. Soc. Am. **71**, 990–995.
MacNeilage, P. F., and Davis, B. L. (**1991**). ''Acquisition of speech production: Frames, then content,'' in *Attention and Performance XIII. Motor Representation and Control*, edited by M. Jeannerod (Erlbaum, New York).
Menn, L. (**1978**). ''Phonological units in beginning speech,'' in *Syllables and Segments*, edited by A. Bell and J. B. Hooper (North-Holland, New York).
Milenkovic, P. (**1997**). CSPEECHSP (University of Wisconsin, Madison).
Nittrouer, S. (**1993**). ''The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study,'' J. Speech Hear. Res. **36**, 959–972.
Oller, D. K. (**1980**). ''The emergence of the sounds of speech in infancy,'' in *Child Phonology: Production*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York).
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.
Piske, T. (**1997**). ''Phonological organization in early speech production: Evidence for the importance of articulatory patterns,'' Speech Commun. **22**, 279–295.
Raphael, L. J. (**1975**). ''The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English,'' J. Phonetics **3**, 25–33.
Schroeder, M. R. (**1967**). ''Determination of the geometry of the human vocal tract by acoustic measurements,'' J. Acoust. Soc. Am. **41**, 1002–1010.
Smith, A., and Goffman, L. (**1998**). ''Stability and patterning of speech movement sequences in children and adults,'' J. Speech Lang. Hear. Res. **41**, 18–30.
Stevens, K. N., and House, A. S. (**1955**). ''Development of a quantitative description of vowel articulation,'' J. Acoust. Soc. Am. **27**, 484–493.
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.
Sussman, H. M., Duder, C., Dalston, E., and Cacciatore, A. (**1999**). ''An acoustic analysis of the development of CV coarticulation: A case study,'' J. Speech Lang. Hear. Res. **42**, 1080–1096.
Tuller, B., and Kelso, J. A. S. (**1984**). ''The timing of articulatory gestures: Evidence for relational invariants,'' J. Acoust. Soc. Am. **76**, 1030–1036.
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.
Waterson, N. (**1971**). ''Child phonology: A prosodic view,'' J. Ling. **7**, 179–211.

Enhancing intelligibility of narrowband speech with out-of-band noise: Evidence for lateral suppression at high-normal intensity

James A. Bashford, Jr.,[a] Richard M. Warren, and Peter W. Lenz
Department of Psychology, University of Wisconsin—Milwaukee, Post Office Box 413, Milwaukee, Wisconsin 53211

(Received 12 April 2004; revised 7 September 2004; accepted 29 October 2004)

Previous studies have shown that the intelligibility of filtered speech can be enhanced by filling stopbands with noise. The present study found that this enhancement occurred only when speech intensity was sufficiently high to degrade performance. Intelligibility decreased by about 15% when narrowband speech was increased from 45 to 65 dBA (corresponding to broadband speech levels of about 60 and 80 dBA), and decreased by 20% at a level of 75 dBA. However, when flanking bands of low-pass and high-pass filtered white noise were added at spectrum levels of −40 to −20 dB relative to the speech, intelligibility of the 75-dBA speech band increased by about 13%. Additional findings confirm that this enhancement of intelligibility depends upon out-of-band stimulation, in agreement with theories proposing that lateral suppressive interactions extend the dynamic range of intensity coding by counteracting effects of auditory-nerve firing-rate saturation at high signal levels. [DOI: 10.1121/1.1835513]

PACS numbers: 43.71.Pc, 43.71.Es, 43.71.An [DOS] Pages: 365–369

I. INTRODUCTION

The ability to perceive speech accurately at high intensities poses an interesting problem for hearing theory. There is evidence that the majority of auditory-nerve (AN) fibers reach their firing-rate limits at conversational speech levels of about 65 dB, and at higher intensities are unable to provide a rate-based encoding of the fine spectral details (Sachs and Young, 1979) and rapid amplitude fluctuations (Palmer and Evans, 1979) that form critical features of speech. However, intelligibility may remain nearly perfect at intensities exceeding 90 dB (e.g., Studebaker, Sherbecoe, McDaniel, and Gwaltney, 1999). Consistent with this finding, Viemeister (1983) has provided behavioral evidence indicating that the dynamic range for rate-based intensity discrimination extends to 100 dB, and he has offered a theoretical account of encoding at high intensities relying upon firing-rate information provided by the small population of AN fibers known to have high thresholds and wide dynamic ranges; these fibers can account for rate-based intensity processing at the upper end of the dynamic range, if it is assumed that input from the larger population of readily saturated, low-threshold fibers is excluded from an analysis at high signal intensities (Siebert, 1968; Viemeister, 1983). Recent physiological models attribute this exclusionary process to mechanisms of lateral suppression (Eriksson and Robert, 1999; Winslow, Barta, and Sachs, 1987), which attenuate input from the frequency-tuned AN fibers when sufficient stimulation occurs in spectral regions adjacent to their best frequencies. Lateral suppressive interactions have been observed physiologically throughout the auditory pathway; they include both the mechanical suppression of vibration within the cochlea, reducing vibratory drive to the haircells (Rhode, 1971), and substantial neural inhibition of input to cells of the cochlear nucleus (Rhode and Greenberg, 1994), as well as higher processing stations (Schreiner and Langner, 1997; Zhang, Suga, and Yan, 1997). It is generally accepted that these lateral interactions can augment a fine-grained spectral analysis by increasing the contrast in firing rates of nerves tuned to the edges of peaks and valleys in a signal's spectrum. However, of interest in the present study is the proposal (Eriksson and Robert, 1999; Winslow *et al.*, 1987) that lateral suppression at high signal intensities primarily attenuates input from low-threshold AN fibers, and consequently produces a shift in the weighting of intensity analysis to favor the input from high-threshold fibers at high signal levels, as required by Viemeister's rate-based account of intensity discrimination at the upper end of the dynamic range (Viemeister, 1983). In the present study we provide behavioral support for the lateral suppression hypothesis, and as a consequence, also provide evidence for a physiological account of previously observed increases in the intelligibility of bandlimited speech produced by the addition of out-of-band noise.

Several studies examining the perception of filtered speech have found that the presence of complementarily filtered noise in the stop bands that filled the spectral gaps improved intelligibility, both for single speech bands (Bashford and Warren, 1987; Lenz, Bashford, and Warren, 2001; Shriberg, 1992) and for widely separated pairs of bands (Healy and Warren, 2003; Warren *et al.*, 1997). This noise-induced enhancement produced by filling spectral gaps with noise has been called ''spectral restoration,'' and was considered a possible counterpart to the phonemic restoration effect, which occurs when brief *temporal* gaps in speech (i.e., broadband interruptions) are filled with broadband noise, resulting in both illusory continuity and increased intelligibility of the interrupted speech. (For a detailed discussion of phonemic restoration, see Warren, 1999). Quantitatively similar

[a]Electronic-mail: bashford@uwm.edu

intelligibility increments, of as much as 15%, have been observed when either temporal or spectral gaps in speech have been filled with noise (e.g., Bashford, Riener, and Warren, 1992; Warren *et al.*, 1997). However, the enhanced intelligibility of temporally interrupted speech produced by stochastic noise has been found to require sentential context (Bashford *et al.*, 1992), suggesting a dependence on high-order processing utilizing linguistic knowledge. In contrast, the addition of complementary noise enhanced the intelligibility of bandlimited isolated words (Bashford and Warren, 1987; Warren *et al.*, 1997) and even improved the discriminability of bandlimited isolated vowels (Shriberg, 1992), suggesting that the effect may have a peripheral basis, such as lateral suppression. The present study tested the hypothesis that bandlimited speech can readily overload the auditory system through firing-rate saturation, due to reduced lateral suppression, and that the addition of noise in the speech-free regions reduces overload and enhances intelligibility by reinstating the lateral suppression effects that would normally be evoked by adjacent speech energy. Our previous studies showing enhancement due to out-of-band noise employed only high-normal presentation levels for speech-bands, with slow-rms peak levels ranging from 67 to 75 dBA. The first experiment of the present study found that increasing the slow-rms peak level of narrowband sentences from 55 dBA to 65 dBA produced a substantial intelligibility decrease, or ''rollover'' of the performance-intensity function; two subsequent experiments taken together indicate that this rollover effect can be substantially reduced when lateral suppression is evoked by out-of-band noise.

II. GENERAL METHOD

A. Subjects

A total of 128 listeners participated in the study: one group of 24 listeners in Experiment 1; groups of 20 and 28 listeners in Experiment 2; and two groups of 28 listeners in Experiment 3 (Differences in numbers of subjects across experiments were determined in part by differences in balancing constraints and the number of conditions employed.). Listeners ranged from 18 to 26 years in age. They were native speakers of English and had no known speech or hearing deficits, as determined by interview and questionnaire responses. All were undergraduate students at the University of Wisconsin—Milwaukee who were paid for their participation.

B. Stimuli

The speech stimuli were derived from the 100 (10 lists of 10) ''everyday'' sentences (e.g., ''I'd like some ice cream with my pie'') published by Silverman and Hirsh (1955). These lists contain 500 keywords that are used for scoring intelligibility. The sentences were read by a male speaker having an average voicing frequency of about 100 Hz and having no obvious regional accent. They were recorded digitally (44.1-kHz sampling, 16-bit quantization) along with an additional 25 practice sentences that were drawn from the high-predictability sublist of the revised Speech Perception in Noise (SPIN) test (Bilger, Nuetzel, Rabinowitz, and Rzeczkowski, 1984; Kalikow, Stevens, and Elliott, 1977). The practice sentences (e.g., ''She made the bed with clean sheets'') were produced by the same speaker as the test sentences. Depending upon the counterbalancing requirements posed by particular stimulus conditions, the number of test sentences presented to a group of listeners varied from 96 to 100.

Prior to filtering, the sentences were transduced by a Sennheiser HD 250 Linear II headphone and their slow-rms peak levels were matched to within 0.2 dBA using a flat-plate coupler in conjunction with a Brüel and Kjaer model 2230 digital sound-level meter set at A-scale weighting (as were all level measurements reported). The sentences were then passed through two successive stages of 4000-order bandpass FIR filtering, using the *fir* function in MATLAB, to produce a 2/3-octave narrow band of speech, centered at 1500 Hz (passband from 1191 Hz to 1890 Hz), with both lowpass and highpass filter slopes exceeding 3000 dB/octave.[1] The narrowband sentences were about 15 dBA lower in slow-rms peak level than their broadband parent sentences.

White noise, lowpass filtered at 20 kHz, was used to produce the three added-noise conditions employed in the study. They were (1) flanking bands of lowpass and highpass filtered noise, providing out-of-band stimulation; (2) narrowband noise matching the spectral limits of the speech band, providing within-band stimulation; and (3) broadband noise. The flanking bands of lowpass and highpass noise were each produced through two successive stages of 4000-order FIR filtering, as with the speech band. The low-frequency limit of the lower flanking band and the high-frequency limit of the higher flanking band matched those of the broadband noise. Separate pairs of flanking noise bands were prepared for each noise spectrum-level employed in the study (ranging from -60 to -10 dB relative to the speech spectrum-level). Noise cutoff-frequencies for each pair were adjusted so that the filter skirts of the flanking noise bands would intersect the skirts of the speech band at an average level 60-dB below that of the speech cutoff frequencies. A similar procedure was followed in preparing the narrowband noise that matched the spectral limits of the speech band: White noise was passed through two stages of 4000-order FIR bandpass filtering, separately for each noise spectrum-level employed, with cutoffs adjusted so that the transitions bands of the noise covered those of the speech band.

C. General procedure

The design of each experiment incorporated repeated measures. Before receiving a given stimulus condition listeners were presented with several practice sentences, which were first presented broadband and then presented bandpass filtered, along with noise when appropriate, in the same manner as the test sentences that followed. The different experimental conditions presented to a given group of listeners were assigned to separate sets of the ''everyday'' test sentences. This assignment varied pseudorandomly, so that, across listeners in a group, each condition was applied an equal number of times to each set of sentences.

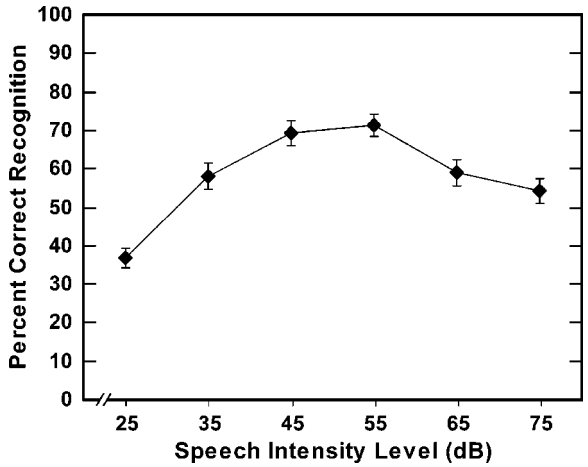


FIG. 1. The mean percent intelligibility scores and standard errors for 2/3-octave narrowband sentences (1500 Hz center frequency) presented at slow-rms peak levels ranging from 25 to 75 dBA.

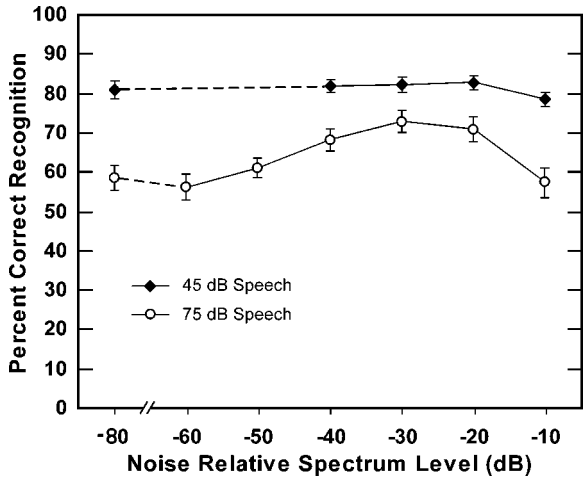


FIG. 2. Mean percent intelligibility scores and standard errors for 2/3-octave narrowband sentences (1500 Hz center frequency) presented at slow-rms peak levels of 45 dBA vs 75 dBA. The narrowband sentences were presented both without noise and along with flanking bands of white noise at spectrum levels ranging from −60 to −10 dB relative to the spectrum level of the speech band. Noise relative spectrum levels of −60 and −50 dB were omitted in the 45-dBA speech condition due to inaudibility.

Listeners were tested individually while seated with the experimenter in an audiometric chamber. The stimuli (both speech and noise when present) were delivered diotically through Sennheiser HD250 Linear II headphones. The number of keywords reported accurately was recorded by the experimenter, who was blind as to the conditions being presented. Testing required a single session lasting approximately 30 minutes.

III. EXPERIMENT 1: THE EFFECT OF INTENSITY ON NARROWBAND SENTENCE INTELLIGIBILITY

A. Method

In order to determine the intensity at which rollover occurs for the narrowband ''everyday'' sentences used in this study, 24 listeners each received a separate set of 16 sentences at each of six different average slow-rms speech peak-levels: 25, 35, 45, 55, 65 and 75 dBA (corresponding to broadband sentence-levels ranging from about 40 to 90 dB). The narrowband speech was presented without noise in this experiment.

B. Results and discussion

Figure 1 presents the mean intelligibility scores and standard errors obtained for the narrowband sentences presented at the six different levels in experiment 1. The data were subjected to an analysis of variance [$F(5,125)=26.1$, $p<0.0001$] followed by Tukey HSD post-hoc tests, which indicated that intelligibility increased with speech band intensity ($p<0.05$) up to presentation levels of 45 and 55 dBA, which did not differ, and then declined significantly ($p<0.05$) at the 65 and 75 dBA levels, with rollover losses of about 15% and 20%, respectively, relative to the 45 dBA condition. This finding of overload at levels of 65 dBA and higher calls into question the generality of previous studies that demonstrated the enhanced intelligibility of bandlimited speech produced by out-of-band noise. As discussed in the Introduction, those studies employed speech levels ranging from 67 to 75 dBA, and hence may have presented stimuli exclusively in the rollover range. Experiment 2 employed speech slow-rms peak levels of 45 dBA and 75 dBA to determine whether the noise-produced enhancement occurs below as well as within the rollover range of intensities.

IV. EXPERIMENT 2: THE EFFECT OF FLANKING NOISE BANDS UPON INTELLIGIBILITY OF NARROWBAND SPEECH PRESENTED AT 45 VS 75 dBA

A. Method

Two separate groups of listeners participated in this experiment. Each of the 28 listeners receiving the 75-dBA speech level was presented with seven separate sets of 14 test sentences: a no-noise set, and six sets having added flanking noise bands at spectrum levels of −60, −50, −40, −30, −20, and −10 dB relative to the speech spectrum level (measured at the center frequency of the speech band). Each of these seven conditions was assigned to each sentence set four times across listeners. An additional group of 20 listeners, who received the 45-dBA speech level, were presented with 20 sentences in each of five conditions: a no-noise condition along with four audible noise spectrum levels in common with those presented to the 75-dBA speech group: −40, −30, −20, and −10 dB (the −60 and −50 dB levels presented with the 75-dBA speech band would have been below threshold for the 45-dBA speech band). These five conditions also were assigned to each sentence set four times across listeners.

B. Results and discussion

Intelligibility scores for the 45- and 75-dBA slow-rms speech peak-level conditions are plotted in Fig. 2 as a function of flanking noise relative spectrum level. It can be seen that scores were substantially higher for the 45-dBA speech band for each of the common conditions, averaging about 80% regardless of noise level. An analysis of variance for the

45-dBA speech data verified that the addition of flanking noise had no reliable effect upon intelligibility [$F(4,76) = 0.84$, $p > 0.5$]. Very different results were obtained in the 75-dBA speech condition. The loss of intelligibility at the higher speech level (more than 20% in the absence of added noise) was substantially reversed when flanking noise was present at intermediate relative spectrum levels. An analysis of variance [$F(6,162) = 9.3$, $p < 0.0001$] followed by Tukey HSD post-hoc tests indicated that intelligibility of the 75-dBA speech band was increased ($p < 0.05$) by the addition of flanking noise at levels of -40, -30, and -20 dB. Intelligibility averaged across these three noise conditions, which did not differ significantly, was about 13% higher than that obtained in the no-noise condition. The lack of enhancement with flanking noise bands at the -10 dB spectrum level presumably was due to appreciable masking of the lower intensity speech components, especially by the lower-frequency noise band.

The enhancement effects produced by flanking noise are consistent with lateral suppression, which would be expected to improve intelligibility only when filtered speech is presented at intensities producing firing-rate saturation in AN fibers, and, hence, rollover. However, given the spread of excitation by the flanking noise bands, it is possible that some of the observed enhancement may have been due to additional stimulation within the frequency range of the speech band. Such within-band stimulation by the noise may have produced some adaptation of firing rate in AN fibers, and shifted their operating ranges to include higher signal levels (e.g., Gibson, Young, and Costalupes, 1985). Experiment 3 examined this possibility by directly measuring the effects of a wide range of noise levels upon the intelligibility of the narrowband sentences for two different noise conditions: (1) narrowband noise that matched the speech spectrum; and (2) broadband noise.

V. EXPERIMENT 3: THE EFFECTS OF WITHIN-BAND AND OF OUT-OF-BAND NOISE AT VARIOUS LEVELS UPON THE INTELLIGIBILITY OF HIGH-INTENSITY NARROWBAND SPEECH

A. Method

Two groups of 28 listeners were presented with the 75-dBA speech band. One group received the speech along with added broadband noise and the other group received added narrowband noise that matched the spectral limits of the speech band. This experiment employed the same range of noise levels employed with the 75-dBA speech in experiment 2, along with the no-noise condition. Listeners received 14 sentences in each of the seven conditions.

B. Results

Intelligibility scores for the narrowband and broadband noise conditions are plotted in Fig. 3 as a function of the noise spectrum level relative to that of the speech. It can be seen that added narrowband noise had no effect upon intelligibility except when considerable masking was produced at the highest noise spectrum levels of -20 and -10 dB. Intelligibility scores for the broadband noise condition substantially exceeded those for the narrowband noise when presented at spectrum levels of -40, -30, and -20 dB, with an average advantage of about 10%. The data for the narrowband and broadband noise conditions were subjected to an analysis of variance, which revealed significant effects of noise bandwidth [$F(1,54) = 7.5$, $p < 0.01$], and noise level [$F(6,324) = 93.2$, $p < 0.0001$], as well as a significant bandwidth$\times$level interaction [$F(6,324) = 2.49$, $p < 0.03$]. Subsequent simple-effects tests of noise bandwidth at each spectrum level confirmed ($p < 0.05$ or better) that intelligibility was significantly higher for the broadband noise condition at spectrum levels of -40, -30, -20, and -10 dB.

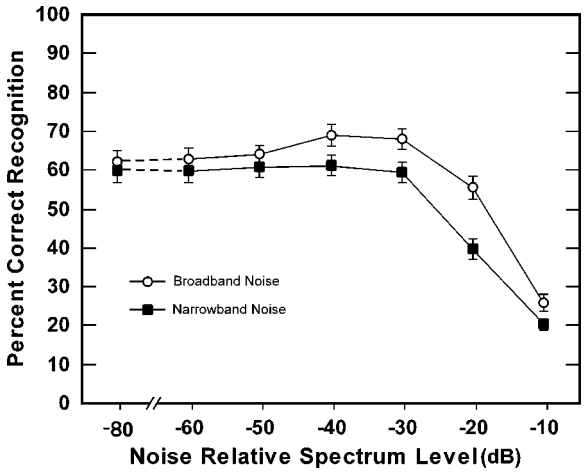


FIG. 3. Mean percent intelligibility scores and standard errors for 2/3-octave narrowband sentences (1500 Hz center frequency) presented at slow-rms peak level of 75 dBA. The narrowband sentences were presented both without noise and along with added noise at spectrum levels ranging from -60 to -10 dB relative to the spectrum level of the speech band. For one group of listeners, the added noise was bandpass filtered to match the spectral limits of the speech band. For the second group of listeners, the added noise was broadband.

VI. DISCUSSION

The results of experiment 3 appear to rule out within-band effects, such as firing-rate adaptation, as a source of the noise-induced recovery from intelligibility-rollover observed in experiment 2. Moreover, the recovery of intelligibility attributable to the out-of-band components of the broadband noise in experiment 3 also indicates that lateral suppression can help preserve the dynamic range of intensity coding when within-band noise masks low-level speech components, as well as increasing vulnerability to firing-rate saturation by the high-level components.

The substantial loss of intelligibility found for the narrowband sentences when they were presented at 75 dBA in these experiments is interesting, considering that the broadband parent sentences containing the narrowband at this level have slow-rms peaks averaging about 90 dB, which permits nearly perfect intelligibility for the broadband parent. The greater rollover effect obtained for the narrowband stimulus clearly is not due entirely to the lack of redundancy present in broadband speech, since the addition of out-of-band noise, providing no speech information, produced a recovery of more than half the intelligibility loss due to rollover.

Evidence from previous studies presenting high levels of filtered speech (Molis and Summers, 2003; Studebaker and Sherbecoe, 2002) suggests that the decrease in intelligibility may be more pronounced for bands consisting of frequencies higher than those employed in the present study, and may be less pronounced for lower frequencies. However, the present study shows that substantial rollover, and recovery due to out-of-band stimulation, does occur in that portion of the spectrum most important for intelligibility (Warren, Riener, Bashford, and Brubaker, 1995).[2]

ACKNOWLEDGMENT

This work was supported by NIH Grant No. DC00208.

[1]The two successive stages of 4000-order FIR filtering resulted in slopes of about 3.2 dB/Hz (3030 dB/octave highpass and 4480 dB/octave lowpass) and a stopband attenuation greater than 75 dB (the use of successive stages reduced spectral artifacts in the stopband). The high filter-order eliminated appreciable contributions from the transition bands, which could prevent accurate measurement of overload (''rollover'') within the passband (see footnote[2]), as well as complicating the interpretation of added-noise effects.

[2]In a study examining effects of filter slopes upon intelligibility, Warren *et al.* (2004) also presented listeners with filtered speech at low and high levels: 45 vs 75 dBA slow-rms peak. They used the same sentences employed in the present study, that were also subjected to narrowband filtering having a center frequency 1500 Hz, but had a considerably narrower width of 1/3-octave rather than 2/3-octave. There was no added noise, and it was found that when filter slopes were sufficiently steep to effectively eliminate information within the transition bands (i.e., greater than 2400 dB/octave), the intelligibility of the narrowband sentences presented at the 45 dBA level exceeded that obtained at the 75 dBA level by approximately the same proportion observed in the present study. However, due to a floor effect imposed by the low intelligibility of the 1/3-octave band, the difference between the 45 and 75 dBA levels was not statistically significant in that study.

Bashford, J. A., Jr., Riener, K. R., and Warren, R. M. (**1992**). ''Increasing the intelligibility of speech through multiple phonemic restorations,'' Percept. Psychophys. **51**, 211–217.

Bashford, J. A., Jr., and Warren, R. M. (**1987**). ''Effects of spectral alternation on the intelligibility of words and sentences,'' Percept. Psychophys. **42**, 431–438.

Bilger, R. C., Nuetzel, 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.

Eriksson, J. L., and Robert, A. (**1999**). ''The representation of pure tones and noise in a model of cochlear nucleus neurons,'' J. Acoust. Soc. Am. **106**, 1865–1879.

Gibson, D. J., Young, E. D., and Costalupes, J. A. (**1985**). ''Similarity of dynamic range adjustment in auditory nerve and cochlear nuclei,'' J. Neurophysiol. **53**, 940–958.

Healy, E. W., and Warren, R. M. (**2003**). ''The role of contrasting temporal amplitude patterns in the perception of speech,'' J. Acoust. Soc. Am. **113**, 1676–1688.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.

Lenz, P. W., Bashford, J. A., Jr., and Warren, R. M. (**2001**). ''Intelligibility of narrowband speech is enhanced by unmodulated flanking noise,'' J. Acoust. Soc. Am. **110**, 2658 (abstract).

Molis, M. R., and Summers, V. (**2003**). ''Effects of high presentation levels on recognition of low- and high-frequency speech,'' ARLO **4**, 124–128.

Palmer, A. R., and Evans, E. F. (**1979**). ''On the peripheral coding of the level of individual frequency components of complex sounds at high levels,'' Exp. Brain Res. **2**, 19–26.

Rhode, W. S. (**1971**). ''Observations of the basilar membrane in the squirrel monkey using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.

Rhode, W. S., and Greenberg, S. (**1994**). ''Lateral suppression and inhibition in the cochlear nucleus of the cat,'' J. Neurophysiol. **71**, 493–514.

Sachs, M. B., and Young, E. D. (**1979**). ''Encoding of steady-state vowels,'' J. Acoust. Soc. Am. **66**, 470–479.

Schreiner, C. E., and Langner, G. (**1997**). ''Laminar fine structure of frequency organization in auditory midbrain,'' Nature (London) **388**, 383–386.

Shriberg, E. E. (**1992**). ''Perceptual restoration of filtered vowels with added noise,'' Lang Speech **35**, 127–136.

Siebert, W. M. (**1968**). ''Stimulus transformations in the peripheral auditory system,'' in *Recognizing Patterns*, edited by P. A. Kolers and M. Eden (MIT Press, Cambridge, MA).

Silverman, S. R., and Hirsh, I. J. (**1955**). ''Problems related to the use of speech in clinical audiometry,'' Ann. Otol. Rhinol. Laryngol. **64**, 1234–1245.

Studebaker, G. A., and Sherbecoe, R. L. (**2002**). ''Intensity-importance functions for bandlimited monosyllabic words,'' J. Acoust. Soc. Am. **111**, 1422–1436.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). ''Monosyllabic word recognition at higher-than-normal speech and noise levels,'' J. Acoust. Soc. Am. **105**, 2431–2444.

Viemeister, N. F. (**1983**). ''Auditory intensity discrimination at high frequencies in the presence of noise,'' Science **221**, 1206–1207.

Warren, R. M. (**1999**). *Auditory Perception: A New Analysis and Synthesis* (Cambridge University Press, Cambridge, UK).

Warren, R. M., Bashford, J. A., Jr., and Lenz, P. W. (**2004**). ''Intelligibility of bandpass filtered speech: Steepness of slopes required to eliminate transition band contributions,'' J. Acoust. Soc. Am. **115**, 1292–1295.

Warren, R. M., Riener, K. R., Bashford, J. A., Jr., and Brubaker, B. S. (**1995**). ''Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,'' Percept. Psychophys. **57**, 175–182.

Warren, R. M., Riener-Hainsworth, K., Brubaker, B. S., Bashford, J. A., Jr., and Healy, E. W. (**1997**). ''Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps,'' Percept. Psychophys. **59**, 275–283.

Winslow, R. L., Barta, P., and Sachs, M. B. (**1987**). ''Rate coding in the auditory nerve,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 212–224.

Zhang, Y., Suga, N., and Yan, J. (**1997**). ''Corticofugal modulation of frequency processing in bat auditory system,'' Nature (London) **387**, 900–903.

Stopped-pipe wind instruments: Acoustics of the panpipes

N. H. Fletcher[a)]
Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia

(Received 13 July 2004; revised 19 September 2004; accepted 23 September 2004)

Stopped-pipe jet-excited musical instruments are known in many cultures, those best-known today being the panpipes or syrinx of Eastern Europe and of the Peruvian Andes. Although the playing style differs, in each case the instrument consists of a set of graduated bamboo pipes excited by blowing across the open tops. Details of the excitation aerodynamics warrant examination, particularly as the higher notes contain amplitudes of the even harmonics approaching those of the odd harmonics expected from a stopped pipe. Analysis shows that the jet offset is controlled by the fluid dynamics of the jet, and is such that appreciable even-harmonic excitation is generated. The theory is largely confirmed by measurements on a player.
[DOI: 10.1121/1.1815132]

PACS numbers: 43.75.Qr, 43.75.Ef, 43.28.Ra [TDR] Pages: 370–374

I. INTRODUCTION

Musical instruments consisting of sets of simple stopped pipes over which the player blows an air jet from the lips to produce the sound have an immensely long history. According to an excellent summary by Marcuse,[1] an example made from seven or eight bird bones has been found in a Neolithic cemetery in the Ukraine and dated to between 2300 and 2000 B.C. Similar instruments made from a wide variety of materials were used in classical Greek culture, from which the names ''panpipes'' and ''syrinx'' are derived,[1] and also in Africa, Asia, and the Americas. The instruments most familiar today are those with pipes made from bamboo and played by musicians from Eastern Europe and from the Peruvian Andes in South America. A typical example of a good-quality instrument from Eastern Europe is shown in Fig. 1. The playing styles in the two places are very different, the best-known recorded performances from Europe being those of the quasiclassical repertoire by Gheorghe Zamfir, while the Andean musicians adopt a percussive style of playing in popular and folk music. The acoustics of both of these playing styles will be examined in the present paper.

II. PHYSICAL PARAMETERS

As shown in Fig. 1, the panpipes studied are very carefully constructed from bamboo, the base of each pipe being terminated at a septum of the stem. The pipes are glued into a slightly curved inlaid wooden holder. Each pipe is lightly beveled at the top inner edge to present a smooth surface to the player's lips, and the larger pipes have been slightly beveled on the inside.

The instrument is able to sound a diatonic scale based on F♯4 (349 Hz). Intonation is subjectively good (it depends upon the player to some extent) except for the shortest pipe, which sounds C7 instead of the expected B♮6. It is not known whether or not this is intentional.

The pipe diameters decrease steadily with increasing pitch, but the slight irregularities of cross section make accurate measurements a little difficult. Figure 2 shows a plot of measured pipe diameter d as a function of pipe length L, both in millimeters. A power-law regression has been fitted with the result

$$d = 1.32L^{0.43}. \tag{1}$$

The purpose of this scaling is twofold. In the first place it shortens the length of the air jet from the player's lips to the outer edge of the pipe, thus facilitating the playing of high notes, as will be discussed in the following. Second, but relatedly, it tends to equalize the loudness over the compass of the instrument.

It is interesting to compare this scaling law with those commonly adopted for organ flue pipes.[2] Such empirical laws are often expressed as the musical interval over which the pipe diameter doubles, or by giving the diameter ratio for pipes differing by an octave, the latter being typically in the range 1.6–1.7. Expressed as in Eq. (1), this gives the approximate rule $d \propto L^{0.75}$. The pipes in the panpipes thus decrease in diameter more slowly than do the pipes in a typical organ rank. Because the musical compass of the panpipes is only 2.5 octaves compared with the 5 octaves of an organ rank, there is not a large effect on tonal balance across the rank, and indeed this would matter little in any case since the panpipes are played as a simple melody instrument.

An example of panpipes from the Peruvian Andes showed a scaling very similar to the European instruments, though the pipe diameters were much less regular, perhaps reflecting the much lower price of the particular instrument measured. A Thai vot, in contrast, had its pipes arranged in a circle around a solid core, and all pipes were of approximately the same diameter. In the vot, also, the pipe entries were all cut obliquely and it was much more difficult for an untutored player to produce the sound.

a)Electronic mail: neville.fletcher@anu.edu.au

FIG. 1. A set of panpipes from Eastern Europe. Note the diameter scaling.

III. ACOUSTIC THEORY

There is a large literature on the acoustics of jet-excited resonators,[3] but most of it refers to unstopped pipes such as organ diapason ranks, to the familiar transverse flute, or to the recorder. While much of this investigation is relevant to the case of stopped pipes, there are some very significant differences. The only paper on panpipe acoustics of which I am aware is a recent contribution on a waveguide model for panpipe sound by Czyzewski and Kostek.[4] In what follows, the simplest possible theory will be put forward, based upon the jet-drive mechanism proposed by Coltman,[5] Elder,[6] and Fletcher.[7] It is recognized that vortex generation may also play a significant role in the process, as initially proposed by Howe[8,9] and elaborated by Hirschberg and his collaborators,[10,11] but this theory is much more complex and is effectively concealed in the simple jet theory through the concept of a mixing region just inside the pipe mouth, in which momentum flux is conserved and energy is dissipated.

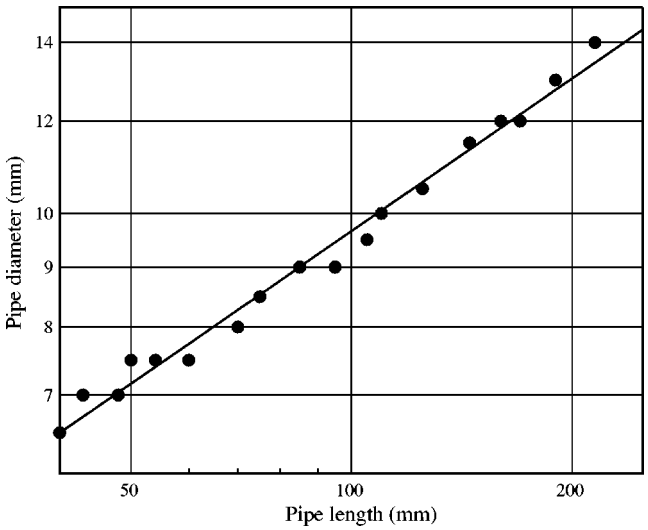


FIG. 2. Pipe scaling of the Eastern European panpipes. The regression line is $d=1.32L^{0.43}$ with both d and L in millimeters.

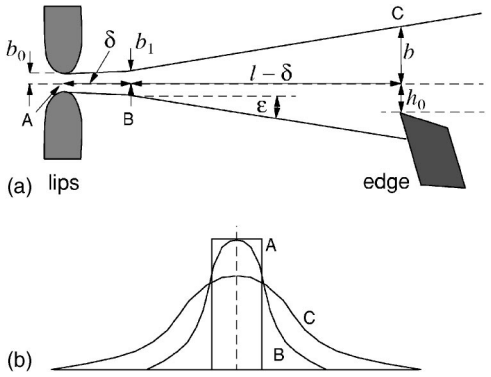


FIG. 3. (a) Development of the jet during propagation across the mouth of the pipe, showing the symbols used in the discussion. In the region between A and B the jet profile changes from ''top-hat'' to Bickley form, and this then broadens linearly by viscous entrainment up to point C. (b) Development of the jet velocity profile, assuming an initial ''top-hat'' shape. Labels A, B, and C correspond in the two parts of the figure.

There is, of course, also considerable energy dissipation by entrainment and vortex formation just outside the mouth of the pipe. While these vortices contribute to sound production in the Howe–Hirschberg model because of their periodic nature, this contribution is small, as in an edge-tone, and is neglected in the mixing-region model.

A very important difference between jet excitation of a closed and an open pipe relates to the steady component of the flow into the pipe. In the case of an open pipe this is of no consequence, because the downstream impedance of the pipe at zero frequency is essentially zero and all the flow can escape. In the case of a stopped pipe, however, the zero-frequency impedance is infinite, so that any steady flow is completely blocked. This has the consequence that the direction of jet flow into a stopped pipe is influenced by the steady balancing component of outflow, and indeed almost completely controlled by this. As a result, the direction of the mean centerline of the jet is essentially fixed by aerodynamic considerations rather than by the player's lip configuration, so that sound production is a relatively stable and simple matter compared, for example, with the difficulty of sounding a comparable end-blown open tube such as the Japanese shakuhachi. The first part of the theoretical development will be devoted to exploring this matter.

IV. AERODYNAMICS

A reasonable approximation to the jet behavior can be deduced by considering the problem in two dimensions. The flow behavior and the symbols used are defined in Fig. 3. When the jet emerges from the player's lips its velocity profile is something between a ''top-hat'' and a parabolic shape, as determined by the effective length of the lip channel. In either case, the central air velocity V_0 is determined simply by the blowing pressure in the mouth and, as in the flute, is typically about 20–40 m/s depending upon the note being played.[12] The flow profile changes shape within a short distance δ after leaving the lips because of viscous drag from the surrounding air, and rapidly assumes the Bickley profile[13]

$$V(z)=V_0\,\mathrm{sech}^2(z/b_1), \tag{2}$$

where z is the coordinate transverse to the jet and b_1 is the jet half-width parameter. It is a reasonable approximation to take the jet center-plane velocity V_0 to remain constant during this quite rapid transition, since viscous effects must diffuse in from the edges of the jet, and this assumption will simplify subsequent calculations.

If $2b_0$ is the width of the player's lip aperture and b_1 the jet half-width parameter after profile adjustment, as shown in Fig. 3, then conservation of momentum flux demands that the quantity

$$J=\int_{-\infty}^{\infty} V(z)^2 dz \tag{3}$$

should remain constant during this profile transition, and this requires that $b_1/b_0=3/2$ if the initial profile is top-hat, and $b_1/b_0=4/5$ if it is parabolic. Integrating across the flow profile in each case to determine the volume flow U in the jet shows that $U=1.5U_0$ if the initial flow profile is top-hat, and $U=1.2U_0$ if it is parabolic, where U_0 is the initial volume flow from the player's lips. Since half of the extra entrained flow comes from above and half from below the jet, it follows that the flow entrained from within the pipe by this jet profile adjustment is βU_0, where $0.1<\beta<0.25$.

As the jet travels across the top of the pipe, it broadens further because of viscous drag, as shown in Fig. 3, and entrains more air. The analysis of Bickley[13] gives an expression for this broadening, but measurements on organ-pipe jets at the sort of pressures used in blowing the panpipes show that while the Bickley profile (2) is maintained to a very good approximation for a distance of around 15 mm in a typical case. The jet actually slows and broadens more than predicted, the broadening being approximately linear with distance and representing a spreading half-angle ϵ of about 6°.[14]

There is further entrainment of air as the jet widens in its progress over the remaining distance $l-\delta$ across the top of the pipe, the total path length being taken as l. If the value of the half-width parameter b has changed by a factor α, then in order to preserve the constancy of the momentum flux integral J, given by Eq. (3), the central velocity of the jet must change by a factor $\alpha^{-1/2}$. Taking these two variables into account, the total volume flux in the jet will change by a factor $\alpha^{1/2}$. In the present case, the jet length is typically about 5–10 mm and the initial jet half-width about 0.5–1 mm, while the spreading makes the final jet half-width about 1.5–2 mm, making $\alpha\approx 2$. The jet volume flow after it has traversed the top of the tube is thus about 1.4 times its initial flow, and half of this increase, or 0.2 times the original flow, has been drawn from within the pipe. To this must be added fraction 0.1–0.25 of the initial flow discussed previously, making a total entrained flow from within the pipe of about $0.4U_0$. To be more specific about this analysis, let the jet half-width after profile relaxation be b_1 and the half-width at the lip of the pipe b, then $\alpha=b/b_1$ and the net entrained flow from within the pipe is

$$U'=\frac{U-U_0}{2}=\left(\frac{\alpha^{1/2}-1}{2}+\alpha^{1/2}\beta\right)U_0=\gamma U_0, \tag{4}$$

which is about $0.4U_0$. It is not helpful to attempt to refine these figures, since they are based on the approximation of two-dimensional flow, and in any case vary considerably from player to player and from note to note.

To balance this entrained flow with the total flow injected from the jet, the jet centerplane must be deflected out of the pipe by an amount h_0 so that

$$V\int_{-\infty}^{-h_0} \mathrm{sech}^2(z/b)dz=Vb[1-\tanh(h_0/b)]=U', \tag{5}$$

where $V=\alpha^{-1/2}V_0$ is the center-plane velocity of the jet at this position and V_0 is the jet velocity as it leaves the lips. From Eq. (4), this requirement reduces to

$$\tanh(h_0/b)=1-2\gamma\alpha^{-1/2}. \tag{6}$$

Inserting the typical values $\alpha\approx 2$ and $\gamma\approx 0.4$ as discussed before, the necessary offset is $h_0\approx 0.5b$. While this value clearly depends in detail upon the player and the note being played, it shows that the jet is deflected so that its midplane passes outside the lip of the pipe by a significant fraction of the jet half-width, and typically by about 0.5 mm for a note near the low end of the range.

V. ACOUSTIC EXCITATION

It is now important to assess the effect of this jet offset on the acoustic excitation of the pipe. This question has been investigated in detail by Fletcher and Douglas[15] in relation to ordinary organ pipe jets, and the same considerations apply for stopped pipes. The actual acoustic output clearly depends upon both the excitation provided by the jet and the resonant reinforcement of the pipe air column which, for stopped pipes, emphasizes the odd harmonics. The essentials of the computation are as follows.

Because of the behavior of sinuous waves propagating on the jet, and the related adjustment of blowing pressure and jet length by the player, as already referred to, the jet deflection responds almost entirely to the amplitude of acoustic flow in the fundamental of the sound. Because, however, the amplitude of the jet deflection is not small compared with the jet width, the jet flow into the pipe contains harmonics of the fundamental, the amplitudes of which are related to the jet center-plane offset and, of course, to the amplitude of the fundamental oscillation. Theory shows, and experiments confirm, that only odd harmonics are generated when the jet offset is zero, but that the amplitudes of the even harmonics in both the flow and the radiated sound become quite large when the jet offset is greater than about $0.5b$ and the deflection amplitude comparable with b. The calculated behavior of harmonics in the injected flow for the case when the deflection amplitude is $2b$ is shown in Fig. 4, and this calculation has been confirmed by experiment[15] for the case of an open pipe.

The radiated acoustic power, however, depends also upon the total acoustic flow excited at the entrance to the pipe by the transverse oscillations of the jet. For a simple jet of velocity V and cross-section S_j, the acoustic volume flow in the pipe is

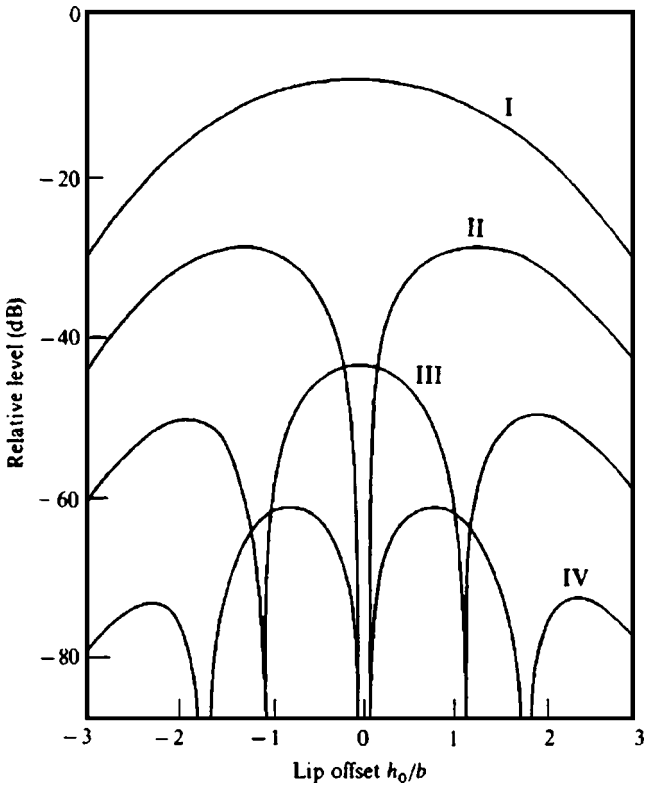


FIG. 4. Jet input flow at harmonics of the fundamental for a Bickley jet profile offset by a fraction h_0/b of the jet width, as calculated by Fletcher and Douglas (Ref. 15). For clarity, each successive curve has been shifted downwards by 10 dB relative to that for the previous harmonic. The calculations were confirmed by experiment on an open pipe.

$$U_p = \frac{(\rho V + j\rho\omega\Delta L)VS_j}{S_p Z_s}, \tag{7}$$

where S_p is the cross-sectional area of the pipe and Z_s is the input impedance of the pipe in series with its end correction.[3,7] This impedance has the approximate value

$$Z_s \approx jZ_0 \cot[(\omega/c - j\sigma)L], \tag{8}$$

where L is the tube length, adjusted for the end correction, $Z_0 = \rho c/\pi r^2$ is the characteristic impedance of the tube, assumed to have radius r, and $\sigma \approx 1.2\times 10^{-5}\omega^{1/2}r^{-1}$ is the wall-loss attenuation. The minimum value of Z_s, corresponding to excitation at an odd harmonic of the tube fundamental, is approximately $\tanh \sigma L \approx \sigma L$, while the maximum value, corresponding to even harmonic excitation, is $\coth \sigma L \approx 1/\sigma L$. The ratio between these two flow amplitude responses is approximately $1:(\sigma L)^2$, which is typically about 10^3 or 60 dB for exactly tuned resonances and antiresonances in ideal pipes of these dimensions. The ratio will be reduced, however, for nonideal pipe walls with higher losses, and for resonances, antiresonances, and playing frequencies that do not align exactly with the ideal model.

Another interesting phenomenon occurs in the South American percussive style of playing. Here the jet is launched in brief bursts and at relatively high pressure, so that it becomes turbulent before reaching the lip of the pipe. The turbulence contains components of all frequencies and is therefore able to excite all the modes of the pipe in a fashion that approximately displays the pipe impedance function, as

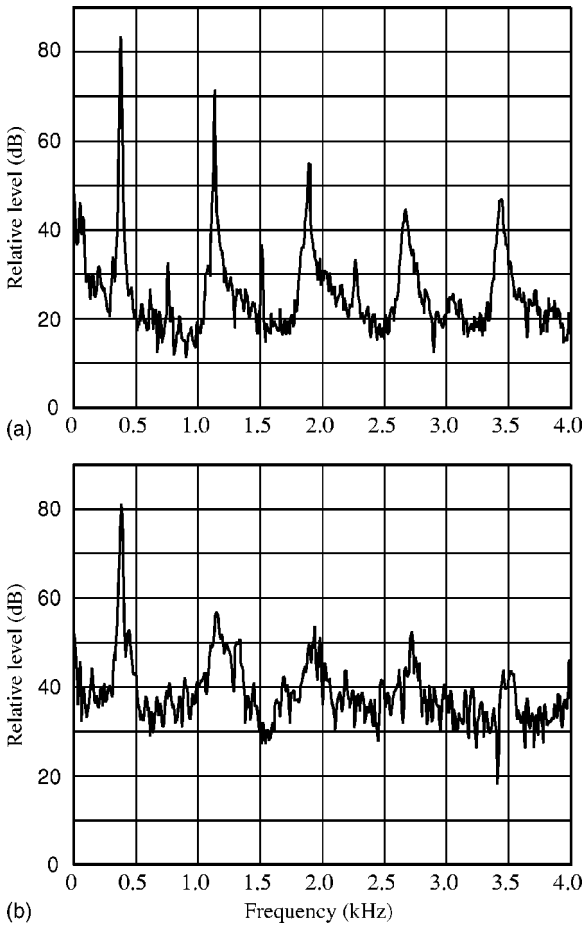


FIG. 5. Radiated sound spectra of the European panpipes for a low note played (a) with normal tone, and (b) percussively.

has been shown before in an experimental study of organ pipe noise background by Verge and Hirschberg.[16] In this playing style the fundamental is essentially the only mode that is excited to coherent oscillation.

VI. EXPERIMENTAL STUDY

A brief experimental study demonstrates the generation of even harmonics in normal playing and also the sound spectrum produced in the percussive playing style. While there is clearly a great deal of scope for adjustment of the jet parameters by experienced players, even those with little aptitude can produce acceptable sounds, and the player used for the experiments came into this second category. Figure 5(a) shows the sound spectrum of a low note played in a normal manner. The odd harmonics are very dominant, with the even harmonic levels being 20–50 dB below those of the neighboring odd harmonics, giving the sound a characteristic ''hollow'' quality. Figure 6(a) shows a similar measurement for a mid-range note. The even harmonics are much stronger, though still lying 10–20 dB below the envelope of the odd harmonics. This trend toward equality is continued for higher notes.

There are two reasons for this progression. The first is that the jet is made shorter for high than for low notes, partly because the pipe diameter is smaller and partly as a conscious technique to facilitate phase adjustment in the sound-

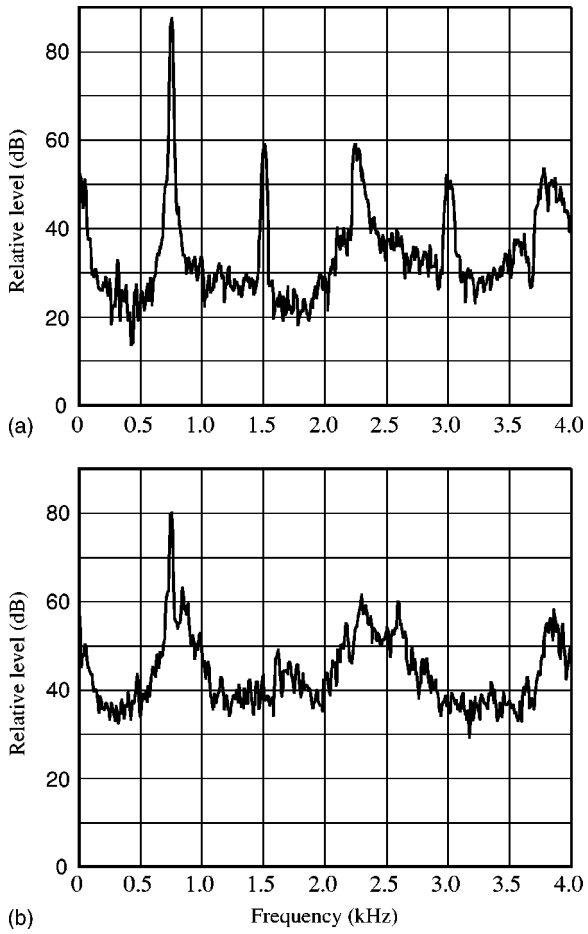


FIG. 6. Radiated sound spectra of the European panpipes for a mid-range note played (a) with normal tone, and (b) percussively.

generation feedback loop. Second, a higher blowing pressure is used for the high notes. Both these changes affect the width of the jet when it impinges upon the edge of the pipe in the same way, high notes having a narrower impact jet than low notes. If b_1 is the half-width of the jet when it relaxes to its steady profile in a distance δ after leaving the player's lips, then the half-width after traveling a total distance l to the opposite edge of the pipe is

$$b = b_1 + \epsilon(l - \delta), \tag{9}$$

where $\epsilon \approx 0.1$, corresponding to the experimental divergence angle of about 6° mentioned previously. Thus b/b_1 is smaller for the short jet lengths than for long. Because of its higher speed, the distance taken for the jet to achieve its steady profile and begin to diverge is greater for fast jets than for slow ones, leaving a smaller value for the divergence length $l - \delta$. From the analysis set out earlier, this means that the fractional offset h_0/b will be greater for high notes than it is for low notes, and consequently there will be greater excitation of even harmonics. On the other hand, it must be conceded that, since the player will generally reduce the height $2b_0$ of the lip opening for high notes,[3,12] this will operate in the opposite direction. Different players may therefore produce a different harmonic balance for high notes.

The spectra produced for the same two notes when played in the percussive style are shown in Figs. 5(b) and 6(b). Only the fundamental is generated in a tonal manner, and there is a noise spectrum with emphasis at odd harmonics of this fundamental, as is to be expected from the resonances (impedance minima, in this case) of the pipe. The broadening of these resonances can be ascribed to the dissipative effect of the jet turbulence. Examination of the ''normal playing'' spectra of Figs. 5(a) and 6(a) shows that this same noise background is present here, though at a much lower level. One might reasonably surmise that skilled players can reduce this noise still further to produce an excellent pure tonal sound.

VII. CONCLUSIONS

It is hoped that this short paper will lead to a rise in interest in the study of a beautiful and subtle wind instrument. It is not claimed that all problems have been solved by the above-presented analysis, but at least it is a start.

[1] S. Marcuse, *A Survey of Musical Instruments* (Harper and Row, New York, 1975), pp. 589–596.
[2] N. H. Fletcher, ''Scaling rules for organ flue pipe ranks,'' Acustica **37**, 131–138 (1977).
[3] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), pp. 503–580.
[4] A. Czyzewski and B. Kostek, ''Waveguide modelling of the panpipes,'' Proceedings of the Stockholm Music Acoustics Conference SMAC-03, 2003, pp. 259–262.
[5] J. W. Coltman, ''Sounding mechanism of the flute and organ pipe,'' J. Acoust. Soc. Am. **44**, 983–992 (1968).
[6] S. A. Elder, ''On the mechanism of sound production in organ pipes,'' J. Acoust. Soc. Am. **54**, 1554–1564 (1973).
[7] N. H. Fletcher, ''Jet-drive mechanism in organ pipes,'' J. Acoust. Soc. Am. **60**, 481–483 (1976).
[8] M. S. Howe, ''Contributions to the theory of aerodynamic sound, with applications to excess jet noise and the theory of the flute,'' J. Fluid Mech. **71**, 625–673 (1975).
[9] M. S. Howe, ''The role of displacement thickness fluctuations in hydroacoustics, and the jet-drive mechanism of the flue organ pipe,'' Proc. R. Soc. London, Ser. A **374**, 543–568 (1981).
[10] M.-P. Verge, R. Caussé, B. Fabre, A. Hirschberg, A. P. J. Wijnands, and A. van Steenbergen, ''Jet oscillations and jet drive in recorder-like instruments,'' Acta Acust. **2**, 403–419 (1994).
[11] B. Fabre, A. Hirschberg, and A. P. J. Wijnands, ''Vortex shedding in steady oscillation of a flue organ pipe,'' Acustica **82**, 863–877 (1996).
[12] N. H. Fletcher, ''Acoustical correlates of flute performance technique,'' J. Acoust. Soc. Am. **57**, 233–237 (1975).
[13] W. G. Bickley, ''The plane jet,'' Phil. Mag. **28**, 727–731 (1937).
[14] N. H. Fletcher and S. Thwaites, ''Wave propagation on an acoustically perturbed jet,'' Acustica **42**, 32–34 (1979).
[15] N. H. Fletcher and L. M. Douglas, ''Harmonic generation in organ pipes, recorders and flutes,'' J. Acoust. Soc. Am. **68**, 767–771 (1980).
[16] M.-P. Verge and A. Hirschberg, ''Turbulence noise in flue instruments,'' Proceedings of the International Symposium on Musical Acoustics, Dourdan, 1995, IRCAM, Paris, pp. 94–99.

Sound radiation from Caribbean steelpans

Brian Copeland
Department of Electrical and Computer Engineering, University of the West Indies, Trinidad and Tobago, West Indies

Andrew Morrison and Thomas D. Rossing
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

(Received 9 August 2002; revised 6 March 2004; accepted 13 April 2004)

Besides radiating sound from the note area being struck, a steelpan radiates from neighboring note areas that vibrate sympathetically, from the areas between notes, and from the skirt [Rossing *et al.*, Phys. Today **49**(3), 24–29 (1996)]. Measurements were taken in an anechoic chamber using a four-microphone intensity probe to visualize the acoustic radiation from selected notes on a double second and a low tenor steelpan. Swept sinusoidal excitation was effected using an electromagnet. Sound intensity maps were drawn for the first three harmonics. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1803574]

PACS numbers: 43.75.Kk, 43.40.Rj, 43.58.Fm [NHF] Pages: 375–383

I. INTRODUCTION

The steelpan is manufactured by systematic deformation of the top of steel drums originally designed for transportation of materials. The majority of instruments are constructed from 55-gallon drums. Each instrument requires from one (tenor pan) to as many as 12 (12-bass) drums. The entire steelpan family covers a musical range of E_1 (41.2 Hz) to G_6 (1567.98 Hz) (Ref. 1) with partials extending to above 3 kHz.

Despite its humble roots, a well-tuned pan is capable of rendering music of extremely high quality. As such, the instrument is now recognized as the most significant acoustic instrument invented in the 20th century. Moreover, because of a unique vibration mechanism that uses a single steel membrane to produce a range of musical notes, there has been a great deal of research interest in the various characteristics of the instrument. Kronman[2] and Rossing,[3] for example, have provided comprehensive descriptions on the history, construction, and performance of the instrument. Rossing,[3] and Rossing, Hampton, and Hansen[4,5] have reported on modal studies on a range of steelpans. Murr *et al.*[6] have reported on the metallurgical considerations in instrument manufacture following on the early work of Imbert.[7]

This paper presents and discusses the results of measurements performed on the sound field of a low tenor and a double tenor steelpan. As far as the authors are aware, the work of Muddeen and Copeland,[8] who studied sound pressure radiation patterns for a low tenor over a range of frequencies, represents the only other work on this issue. The present study goes further by looking at the distribution of active intensity (AI), reactive intensity (RI), and sound pressure level (SPL) in the near field of the instrument. As with other instruments,[9] studies of this type provide greater understanding of the way in which the steelpan reacts with its acoustical environment. These studies can also suggest improvements on various performance aspects, including room acoustics and microphone types and positioning. In the case of the steelpan, such studies may also provide clues as to how the physical form of the instrument may be improved. This last point is significant when one considers that the instrument is relatively new, having first appeared around 1935, and has evolved largely by trial and error.

The sound field is primarily determined by two attributes of the particular instrument:

(1) The modal characteristics: mode frequencies and relative phases as well as modal shapes and locations.
(2) The geometry of the instrument: The most important parameters here are the pan diameter, which is typically of the order of 55 cm, skirt length, and playing surface profile. These impact on the diffractive and reflective aspects of sound propagation.

The steelpan is still primarily manufactured by hand. In the ''fine tuning'' stage, the most critical phase of the manufacturing process, the shape of individual notes is repeatedly adjusted until the sound of the whole instrument is to the tuner's satisfaction. Tuning is an iterative process since an adjustment made to one note often throws another out of tune. It is important to note that the tuner has a significant degree of control over the timbre of individual notes. In the now common ''harmonic tuning'' process, for example, sections of the note are shaped to produce and control note partials at the second or even the third harmonic.[5] Some tuners also use coupling between notes to control the level of the partial. The ability to control the frequencies and levels of note partials adds to the uniqueness of the pan as a musical instrument; as far as the authors know, electronic synthesizers are the only other type of instrument that allows this degree of control of timbre. On the downside, this ability can result in a degree of dissonance as current tuning technology cannot guarantee the maintenance of the desired harmonic relationship between partials if or when the instrument loses tune.

Given all of the above, and taking into consideration the variability in raw materials, construction, and tuning styles, it might be expected that modal characteristics would vary somewhat over instruments in a given range. However, instrument geometry is fairly uniform for a given range of

steelpan. As such, the results presented here can therefore be taken as typical.

The paper proceeds as follows: A brief review of sound intensity (SI) is provided in Sec. II. Section III describes the relevant aspects of the instruments studied while the experimental setup is described in Sec. IV. Sample results and discussions on these results are presented in Secs. V and VI.

II. REVIEW OF SOUND INTENSITY

The literature carries many excellent expositions on various aspects of sound intensity and its value in determining propagation in a sound field.[10–13] In this section, however, we will review those aspects that are significant to the interpretation of the results presented. Note that bold symbols are used to represent vector quantities.

Assuming sinusoidal excitation at frequency ω rad/s, solution of the wave equation yields an instantaneous pressure

$$p(\mathbf{r},t)=P(\mathbf{r})\cos[\omega t-\phi(\mathbf{r})], \tag{1}$$

where $P(\mathbf{r})$ and $\phi(\mathbf{r})$ are real functions of the spatial position vector $\mathbf{r}$. The particle velocity vector is given by[10]

$$\begin{aligned}\mathbf{u}(\mathbf{r},t)&=-\frac{1}{\rho_0}\int\nabla p(\mathbf{r},t)\,dt\\&=\frac{1}{\omega\rho_0}P(\mathbf{r})\nabla\phi(\mathbf{r})\cos[\alpha(\mathbf{r},t)]\\&\quad-\frac{1}{\omega\rho_0}\nabla P(\mathbf{r})\sin[\alpha(\mathbf{r},t)],\end{aligned} \tag{2}$$

where ρ_0 is the air density and $\alpha(\mathbf{r},t)=\omega t-\phi(\mathbf{r})$. The *instantaneous total intensity* (in W/m^2) is defined by

$$\begin{aligned}\mathbf{I}(\mathbf{r},t)&=p(\mathbf{r},t)\mathbf{u}(\mathbf{r},t)\\&=\frac{1}{\omega\rho_0}P^2(\mathbf{r})\nabla\phi(\mathbf{r})\cos^2\alpha(\mathbf{r},t)\\&\quad-\frac{1}{\omega\rho_0}P(\mathbf{r})\nabla P(\mathbf{r})\sin\alpha(\mathbf{r},t)\cos\alpha(\mathbf{r},t)\\&=\mathbf{A}(\mathbf{r},t)+\mathbf{R}(\mathbf{r},t),\end{aligned} \tag{3}$$

where

$$\mathbf{A}(\mathbf{r},t)=\overline{\mathbf{I}}(\mathbf{r})[1+\cos 2\alpha(\mathbf{r},t)], \tag{4a}$$

$$\mathbf{R}(\mathbf{r},t)=\mathbf{Q}(\mathbf{r})\sin 2\alpha(\mathbf{r},t). \tag{4b}$$

$\mathbf{A}(\mathbf{r},t)$ is known as the *instantaneous active intensity* (AI) while $\mathbf{R}(\mathbf{r},t)$ is known as the *instantaneous reactive intensity* (RI). $\mathbf{I}(\mathbf{r})$ is the *time averaged intensity* or *mean intensity*

$$\overline{\mathbf{I}}(\mathbf{r})=\frac{1}{T}\int_0^T\mathbf{I}(\mathbf{r},t)\,dt=\frac{1}{2\omega\rho_0}P^2(\mathbf{r})\nabla\phi(\mathbf{r}) \tag{5}$$

$T=2\pi/\omega$ and $\mathbf{Q}(\mathbf{r})$ is given by

$$\mathbf{Q}(\mathbf{r})=-\frac{1}{2\omega\rho_0}P(\mathbf{r})\nabla P(\mathbf{r})=-\frac{1}{4\omega\rho_0}\nabla P^2(\mathbf{r}). \tag{6}$$

Following are key characteristics of sound intensity

(1) $\mathbf{A}(\mathbf{r},t)$ arises from the product of the in-phase components of p and $\mathbf{u}$ and therefore exclusively represents the energy dissipation component in the acoustic environment; it is in fact easy to show that its time average is exactly $\mathbf{I}(\mathbf{r})$, the dc component in Eq. (4a). Note that both the dc and sinusoidal components are of the same amplitude. This fact is used to identify the AI contribution of each frequency component in sound waves with more complex spectra. Wave propagation to the far field is effected by the AI component.

(2) Equation (4b) shows that $\mathbf{R}(\mathbf{r},t)$ is purely sinusoidal and therefore has a time average of zero. It arises from the product of quadrature components of p and $\mathbf{u}$. $\mathbf{R}(\mathbf{r},t)$ represents the movement of stored energy between different parts of the acoustic environment. In fact, Fahy[10] shows that in the special case of a monochromatic two-dimensional sound field, $\nabla\mathbf{Q}(\mathbf{r})$ is directly proportional to the difference between the mean kinetic and potential energy densities. In most cases the RI component is significant close to the source but drops off rapidly as one moves away from the source. This makes intuitive sense when one considers the simple compression model at the source/air interface for a stiff vibrating plate: Maximum or minimum source deflection of the source coincides with maximum or minimum air pressure and zero particle velocity.

(3) Equations (3) and (4) show that $\mathbf{I}(\mathbf{r},t)$ has a Fourier series decomposition with just a single nonzero frequency at 2ω and coefficient

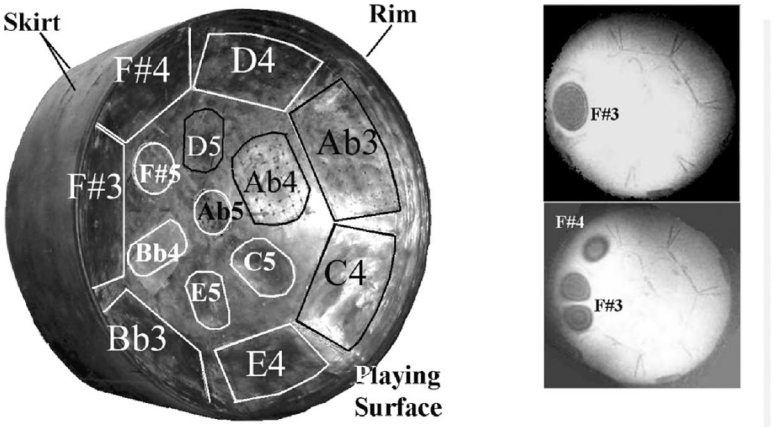
$$\mathbf{C}(\mathbf{r})=\overline{\mathbf{I}}(\mathbf{r})+j\mathbf{Q}(\mathbf{r}). \tag{7}$$

$\mathbf{C}(\mathbf{r})$ is defined in the literature[11,13] as the *complex intensity*.

(4) A vector plot is often used to show the variation of time averaged AI magnitude and direction over a measurement grid. From Eq. (5) we see that the instantaneous as well as the time averaged AI vectors always point in the direction of positive phase gradient $\nabla\phi(\mathbf{r})$, with instantaneous AI magnitudes varying from 0 to $2\|\mathbf{I}(\mathbf{r})\|$, where $\|\mathbf{x}\|$ denotes the magnitude of the vector $\mathbf{x}$. AI vectors are therefore always perpendicular to the wave front, the latter being defined as lines of uniform phase.

(5) From Eq. (6) we see that instantaneous RI vectors at a point $\mathbf{r}$ are always aligned with the direction of square pressure gradient, $\nabla P^2(\mathbf{r})$, but change direction at a rate of 2ω rad/s. A vector map is often used to show the variation of RI peak magnitude and direction over a measurement grid. Because of the negative sign in Eq. (6), these vectors are plotted in the direction of decreasing square pressure. Thus, the convergence of RI vectors indicates a pressure minimum at the relevant point; similarly the divergence of RI vectors indicates a pressure maximum at the relevant point.

(6) The AI and RI vectors combine to give the total intensity vector. It is easily shown from Eq. (4) that, under single frequency excitation, instantaneous total intensity vectors trace out ellipses over time.



(a) The drum and measurement reference frame

(b) Interferograms for excitation at 178 Hz (top) and 366 Hz

FIG. 1. Left side of Alexis double second pan. (a) The drum and the measurement reference frame, (b) holographic interferograms for excitation at 178 Hz (top) and 366 Hz (bottom).

(7) AI and RI are quoted in dB referred to $10^{-12}\,\mathrm{W/m^2}$. These reference standards are adhered to in all results described below.

Despite the fact that most commercial SI measurement systems provide only time-averaged or peak amplitude measurements, it is possible to reconstruct the instantaneous vectors from these measurements. Equations (1) and (2) clearly show that all one requires is the spatial phase function $\phi(\mathbf{r})$. This can be obtained indirectly from the mean intensity and pressure. From Eq. (5)

$$\nabla\phi(\mathbf{r}) = \frac{2\omega\rho_0}{P^2(\mathbf{r})}\overline{I}(\mathbf{r}), \tag{8}$$

which gives the phase gradient. This can now be integrated over the surface with arbitrary initial condition to give the relative phase distribution. Equations (1)–(3) can then be used to reconstruct the instantaneous variables as well as the wavefront (contours of uniform phase).

III. INSTRUMENTS STUDIED

A. Clifford Alexis double second pan

Figure 1(a) shows a photograph of the left side drum of the Alexis 15-note double second pan used in the study. The study focused on the $F_3^{\#}$ note.

The double second pan covers the alto range and, as its name suggests, is comprised of two drums. The sample studied covers the 28-note range from $F_3^{\#}$ to A_5. Each drum of the pair has 14 notes; chromatically adjacent notes are placed on different drums to minimize dissonance while providing good ergonomics. The drum has a diameter of 56 cm, a skirt length of 23 cm, and a bowl depth of 16 cm. The note areas of Fig. 1(a) have been highlighted for easy identification; axis alignments have been included for future reference.

Figure 1(b) shows holographic interferograms reported in Rossing, Hampton, and Hansen[5] for the $F_3^{\#}$ note area at 178 and 366 Hz, associated with the first and second modes, respectively. As in Rossing, Hampton, and Hansen[5] we use the notation (n,m) to indicate a pattern with n radial nodes and m circumferential nodes. At 178 Hz, the $F_3^{\#}$ note displays the expected (0,0) pattern expected for a surface excited at its fundamental frequency. At 366 Hz, the $F_3^{\#}$ note experiences the greatest displacement and displays the anticipated (0,1) nodal pattern for a surface excited at its second modal frequency while the $F_4^{\#}$ note area displays the expected (0,0) nodal pattern for a surface excited at its fundamental frequency. Sound intensity distributions at 178 and 366 Hz would be influenced by the relative deflection magnitudes and phases of the respective vibrating areas shown.

It should be noted that the nonlinear nature of the instrument suggests the possibility of the existence of other frequency components. Rossing, Hampton, and Hansen,[4] for example, have observed and attempted to characterize the octave component produced by single frequency sinusoidal excitation of varying intensity. For the sake of discussion, however, it is assumed that the patterns shown all occur at the applied frequency.

B. Bertrand Kellman low tenor pan

The Kellman tenor pan studied employs the now standard note layout based on the circle of fourths and fifths (Fig. 2). As can be seen from the figure, this pan has an outer ring of all 12 notes in the fourth octave as well as an inner ring of all notes of the fifth octave. The innermost ring has only the first five notes of the sixth octave because of space restrictions.

The pan is so named because it evolved from drums with a very limited (tenor) range; it is more correctly called a soprano pan in Europe and North America. The fourths and fifths note layout, introduced in the 1960s, results in a significant reduction in dissonance due to internote coupling by ensuring that each note is surrounded by its octave, fifth and fourth.[2] In addition, the use of the circle of fifths makes this instrument ideal for teaching music. Fourths and fifths tenors originally used a note range starting from D_4 so that the lowest C note was C_5. With the expansion of note range down to C_4 the original tenors became known as high tenors and the new instruments low tenors or low-C tenors.

The sample studied has a diameter of 58.5 cm, a skirt length of 18 cm, and a bowl depth of 22.5 cm. Sound intensity studies were conducted for D_4 note excitation; as for the Alexis pan, modal patterns for the first, second, and third

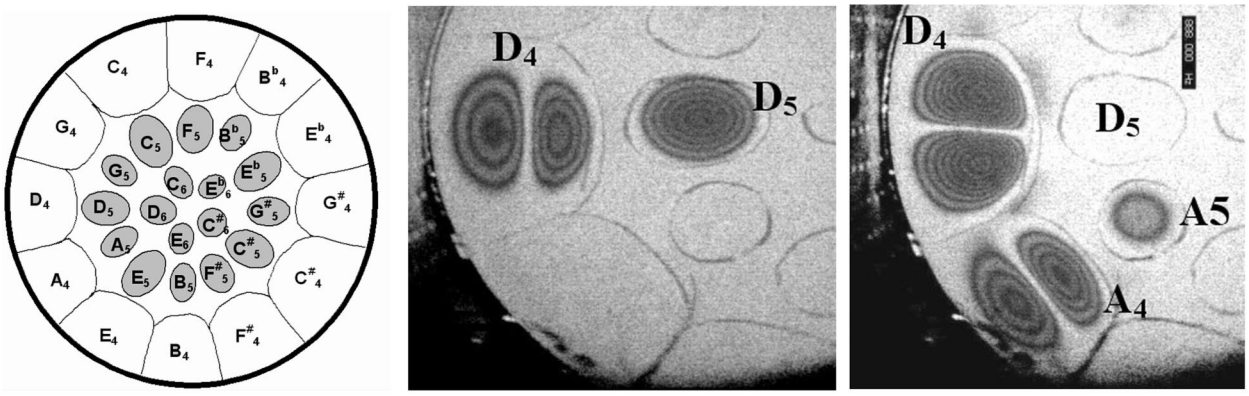


FIG. 2. (a) Kellman low tenor pan with fourths and fifths (circle of fifths) note layout, (b) modal pattern for D4 note excited at 587 Hz, and (c) modal pattern for D4 note excited at 880 Hz.

harmonics were obtained using holographic interferometry. These are shown in Figs. 2(b) and (c). Note that for performance purposes the instrument is positioned with the C note at the bottom of the pan; the figure has been rotated to better display the orientation used in the study.

Excitation at the frequency of the second mode (587 Hz) produces a (0,1) pattern on D_4 and a (0,0) pattern on D_5. Excitation at the frequency of the third mode (880 Hz) results in a (1,0) pattern on D_4, a (0,1) pattern on A_4 note, the fifth, and a (0,0) pattern on A_5. Given the layout and current tuning techniques, these results are expected. For the third partial, the displacement levels, as estimated from the holographic fringe count on the modal patterns, are in the ratio of 8:3:1. Thus the A_4 and note A_5 vibration amplitudes are 8.5 and 18 dB, respectively, below that of the D_4 note. As we shall see, these modal patterns have a significant impact on the sound intensity distribution patterns.

IV. EXPERIMENTAL SETUP AND PROCEDURE

The Alexis double tenor pan was suspended with its face vertically oriented in a metal frame of dimension 160 cm $\times$160 cm$\times$160 cm constructed of 2 cm aluminum tubing. One set of measurements was taken on a grid parallel to the *x-y* plane (Fig. 1) incorporating a radial line through the center of the $F_3^{\#}$ note and the cylindrical axis of the pan. A second set of measurements was made on a similar grid parallel to the *y-z* plane, passing through the center of the $F_3^{\#}$ note. Grid spacing was set at 7 cm, determined after consideration of the wavelengths of the note partials. Each measurement plane was 105 cm$\times$105 cm giving a total of 256 measurement points. The excitation mechanism involved a coil positioned over a 5 mm diameter NdFeB magnet placed on the outer section of the note. A 0–2 kHz chirp (swept sine) excitation signal was used with a repetition time of 1 s. Measurements were taken over three separate regions on the grid and the resulting data subsequently combined. Measurements within 7 cm or so of the pan were limited by the physical geometry of the microphone.

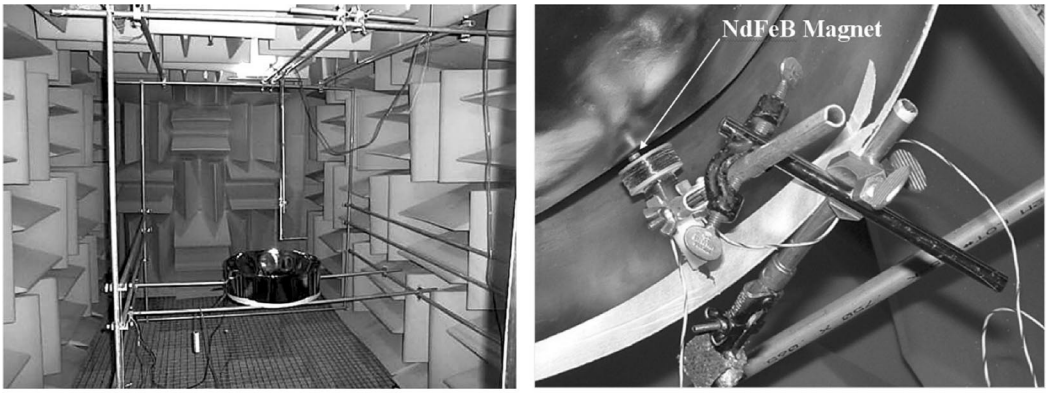
Sound intensity measurements for the Kellman tenor were taken on a plane 5 cm above and parallel to the face of the pan, centered on the cylindrical axis of the pan (Fig. 3). The same grid dimension, spacing, and excitation mechanism was used as for the Alexis tenor; the measurement grid was centered on the cylindrical axis of the pan.

The AI, RI, and SPL spectra were acquired using Ono Sokki equipment comprising the CF-0357 SI Analysis software, the CF-350 FFT Analyzer, and the M6410 4-channel microphone. Only the low frequency pair of microphones were utilized on the M6410; this pair has a separation of 5 cm and a stated frequency response of 50 to 2000 Hz. Intensity measurements were taken in three orthogonal directions and AI, RI, and SPL data extracted for each frequency of interest. This data was further analyzed and plotted using software written in MATLAB.

Although the spectra revealed a great many frequency components, particularly at high frequency, this paper will focus only on the results for the first two partials for the Alexis pan and the first three partials for the Kellman pan. A significant third partial component was not observed for the $F_3^{\#}$ on the Alexis pan. The limitation in frequency range for analysis is not entirely restrictive since, as observed from the acquired spectra, the electromagnetic excitation system used imparted more high frequency energy to the note than that provided by pan sticks used for this range of instrument. In other words, the forcing function used was quite differently from that used in normal pan performance. It is well known[2] that the rubber ends of pan sticks aid in the suppression of the higher (non-musical) frequency modes.

V. RESULTS AND DISCUSSION FOR THE ALEXIS DOUBLE SECOND PAN

Figures 4 and 5 show the AI, RI, and SPL plots for the first and second harmonics, respectively, for the double second pan. A sketch of the pan is included as wireframe drawings or semitransparent images so as to establish the correct spatial orientation. A star is also included to indicate the location of the excited note and the bar in Fig. 4(d) provides



(a) Kellman tenor on rig in the anechoic chamber at NIU

(b) Electromagnetic exciter positioned under the pan.

FIG. 3. Measurement setup. (a) Kellman tenor on measurement rig in the anechoic chamber, (b) electromagnetic exciter positioned under the Kellman pan. The NdFeB magnet is just visible above the exciter coil.

the color reference for all plots shown. References to the left, right, top, and bottom of the pan will be taken in regard to the perspective shown in Fig. 1. The *y-z* grid is presented as viewed from the left of the pan.

The intensity plots shown here utilize solid cones to represent intensity vector directions. Each cone in the graph represents a single point of measurement. The use of color to represent the magnitude variation of the displayed parameter further enhances the presentation of results. Finally, the measurement plane has been rendered semitransparent to reveal vector orientation beneath the surface.

A. First partial (fundamental)

The AI vectors on the *x-y* grid of Fig. 4(a) clearly show the transport of acoustical energy away from the source. All of the AI vectors are slanted upwards from the *x-y* grid; their general orientation suggests that the energy emanates mainly from the $F_3^{\#}$ note area. AI plots for the *y-z* grid show no discernible pattern; this may be as a result of the fact that the plane intersects the pan along a chord that is too short relative to this wavelength. AI is strongest in the area of the $F_3^{\#}$ and $F_4^{\#}$ notes to the front of the pan. This, of course, is quite expected, particularly in light of the modal patterns shown in Fig. 1(b). AI is weakest, by 30 to 40 dB of maximum, in the left and right skirt areas for the fundamental.

One would expect the RI field in the vicinity of vibrating surfaces to display circulatory patterns.[10] RI would be greatest in the vicinity of the surface antinodes (maximum average pressure) and weakest in the area of the surface nodes (minimum average pressure). In general, a snapshot of the instantaneous RI vectors would reveal streams that emanate from an area of maximum (positive) compression, running across a nodal area and terminating in a region of minimum (negative) compression. The implication is that a component of reactive energy oscillates back and forth between these regions. The *x-y* RI plots (Fig. 4) indicate circulatory patterns around the left and right side; this is also evident to the rear of the pan on the *y-z* grid, which seems to indicate RI flow around the top and bottom of the pan.

The *x-y* vector pattern of Fig. 4 supports the conclusion that the $F_3^{\#}$ note is the main radiator, although it is clear that most of the pan is involved in the process. The RI vectors converge from both the front and rear on to the middle of the skirt suggesting the existence of pressure lows on the left and right sides. Indeed, the corresponding pressure plot confirms that the pressure levels in the area of the skirt are of the order 30–35 dB less than the peak pressure. The modal plot of Fig. 1(b) shows that the $F_3^{\#}$ note resembles a monopole source; the zones of high and low compression therefore alternate between the front and rear of the pan. One observes that RI is strongest in the excited note area. There is an analogy here from what one would expect of RI patterns in the vicinity of a low frequency speaker operated outside of an enclosure. Since the AI is also low in the skirt areas adjacent and opposite to the excited note there is a clear indication that these areas do not contribute significantly to the sound field at this frequency.

The AI and SPL plots suggest the existence of one major lobe to the front and rear. The front lobe is almost symmetrical about the cylindrical axis while the rear lobe seems to be directed to the top right of the instrument. This conclusion is made based on the general notion of AI being responsible for propagation in the far field and the observed orientation of the strongest AI vectors in close proximity to the pan.

B. Second partial (octave)

Figure 5(a) shows *x-y* grid AI vectors oriented away from the note pair in an upward direction in the left and top left areas of the figure and a downward direction elsewhere. Again the vectors seem to emanate from the excited area at this frequency. Results on the *y-z* grid confirm that levels are much more intense for the octave than the fundamental. As noted previously, this is a direct consequence of the means of excitation. In addition, the highest magnitude AI vectors on the *y-z* grid seem to emanate from the $F_3^{\#}/F_4^{\#}$ note areas from both the front and rear of the pan. Vector orientations on the *y-z* grid suggest that the $F_4^{\#}$ note (top of the pan in this perspective) is the dominant contributor to radiation at this frequency.

The above conclusion is fully supported by the RI plots on Fig. 5. The reader will once again note the circulatory pat-

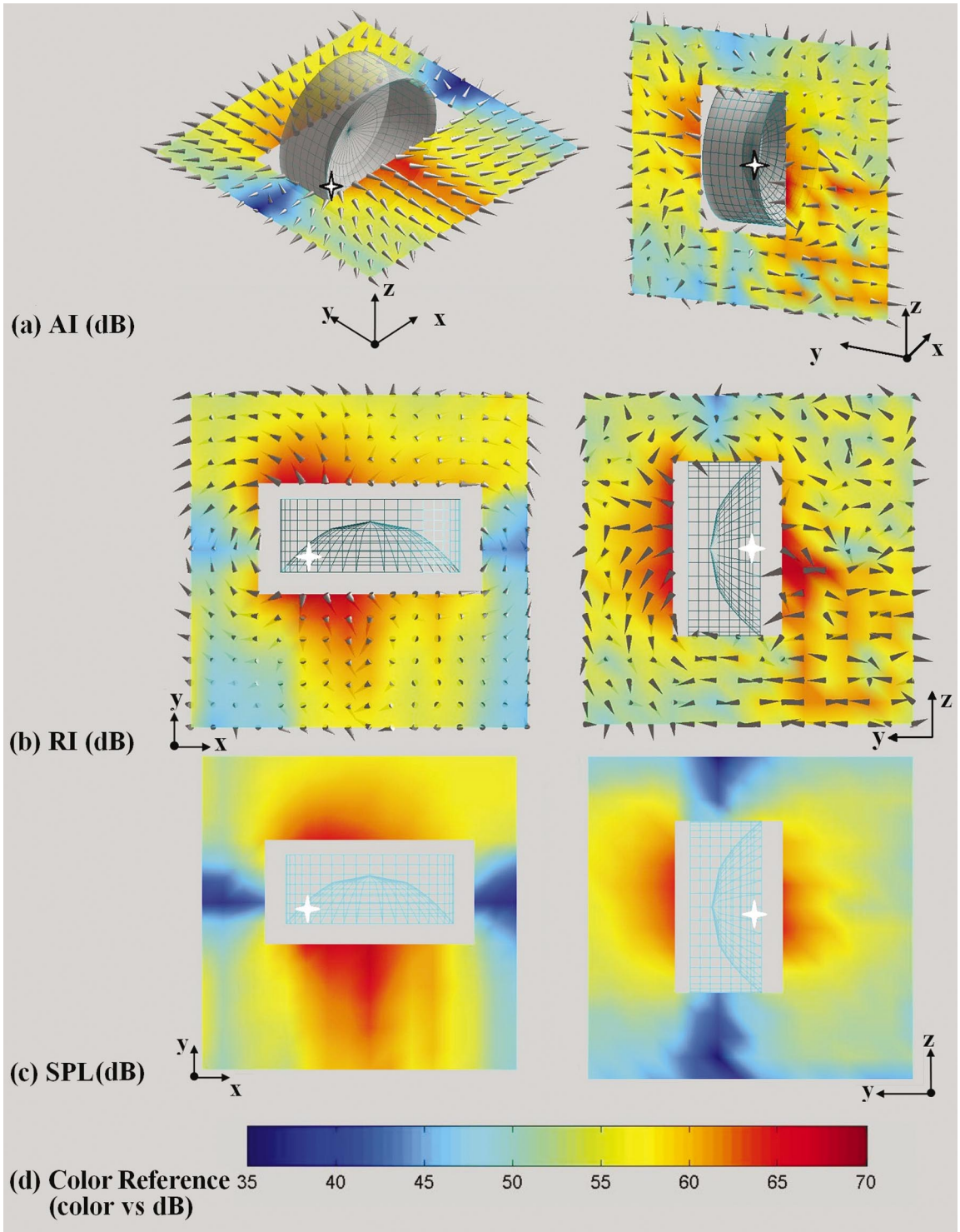


FIG. 4. Sound intensity and sound pressure level plots in x-y and y-z grids surrounding the Alexis double second pan for the F#3 excited at the fundamental frequency. The color reference bar applies to all plots in this and subsequent figures. Coordinate directions are shown. (a) Active intensity (AI), (b) reactive intensity (RI), (c) sound pressure level (SPL), and (d) color reference bar. All measurements are given in decibels.

terns formed by the RI vectors on the x-y grid. These now emanate from the area of the pan occupied by the $F_4^{\#}$ note. The RI plot for the y-z grid is even more conclusive; the circular vector field pattern now exists to the top of the pan, again in the location of the $F_4^{\#}$ note.

AI and RI vector patterns also suggest the existence of two distinct areas of radiation, one in the region of the $F_4^{\#}$ note and another more significant source directly opposite in the area of the A_3^b and C_4 notes. Figure 1 shows no vibrating source in this region although modal studies have revealed that there is significant vibration of the A_4^b note at very high excitation levels. One is, however, lead to the conclusion that the effect is produced by pan geometry through reflection and/or diffraction of the wave propagating from the $F_4^{\#}/F_3^{\#}$ area. RI vector animation seems also to support this view.

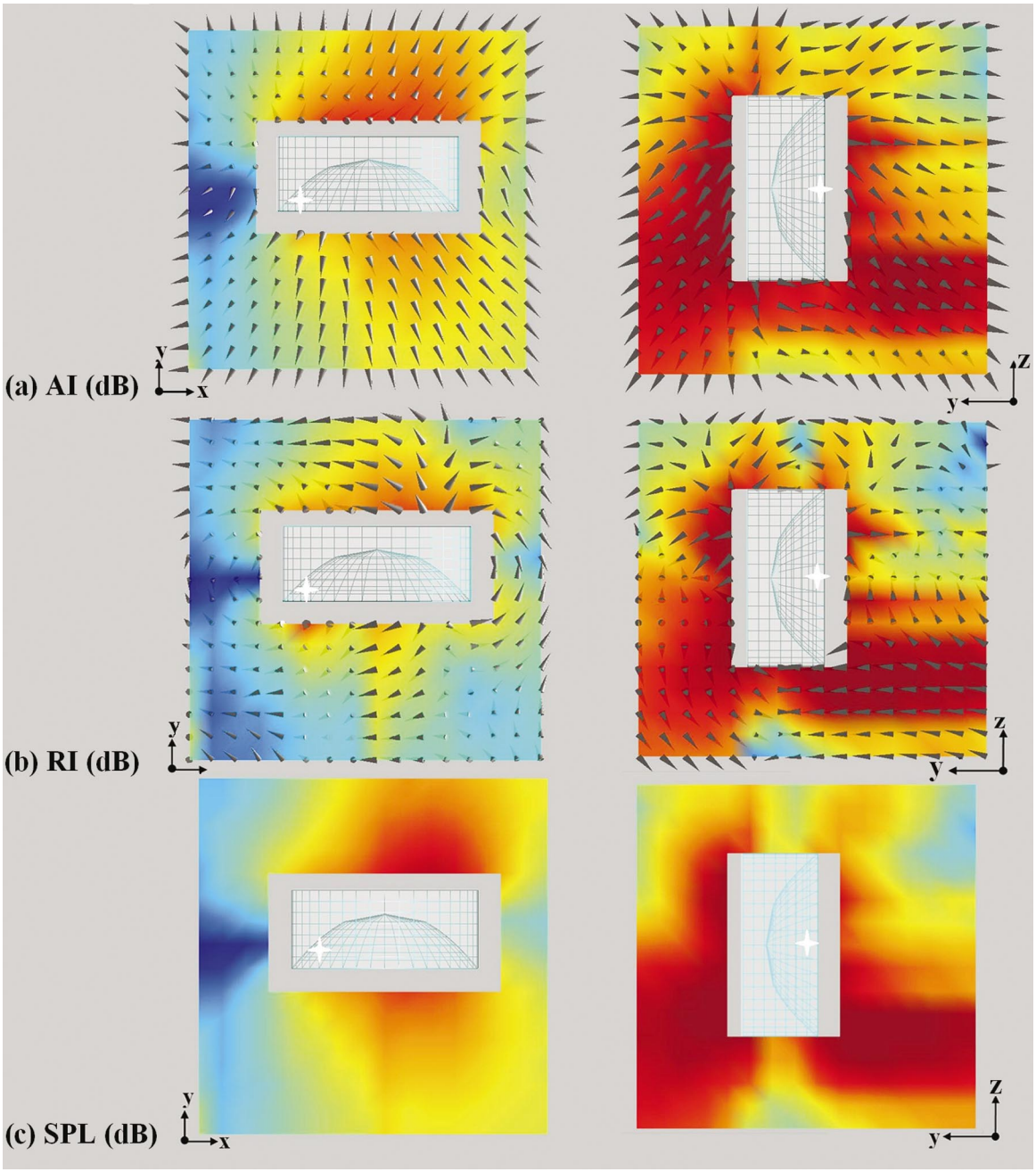


FIG. 5. Sound intensity and sound pressure level plots in x-y and y-z grids surrounding the Alexis double second pan for the F#3 note excited at the octave. Coordinate directions are shown. (a) Active intensity (AI), (b) reactive intensity (RI), and (c) sound pressure level (SPL). Colors refer to color reference bar in Fig. 4.

The AI and SPL plots obtained at this frequency suggest the existence of two major lobes to the front and one to the rear, the intensity being stronger to the rear. The rear lobe is oriented toward the right lower section of the pan, away from the $F_3^{\#}/F_4^{\#}$ areas. The frontal lobes seem to be directed toward the right upper section and the left lower section, respectively, with the latter appearing to be the more intense one. AI vector orientation [Fig. 5(a)] in the vicinity of the top left edge of the pan seems to suggest that the hornlike cavity created by the junction of the skirt and the rear of the playing surface may have a small but significant contribution to the radiation patterns. However, a conclusive assessment is not possible without a more detailed study on this aspect of the sound field.

The sound pressure plots shown Figs. 4(c) and 5(c) largely support the conclusions drawn from the other two plots. Pressure is greater in front of the pan for the fundamental and in the back for the second partial. This suggests that sound heard from the rear of the pan may be brighter than in front. The variation in directivity with frequency also clearly indicates how the instrument timbre depends on the position of the listener.

VI. RESULTS AND DISCUSSIONS FOR THE KELLMAN TENOR

Figure 6 shows color plots of AI, RI, and SPL for the first three partials. An outline of the pan has been included to

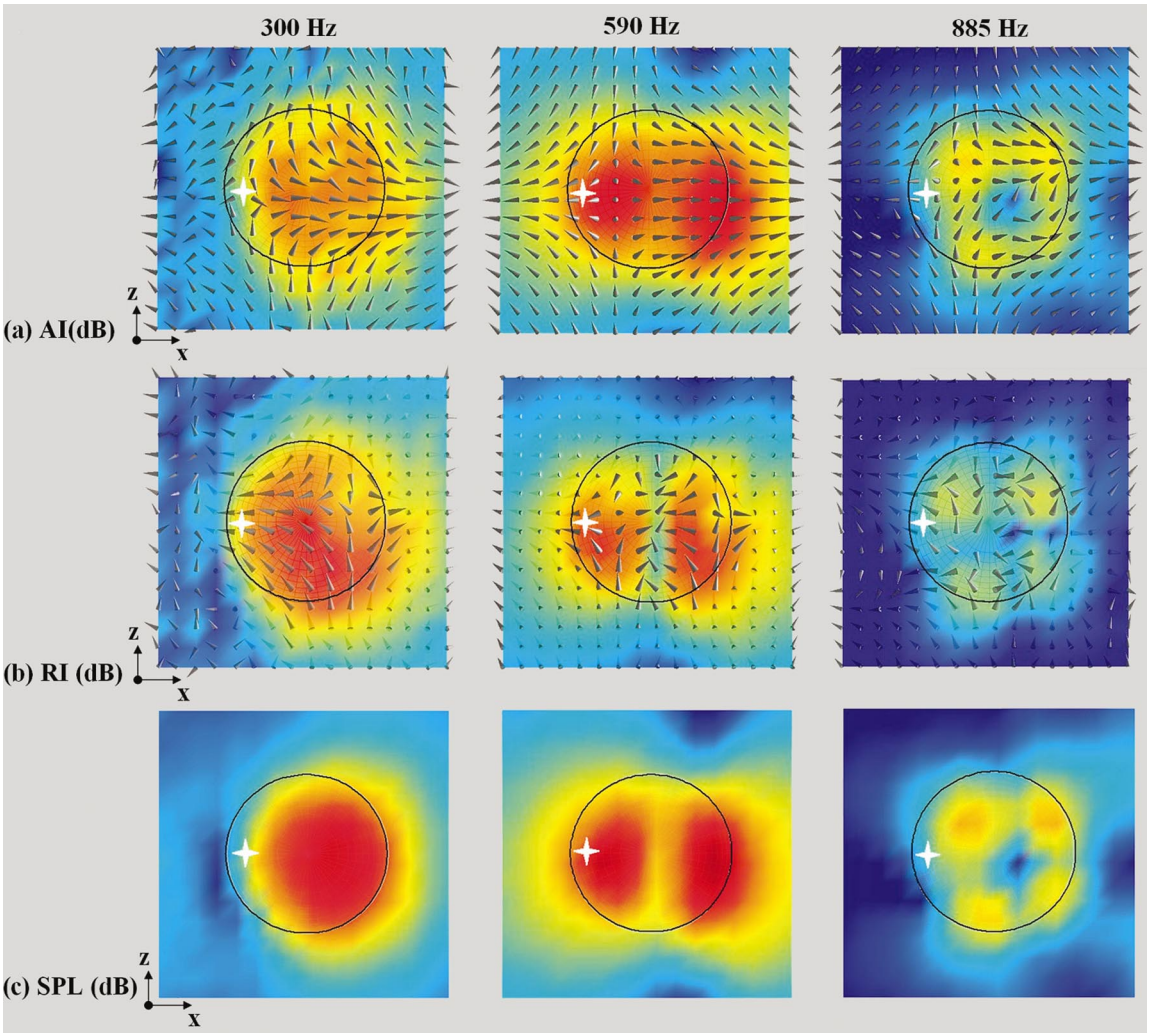


FIG. 6. Sound intensity and sound pressure levels in the x-z plane just above the Kellman low tenor pan for the D4 note (white cross) excited at its fundamental (300 Hz), octave (590 Hz), and third harmonic (885 Hz) frequencies. (a) Active intensity (AI), (b) reactive intensity (RI), and (c) sound pressure level (SPL). Colors refer to color reference bar in Fig. 4. Outline of the pan, centered in each frame, is barely visible.

provide the proper orientation. The D_4 note is located at the left center of the illustrated pan. The results clearly indicate that the acoustic radiation patterns assume monopole, dipole, and quadrupole patterns at the first, second, and third partials (fundamental, octave and third harmonic), respectively.

It must be emphasized that the patterns seem to emanate from the entire pan and NOT directly from the D_4 note. This is clearly supported by the AI and RI plots which show vectors of high magnitude over much of the grid directly over the pan.

The vector orientations in the AI plots seem to suggest the existence of a major frontal lobe for the first partial oriented to the side of the pan away opposite to the D_4 note. In particular, the results show that the SPL is lowest in the area of the rim closest to the excited note. For the octave, the left lobe appears to be aligned parallel to the cylindrical axis while the right lobe appears to be oriented to the area directly opposite the D_4 area. This is repeated for the quadrupole patterns produced at the third partial—the two lobes to the left seem to be aligned with the cylindrical axis while the other two lobes are slanted to the right. The upper pair of lobes is more intense than the lower pair.

As was the case for the Alexis pan, the RI vectors form circulatory patterns around the instrument. For the first partial the largest vectors originate in the area of the pan and enter the grid outside the pan. This suggests the existence of a circulatory pattern between the front and back. For the second partial there seems to be three sets of circulatory patterns (viewed along the y axis). There is a band of low intensity running vertically through the middle of the pan. RI vectors originate from both high intensity zones and converge on this area; this pattern occupies the middle third of the pan. The remaining RI vectors in the strong intensity areas closer to the rim appear to set up circulatory patterns on each side of the middle pattern from the bowl to the outside of the pan. Similar results hold for the quadrupole—RI vectors emanate from the areas of strong intensity and converge in areas of weak reactive intensity.

The reader is referred to Fig. 2(b), which shows that two vibrating sections on D_4 exhibit approximately the same amount of deflection; their antiphase relationship would clearly set up a null at distances equidistant from each section—a classic dipole. The contribution of the vibration of the D_5 note should simply produce a shift in the location

and disposition of the null. This should explain the line of low intensity in all second partial results.

For the third partial, the modal orientations shown in Fig. 2(c) should also significantly contribute to the quadrupole pattern. This was clarified by taking sound intensity measurements after the A_4 and A_5 notes were muffled. With the D_4 note now the only significant radiator, the two leftmost lobes disappeared leaving a dipole pattern with a horizontally aligned null line—clearly resulting from the (1,0) D_4 modal pattern.

The patterns intensity distributions for this pan have a high degree of symmetry. This can probably be attributed to the high degree of physical symmetry in the instrument design layout—a characteristic inherited from the fourths and fifths layout template. The tuning technique, which defines the alignment of the various modes, also plays a very significant role in this regard.

On the issue of instrument timbre, it is more obvious here than in the case of the Alexis pan that only the fundamental contributes significantly to the sound field along the cylindrical axis of the pan. This suggests that the sound of the pan tends to increase in brightness as one moves towards the rim. This is of significance to the issue of live performance and microphone placement for ''good'' reproduction. In the case of live performances, the front of the instrument, which has the most intense radiation, faces the pannist and the rear of the instrument faces the audience.

VII. CONCLUSION

This paper has presented the results of sound intensity measurements on a double second pan made by Clifford Alexis and a low tenor made by Bertrand Kellman. The study on the Alexis pan suggests that sound intensity emanating from the skirt regions next to and directly opposite to the excited note is significantly lower than that from the either side of the playing surface. It also shows that the sound from the rear of the pan is ''brighter'' than that from the front. In addition, in both cases, higher partials are subdued along the cylindrical axis to the front of the instrument. For the frequencies analyzed, first partial projection is essentially symmetric about the cylindrical axis while higher partials tend to be projected directly away from the excited note.

These results were obtained in an anechoic chamber and are therefore representative of a free-field analysis. In practice the pans would be positioned as much as 30° to 45° to the vertical. Apart from the effect of the acoustics of the performance environment, there would be some absorption due to the presence of the player who is invariably positioned in front of the instrument.

This study would clearly benefit from volumetric (full three-dimensional) measurement. The major problem here, of course, is the sheer effort of gathering the quantity of data required for meaningful assessment. This, however, can be very accurately addressed using near-field acoustic holography methods[14] that facilitate accurate estimation of the sound field using measurements on a well-defined grid. Another possible and potentially interesting study would be that of the sound field in closer proximity to the pan; the current study was restricted in this regard by the physical geometry of the Ono Sokki M6410 microphone.

ACKNOWLEDGMENTS

The authors wish to acknowledge the kind support of Clifford Alexis and Bertrand Kellman whose pans were used in this study. This work was funded in part by a University of the West Indies British Petroleum Trinidad and Tobago Senior Fellowship Award.

[1] Trinidad and Tobago Bureau of Standards, *Results of Steelpan Survey 1990*, Trinidad and Tobago Bureau of Standards, Macoya, Trinidad and Tobago, 1990.
[2] U. Kronman, *Steelpan Tuning*, Musikmuseet, Stockhom, 1991.
[3] T. D. Rossing, *Science of Percussion Instruments, Series in Popular Science-Vol. 3*, World Science, Singapore, 2000.
[4] T. D. Rossing, D. S. Hampton, and U. Hansen, ''Music from oil drums: The acoustics of the steel pan,'' Phys. Today **49**(3), 24–29 (1996).
[5] T. D. Rossing, D. S. Hampton, and U. Hansen, ''Vibrational mode shapes in caribbean steelpans. I. Tenor and double second,'' J. Acoust. Soc. Am. **108**, 803–812 (2000).
[6] L. E. Murr, E. Ferreyra, J. G. Maldonado, E. A. Trillo, S. Pappu, C. Kennedy, J. DeAlba, M. Posada, D. P. Russell, and J. L. White, ''Materials science and metallurgy of the Caribbean steel drum-Part I. Fabrication, deformation phenomena and acoustic fundamentals,'' J. Mater. Sci. **34**, 967–979 (1999).
[7] C. Imbert, ''Investigation of commercial production techniques of steelpan instruments,'' Project No. E-SBR-73-5-1, Caribbean Industrial Research Institute (CARIRI), Tunapuna, Trinidad, 1977.
[8] F. Muddeen and B. Copeland, ''The polar response of a tenor steelpan,'' *Proceedings of The First International Conference on the Science and Technology of the Steelpan*, (ISBN 976-620-164-1), edited by Anthony Achong (Department of Physics, University of the West Indies, St Augustine, Trinidad and Tobago), **1**, 115–127 (2002).
[9] L. M. Wang and C. B. Borroughs, ''Acoustic radiation from bowed violins,'' J. Acoust. Soc. Am. **110**, 543–555 (2001).
[10] F. J. Fahy, *Sound Intensity*, Elsevier Science, London, 1989.
[11] J. Adin Mann III, J. Tichy, and A. J. Romano, ''Instantaneous and time-averaged energy transfer in acoustic fields,'' J. Acoust. Soc. Am. **82**, 17–30 (1987).
[12] M. J. Crocker and F. Jacobsen, in ''Sound Intensity,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker, Wiley, New York, 1997.
[13] F. Jacobsen, ''A note on instantaneous and time-averaged active and reactive sound intensity,'' J. Sound Vib. **147**, 489–490 (1991).
[14] E. G. Williams, *Fourier Acoustics: Sound Radiation and Near-Field Acoustical Holography*, Academic, San Diego, 1999.

Designing musical structures using a constrained optimization approach

J. Petrolito and K. A. Legge
Department of Physical Sciences and Engineering, La Trobe University, Bendigo, Australia

(Received 25 August 2004; revised 8 October 2004; accepted 10 October 2004)

The paper discusses the application of numerical optimizing techniques to musical structures. It is shown that the natural tendency to emulate the instrument maker by optimizing the characteristic frequencies of the structure is not always the most appropriate strategy, and that in cases where one or more possible solutions exist it is more appropriate to specify the frequency requirements as a constraint. The strategy is illustrated using a xylophone bar. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1828572]

PACS numbers: 43.75.Kk, 43.40.Cw [NHF] Pages: 384–390

I. INTRODUCTION

The optimization of a musical instrument has traditionally been the task of the maker during the construction phase. The instrument maker generally crafts structures such that they respond to excitation with sounds that please. The optimal design for the instrument maker is therefore one for which the tuning requirements are best obtained. The optimization of physical structures such as a building or bridge, on the other hand, is generally performed by a structural engineer during the design phase. The optimal design for the engineer is typically one that minimizes cost or amount of construction material used, and a numerical optimization strategy is normally employed in the design. In this case, the behavior of the structure is described in terms of a system of differential equations and a solution to the equations is sought numerically.

Numerical techniques are not new to musical acoustics. For instance, in recent years the study of percussive bars, such as the xylophone bar, has been performed using finite-element techniques,[1–7] whereby the bar is discretized into many small parts to approximate the profile. Similarly, finite-element analysis has been used to study a diverse set of musical problems, such as the design of a pentangle,[8] the vibrational patterns of a violin-shaped air cavity,[9] and the design of a bell profile.[10,11]

Numerical optimization is a design approach that can be used in conjunction with finite elements or any other numerical technique. Like any new tool, however, it is tempting to use it in the way that is most familiar, and to thereby overlook its full power. The designers of musical structures have foremost in their minds that they require optimal tuning and this, it would seem, inevitably leads to the use of an optimizing function in terms of the required frequencies. For example, Schoofs *et al.*[10] used such an optimization procedure for the tuning of church bells, and more recently Henrique and Antunes[7] used a similar procedure in the design of mallet percussion instruments. Such a strategy inevitably leaves the user with the task of determining the suitability of the optimal solution, that is, how close to the desired frequency should each mode be, and which mode is more important to get correct?

In a previous paper,[2] we outlined a numerical optimization technique for the design of xylophone and marimba bars such that the bar, described as a series of piecewise-constant height elements, responded with predetermined frequency characteristics. We have also discussed[12] the role of the optimizing function and constraints on the numerical procedure. In the current paper, we present a general procedure for applying numerical optimizing techniques to musical structures by examining the assumptions made by both the musician and the mathematician. We seek to provide a suitable balance between constraints and desirable attributes such that a suitably tuned structure is produced. The paper focuses on the general procedure of optimization of any musical structure and refers to the xylophone or marimba bar as an illustration.

II. MATHEMATICAL MODELS

The goal is to design a structure that responds with specified frequencies. The first step is to formulate the equations that govern the motion of the vibrating structure. These equations depend on the model adopted, and the first decision is therefore the choice of model to be used to described the structure. The structure may be modeled using a one-, two-, or three-dimensional theory as appropriate.

For example, the dominant response of a struck xylophone bar is transverse to its longitudinal axis, and the motion is therefore beam-like in nature. Hence, a one-dimensional model that can account for shear deformation is usually considered suitable.[2] However, when the profile of the bar entails sudden jumps in height, the one-dimensional model fails to accurately account for the complex stress system set up around the sharp corners. In this case, a two-dimensional model is more accurate.[13] Similar consideration would apply to any musical structure. For example, a bell requires a three-dimensional model.

We discuss two models for a xylophone bar below.

A. One-dimensional model for a xylophone bar

For the one-dimensional model of the bar, Timoshenko's beam theory[14] is adopted to account for the effect of shear deformations on the natural frequencies, which can be sig-

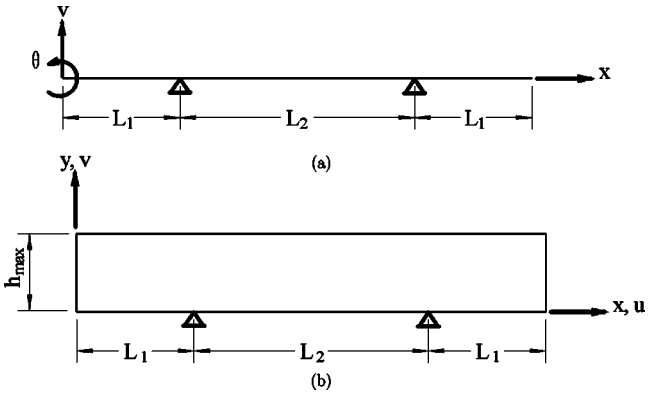


FIG. 1. Models for a xylophone bar. (a) One-dimensional model. (b) Two-dimensional model.

nificant for shorter bars. The primary variables are the transverse displacement, $v(x)$, and rotation, $\theta(x)$, which are assumed to be independent [see Fig. 1(a)].

As we are only interested in the natural frequencies, we can formulate the problem in terms of the amplitudes of the displacement and rotation of the beam as it vibrates, which are denoted by $\bar{v}(x)$ and $\bar{\theta}(x)$. The theory results in a fourth-order system of coupled ordinary differential equations, namely

$$kGA(\bar{v}'' - \bar{\theta}') + \rho A\Omega^2 \bar{v} = 0, \tag{1a}$$

$$EI\bar{\theta}'' + kGA(\bar{v}' - \bar{\theta}) + \rho I\Omega^2 \bar{\theta} = 0, \tag{1b}$$

where E is Young's modulus, $G = E/[2(1+\nu)]$ is the shear modulus, ν is Poisson's ratio, I is the section modulus, A is the cross-sectional area, ρ is the material density, k is the shear correction factor, which is usually taken as 5/6 or $\pi^2/12$, and Ω is the natural frequency. A prime denotes differentiation with respect to x. Since the task involves varying the geometry to obtain a suitable tuning regime, the sectional properties A and I are also functions of x.

Finally, the boundary conditions at $x = L_1$ and $x = L_1 + L_2$ are considered. The supports can be modeled by simple springs that allow motion in the vertical direction only. If each spring constant is selected as an option, the supports can be selected as any condition within the range of simple support to completely free. The calculation of the natural frequencies leads to the solution of a linear eigenvalue problem.

B. Two-dimensional model for a xylophone bar

If instead a plane stress theory for the two-dimensional model of the bar is adopted, the primary variables are the axial displacement, $u(x)$, and the transverse displacement, $v(x)$ [see Fig. 1(b)].

In terms of the amplitudes of vibration, $\bar{u}(x,y)$ and $\bar{v}(x,y)$, the governing partial differential equations are[15]

$$\frac{E}{1-\nu^2}\left[\bar{u}_{,xx} + \frac{(1-\nu)}{2}\bar{u}_{,yy} + \frac{(1+\nu)}{2}\bar{v}_{,xy}\right] + \rho\Omega^2\bar{u} = 0, \tag{2a}$$

$$\frac{E}{1-\nu^2}\left[\bar{v}_{,yy} + \frac{(1-\nu)}{2}\bar{v}_{,xx} + \frac{(1+\nu)}{2}\bar{u}_{,xy}\right] + \rho\Omega^2\bar{v} = 0, \tag{2b}$$

where a comma denotes partial differentiation. The boundary conditions are the same as for the one-dimensional model. A comparison of optimization results for a xylophone bar modeled as a one-dimensional system and as a two-dimensional system has been presented previously.[13]

Whatever the model, analytical solutions of the governing equations can only be obtained for simple boundary conditions and geometries.[14] As the tuning of a musical structure requires varying the sectional properties, analytical solutions are generally not possible, and hence a numerical solution is required. Moreover, the problem is not simply one of solving an eigenvalue problem for a particular structure, but rather a case of finding a structure that will produce the required frequencies.

III. STRUCTURAL GEOMETRY

The design of a structure that will respond in a predetermined manner involves describing the geometry of the structure by a number of parameters. These parameters are the primary unknowns for the problem and will be varied to produce the required outcome.

The following examples illustrate ways in which the undercut of a xylophone bar may be described. The height of the bar is defined as the function $h(x)$. Clearly, a wide range of geometries can be adopted within this general framework,[3] and three distinct geometries are presented below. In each case, the cut section is restricted to be between the two supports for convenience. In addition, a symmetrical profile is assumed in all cases, and hence only half the beam is considered in any description.

A. Piecewise-constant heights

A common attempt is to discretize the beam as a series of piecewise-constant heights.[2] Figure 2(a) shows the beam discretized into N sections each of length, l_i, and height, h_i, $i = 1,2,...,N$. The response of the beam is thus a function of h_i, with constraints imposed on the heights. For ease of construction, the height variations are restricted so that no material is added, and hence a maximum height $h_{\max}$ is stipulated. Furthermore, to ensure that the beam remains structurally sound, a minimum height $h_{\min}$ is also stipulated. Hence, the variables h_i are bounded in the range

$$h_{\min} \leqslant h_i \leqslant h_{\max}, \quad i = 1,2,...,N. \tag{3}$$

B. Piecewise-linear heights

The simplest extension of the previous geometry is to require that there are no jumps in the profile of the beam.[3] Hence, the beam is again discretized into N sections each of length l_i. However, this time the profile has linearly varying heights. Figure 2(b) shows an example of this discretization.

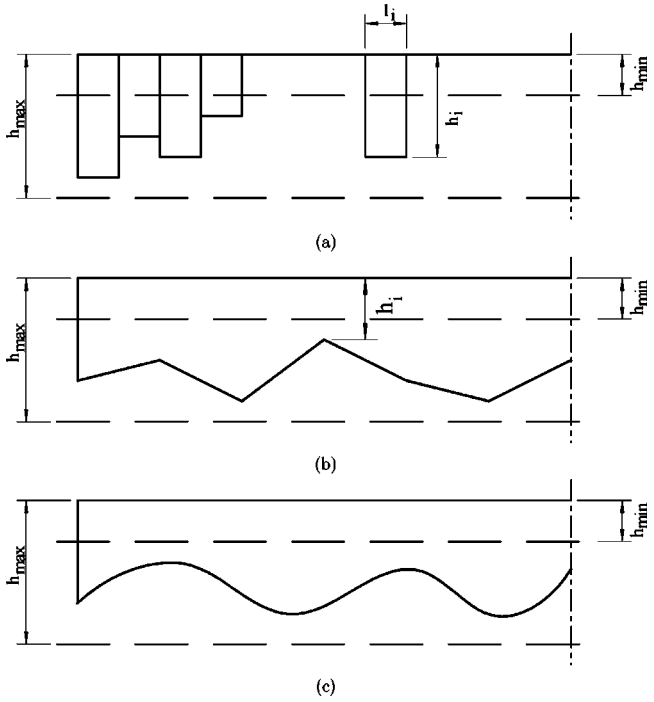


FIG. 2. Discretization of longitudinal profile. (a) Piecewise-constant heights. (b) Piecewise-linear heights. (c) Sine curves.

C. Sinusoidal functions

Smoother profiles can be generated by introducing more complex geometries such as piecewise-quadratic or cubic functions or sinusoidal profiles[3] as shown below. Hence, the beam is no longer discretized in appearance [see Fig. 2(c)]. A variety of approximations was explored, although just one is detailed below. To provide a general approximation to the height, the height is assumed to be a combination of a linear function plus a truncated sine series. Hence, the height of the beam between the two supports is taken as

$$h(x)=h_{\max}-\frac{2a_0x}{L_2}-\sum_{n=1}^{N-1}a_n\sin\frac{2n\pi x}{L_2},\quad x\leqslant\frac{L_2}{2},\qquad(4)$$

where the a_n are the amplitudes of the sinusoids for $n\geqslant 1$, and x is measured from the left support.

IV. NONLINEAR EQUATION SOLUTION STRATEGY

In general, the problem involves the tuning of a number of frequencies. Hence, a simple solution strategy could be developed as follows. If N_f natural frequencies are specified, the number of variables is taken as $N=N_f$. The frequencies are thus nonlinear functions of the variables, that is

$$\Omega_i=f_i(h_j),\quad i,j=1,2,...,N.\qquad(5)$$

For given values of the frequencies, these equations can be solved to find the unknown parameters using a standard technique for nonlinear, simultaneous equations.[16]

This method is not ideal for the following reasons. First, there need not be a solution to the nonlinear system, or the solution may not be unique, although the latter possibility is not critical. Second, and more importantly, the constraint $N=N_f$ does not allow sufficient latitude for describing the shape of the structure. Hence, a more general technique is required that allows more freedom in defining the shape, and this can be achieved by using an optimization approach as outlined below.

V. OPTIMIZATION STRATEGIES

Optimization generally implies that one is choosing the better solution, but it does not necessarily imply that the solution is perfect. For a musical structure, however, the required frequencies are critical, and the preference would be to get them perfect. In this section, we outline the crux of our argument, namely the choice of optimizing strategy. We illustrate the difference between a strategy that optimizes the frequency requirements and a strategy that constrains the frequency requirements. We have shown previously[12] that the optimizing function for the piecewise-constant height xylophone bar need not be the most obvious choice. We repeat our arguments here and generalize to include all geometries.

From a computational perspective, it is also important to determine the scope of the optimization, that is, is a local or a global optimum solution required? Global optimization techniques generally require significantly more computational effort.[17]

A. Optimization of characteristic frequencies

To emulate the instrument maker numerically, it seems natural to optimize the characteristic frequencies. Hence, we are attempting to tune a musical structure so that the lowest N_f frequencies are specified as $\bar{\Omega}_i$, $i=1,2,...,N_f$. Numerically, we need to define a function, f, of the required frequencies, and minimize it. The exact form of f is somewhat arbitrary. One possibility is to define the optimizing function as

$$f(h_j)=\sum_{i=1}^{N_f}w_i(\bar{\Omega}_i-\Omega_i(h_j))^2,\quad j=1,2,...,N,\qquad(6)$$

where w_i are positive weights that reflect the relative importance of the required frequencies.

The form chosen in Eq. (6) ensures that f is positive and differentiable, and that it has a minimum value of zero. However, since the function generally has more than one local minimum, a local optimization procedure is not guaranteed to find the desired value of zero, and this difficulty was noted by Schoofs *et al.*[10]

In general, f is a multidimensional function, but to illustrate the potential difficulties we will consider the one-dimensional case. Figure 3 depicts a fictitious function f as a function of one variable. Points 1, 2, and 3 are the local minima, with point 3 being the global minimum for the range plotted. A local optimization procedure may find any one of these points. However, whether any of the minima are acceptable solutions depends on how close to zero the value of the optimizing function is for a particular solution.

Hence, by constricting the procedure to one that numerically emulates the maker, one may not have gained much of an advantage. There is still a need to consider the suitability of the optimal solution, that is (a) how small does the minimal function need to be to produce a musically acceptable

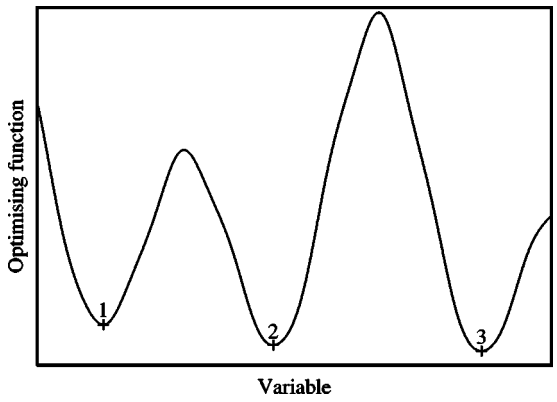


FIG. 3. One-dimensional optimizing function.

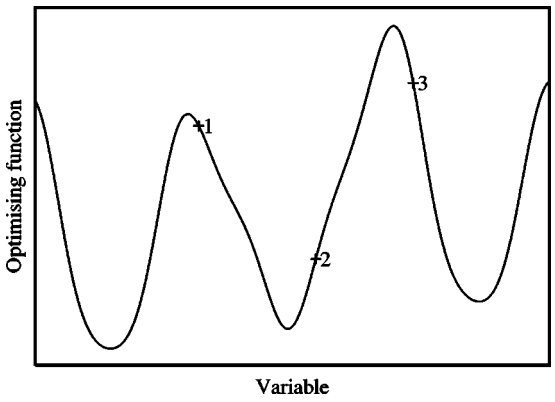


FIG. 4. One-dimensional optimizing function with constraints.

result, and (b) what weighting is given to the various frequencies, or in other words, which mode is more important?

B. Constrained optimization

Instead of the seemingly ''natural'' approach that attempts to emulate the instrument maker, an alternative strategy is to set the frequency regime as a constraint, and to choose a less obvious function to optimize. This approach, known as constrained optimization, allows considerable freedom in the choice of optimizing function.[12]

Consider the following optimization problem:

$$\text{Minimize } f(h_i) \text{ subject to } g_j(h_i)=0,$$

$$i=1,2,...,N, \quad j=1,2,...,N_c, \tag{7}$$

where N_c is the number of constraints for the problem. The numbers N and N_c define the dimensions of the problem, and $N \geqslant N_C$ is required for a solution. The number of possible solutions for the problem is determined as the difference $N-N_c$, and in general, more solutions are obtained as $N-N_c$ increases.

In this paper, the first N_f constraints are associated with the required frequencies, that is, we set

$$g_j=\overline{\Omega}_j-\Omega_j(h_i), \quad j=1,2,...,N_f. \tag{8}$$

The remaining N_c-N_f constraints can be associated with any other factors that are deemed appropriate, for example, manufacturing constraints.

The constrained optimization problem is thus uniquely specified by three things, namely, the geometry parameters, the optimization function, and the constraint functions, and different solutions may be obtained by varying any of these. In principle, f can be any function, and this leads to an almost limitless range of possibilities. For example, issues with musical significance such as the relative amplitude and damping of vibration modes could be considered providing they can be couched in appropriate mathematical terms, that is, as functions of the geometry parameters. Our point is that, for musical structures where correct frequencies are required, the required frequencies should not be associated with f but rather with the constraints g. Since the first N_f constraints are associated with the required frequencies, any optimum solution generated is at least musically acceptable.

To compare this approach with the previous one, we again consider the one-dimensional case. Figure 4 depicts another fictitious function to be optimized. In this case, the function is constrained so that points 1, 2, and 3 satisfy the constraints, namely the frequency regime requested. From the point of view of the musician, any of these solutions is viable if it produces a nice sound. In fact, there may be no reason to optimize at all as any of the points are musically acceptable.

C. Optimizing the xylophone bar

To illustrate the constrained optimization procedure for a structure for which many solutions exist, the optimization of a xylophone bar is again considered. The numerical techniques used to solve the problem have been detailed previously.[1] The optimization technique used is a sequential quadratic programming approach, and the frequencies were calculated using a thick beam finite element. Since the method is iterative, an appropriate error tolerance should be specified to determine the stopping criterion, for example, to dictate how accurately the constraints are satisfied. The specified error tolerance should reflect the required accuracy of the solution.

As noted above, f can, in principle, be any function deemed appropriate to the problem. Four possible forms of f are discussed below.

1. Optimizing the piecewise-constant height xylophone

A structural engineer may typically use optimization to ensure that the cost of construction of the structure is minimized. This approach can be simulated by choosing an optimizing function that, for example, minimizes the amount of material that needs to be removed from the bar, so that

$$f(h_i)=\sum_{i=1}^{N}(h_{\max}-h_i)l_i, \tag{9}$$

since the width of the beam remains constant. However, such considerations may be of less concern to the instrument maker.

On the other hand, the musician may have other reasons to optimize. For example, the limitation of the one-dimensional model used to describe the bar is such that sudden jumps in height between adjacent sections are

undesirable.[2] The choice is either to use a more complex two-dimensional model[13] or to use the one-dimensional model and choose an optimizing function that produces a smoother profile. For example, the optimization criterion can be adapted to generate smooth profiles by stipulating that the differences between two adjacent heights be minimized. One possible way of achieving this is to choose

$$f(h_i)=\sum_{i=1}^{N-1}(h_{i+1}-h_i)^2+\alpha\sum_{i=1}^{N-2}(h_{i+2}-h_i)^2, \tag{10}$$

where α is a selectable weighting factor. The form of the above criterion has been chosen to ensure that f is always positive and differentiable, and the second term is included to achieve a smoother overall profile.

2. Optimizing the piecewise-linear height xylophone

Smoother profiles can also be obtained with a different choice of profile. For the piecewise-linear profile, the beam is again discretized into sections of length, l_i. However, in this case the profile has linearly varying heights. The optimization criterion used is such that there is a minimum difference in slope between adjacent sections. Hence, for constant l_i, f is taken as

$$f(h_i)=\sum_{i=1}^{N-1}(h_{i+1}-2h_i+h_{i-1})^2, \tag{11}$$

where the slope difference has been squared for convenience.

3. Optimizing the sinusoidal profile xylophone

Further smoothness of the profile can be generated by introducing more complex geometries such as the sine series approximation. The optimization criterion can then be chosen for ease of manufacture by, for example, ensuring that the curvature of the cut is minimized. Hence, in this case f is taken as

$$f(a_n)=\sum_{n=1}^{N-1}n^4a_n^2, \tag{12}$$

where f is proportional to the sum of the squared second derivatives for constant l_i.

D. Global optimization

A local optimization technique will only produce a possible optimal solution rather than the true global optimum. If the global optimum is sought, a different optimization technique is required.[17] However, global optimization techniques are computationally more expensive, and hence the benefits of global optimization need to be justified relative to the computational cost. In addition, the relative merits of different approaches to global optimization for this class of problem need to be assessed. For example, direct search methods do not appear to be particularly suitable for current formulation. The most time-consuming part of the procedure is the calculation of the natural frequencies for a given profile. Since it is essential that these be correct, a random search that does not produce the required frequencies is of little use.

In the first optimizing approach outlined above, where the frequencies are optimized, a local optimization procedure may not be sufficient. A local optimization procedure may find the local minimum for the optimizing function, but the value of f may not be sufficiently close to zero. In this case, a computationally expensive global optimization procedure may need to be performed.

In the constrained optimization approach, where the frequency regime is a constraint, and considerable freedom is allowed in the choice of optimizing function, optimization of relatively simple structures, such as the xylophone bar, can result in a choice of musically acceptable solutions. In this case, global optimization is more usefully performed with the function value associated with each solution forming the basis of any comparison when searching for a global optimum. It should, however, be noted that the global optimization criterion need not be the same as the local optimization criterion, and this again leads to many possibilities.

In this paper, we consider two such approaches. In each case, a specified number of acceptable profiles is generated using random starting points for the parameters and local optimization. The global optimum was then chosen based on one of two criteria, namely

(1) the solution that gave the smallest value of the optimizing function f, or
(2) the solution for which minor changes in the profile parameters had the least effect on the frequency regime, ensuring that slight manufacturing errors have minimal effect.

VI. EXAMPLE

To illustrate the full procedure, we consider the design of suitable undercuts for a xylophone bar and present the range of possible profiles that satisfy the design criteria. The procedure requires well-defined physical properties, and for the one-dimensional model the properties are assumed to be isotropic. Hence, we shall illustrate the design procedure on a xylophone bar constructed of aluminum. The bar is sized as for a typical marimba bar, and the requested frequency regime follows the ratios for a possible marimba tuning regime (1:4:10).

We began with a solid piece of aluminum with dimensions $L_1=120\,\text{mm}$ and $L_2=300\,\text{mm}$. It had a width of 65 mm and an uncut thickness of 20 mm, and a minimum thickness of 4 mm was stipulated. The density and Young's modulus for aluminum were taken as $\rho=2700\ \text{kg}\cdot\text{m}^{-3}$ and $E=71\,\text{GPa}$, respectively. Poisson's ratio was assumed to be 0.3, and a shear correction factor of $\pi^2/12$ was used. The xylophone bar is assumed to be supported by spring-like supports, as indicated in Fig. 1(a), with a stiffness of $2.9\times10^4\ \text{N}\cdot\text{m}^{-1}$. For convenience, the height changes were restricted to the section of the bar between the supports. The tuning regime required was such that the first three natural frequencies should be 110, 440, and 1100 Hz, respectively.

The constrained optimization approach was used to solve the problem using ten parameters ($N=10$) and one of the following criteria:

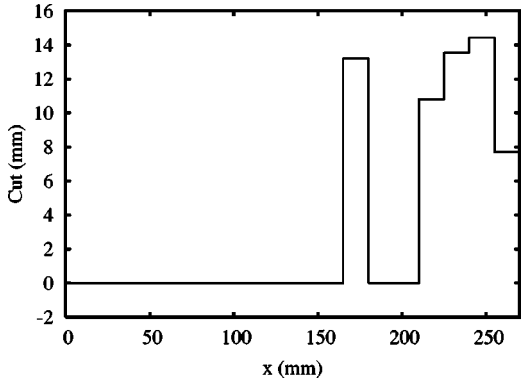


FIG. 5. Results for local optimization—case 1.

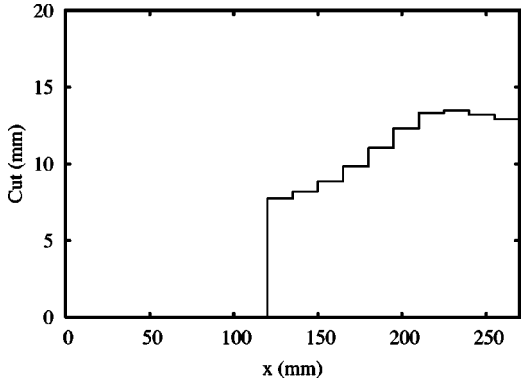


FIG. 6. Results for local optimization—case 2.

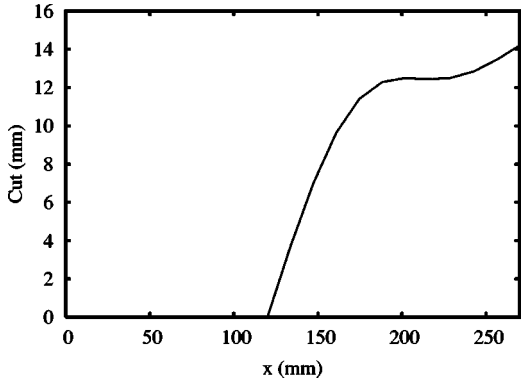


FIG. 7. Results for local optimization—case 3.

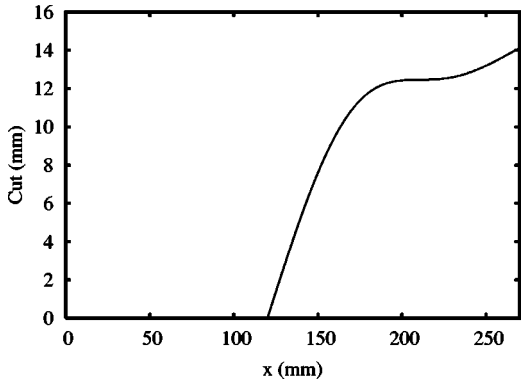


FIG. 8. Results for local optimization—case 4.

(1) piecewise-constant heights and using Eq. (9) for the optimizing function;

(2) piecewise-constant heights and using Eq. (10) with a value of $\alpha=0.5$ for the optimizing function;

(3) piecewise-linear heights and using Eq. (11) for the optimizing function, or

(4) a sinusoidal undercut with Eq. (12) as the optimizing function.

In all cases, a solution was initially sought using the uncut bar as a starting point for the iteration, that is, with all the h_i set to $h_{\max}$. To generate further local optimal solutions, the initial values of the heights were randomly generated in the range $h_{\min}$ to $h_{\max}$, and these were used as the starting point for the iteration. It should be noted that it is not possible to guarantee that a different solution is found using this procedure. Figures 5–8 depict a possible solution for each of the geometries. Only half the profile is plotted, and the complete profile is symmetric about the centerline. Since each is a solution, this shows that the choice of optimizing function is arbitrary, and that there are many possible profiles that satisfy the tuning requirements. It is interesting to note the similarity in shape for the piecewise-linear and sinusoidal solutions. This is perhaps not surprising, since both solutions are attempting to find a smooth profile.

To investigate the benefits of global optimization, 100 distinct iterations were performed using criterion 2 above. In the first instance, the solution giving the smallest value of the optimizing function from Eq. (10) was selected. In the second instance, the solution for which minor random changes in the parameters had the least effect on the frequency regime was selected. This produced 21 and 23 distinct solutions, respectively, from which the two globally optimal solutions were selected. These two solutions are shown in Figs. 9 and 10.

Again, it should be noted that the two global solutions are close, and theoretically either solution is acceptable. However, in attempting to construct one of these profiles, the solution that is least sensitive to manufacturing errors may be more suitable.

VII. CONCLUSIONS

The optimization of a musical instrument is a task usually performed by the maker during its manufacture. In contrast, the optimization of many nonmusical structures is

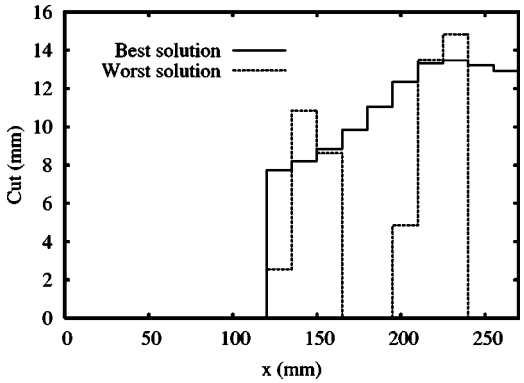


FIG. 9. Results for global optimization—heights.

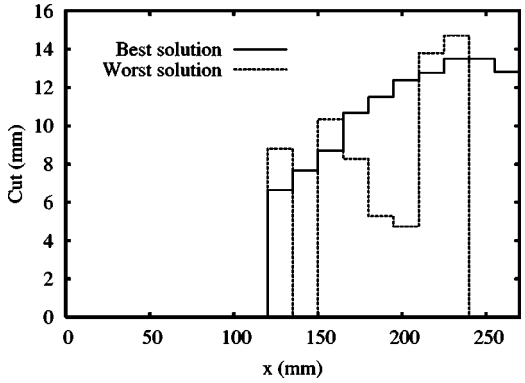


FIG. 10. Results for global optimization—sensitivities.

solved using a mathematical approach. In this paper, we have discussed the application of mathematical optimization strategies to structures such as musical instruments, showing that the procedure is not altogether intuitive. We have illustrated the strategy by considering the example of a xylophone bar. Current work by the authors is extending this approach to other musical structures in two and three dimensions.

[1] J. Petrolito and K. A. Legge, ''Optimising the xylophone: A numerical approach,'' in Proceedings, Computational Techniques and Applications Conference, pp. 625–630, Melbourne, 1995.
[2] J. Petrolito and K. A. Legge, ''Optimal undercuts for the tuning of percussive beams,'' J. Acoust. Soc. Am. **102**(**4**), 2432–2437 (1997).
[3] K. A. Legge and J. Petrolito, ''Tuning marimbas using numerical optimisation,'' in Proceedings, 5th International Congress on Sound and Vibration, Adelaide, 1997, Vol. 3, pp. 1641–1647.
[4] F. Orduna-Bustamante, ''Nonuniform beams with harmonically related overtones for use in percussion instruments,'' J. Acoust. Soc. Am. **90**, 2935–2941 (1991).
[5] J. Bretos, C. Santamaria, and J. Aloonso Moral, ''Finite element analysis and experimental measurements of natural eigenmodes and random responses of wooden bars used in musical instruments,'' Appl. Acoust. **56**, 141–156 (1999).
[6] B. H. Suits, ''Basic physics of xylophone and marimba bars,'' J. Acoust. Soc. Am. **69**, 743–750 (2001).
[7] L. L. Henrique and J. Antunes, ''Optimal designs and physical modelling of mallet percussion instruments,'' Acta Acust. (Beijing) **89**, 948–963 (2003).
[8] N. H. Fletcher, ''Tuning a pentangle—A new musical vibrating element,'' Appl. Acoust. **39**, 145–163 (1993).
[9] J. Bretos and C. Santamaria, ''Vibrational patterns of a violin-shaped air cavity obtained by finite element modeling,'' Acta Acust. (Beijing) **85**, 584–586 (1999).
[10] A. Schoofs, F. van Asperen, and P. Maas, ''Computation of bell profiles using structural optimization,'' Music Percept. **4**(3), 245–254 (1987).
[11] N. McLachlan, B. K. Nigjeh, and A. Hasell, ''The design of bells with harmonic overtones,'' J. Acoust. Soc. Am. **114**(1), 505–511 (2003).
[12] J. Petrolito and K. A. Legge, ''Optimisation of musical structures,'' in Proceedings, NAFEMS World Congress, Lake Como, Italy, 2001, Vol. 1, pp. 373–382.
[13] J. Petrolito and K. A. Legge, ''Two-level procedure for tuning stepped beams,'' in Proceedings, Australasian Conference on Structural Optimisation, Sydney, 1998, pp. 173–179.
[14] C. L. Dym and I. H. Shames, *Solid Mechanics: A Variational Approach* (McGraw-Hill, New York, 1973).
[15] J. L. Davis, *Wave Propagation in Solids and Fluids* (Springer, New York, 1988).
[16] J. E. Dennis and R. E. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
[17] Z. Michalewicz and D. B. Fogel, *How to Solve It: Modern Heuristics* (Springer, Berlin, 2000).

Predicting the similarity between expressive performances of music from measurements of tempo and dynamics

Renee Timmers[a)]
Austrian Research Institute for Artificial Intelligence, Freyung 6/VI, 1010 Vienna, Austria

(Received 27 October 2003; revised 19 October 2004; accepted 26 October 2004)

Measurements of tempo and dynamics from audio files or MIDI data are frequently used to get insight into a performer's contribution to music. The measured variations in tempo and dynamics are often represented in different formats by different authors. Few systematic comparisons have been made between these representations. Moreover, it is unknown what data representation comes closest to subjective perception. The reported study tests the perceptual validity of existing data representations by comparing their ability to explain the subjective similarity between pairs of performances. In two experiments, 40 participants rated the similarity between performances of a Chopin prelude and a Mozart sonata. Models based on different representations of the tempo and dynamics of the performances were fitted to these similarity ratings. The results favor other data representations of performances than generally used, and imply that comparisons between performances are made perceptually in a different way than often assumed. For example, the best fit was obtained with models based on absolute tempo and absolute tempo times loudness, while conventional models based on normalized variations, or on correlations between tempo profiles and loudness profiles, did not explain the similarity ratings well. [DOI: 10.1121/1.1835504]

PACS numbers: 43.75.St, 43.75.Cd [SEM] Pages: 391–399

I. INTRODUCTION

Measurements of performances, with an emphasis on piano performances, have become an important means of understanding musical expression. For example, measurements have shown the extensive use of subtle variations in tempo, timing, articulation, and dynamics, as well as the consistency of such variations in repeated performances (both already observed by Seashore, 1938), and the controllability of the variations (e.g., Kendall and Carterette, 1990; Palmer 1989, 1996). Comparisons between variations have given some insights into the diversity of interpretations. Repp (1990, 1992a) found, for example, that the diversity tends to be greater among professional musicians than among piano students, and that the diversity tends to be smaller at the phrase level than below the phrase level.

Though measurements provide a detailed account of what is physically happening in a performance, the concern central to this paper is to what extent they reflect the psychological reality of performers and listeners. The validity for performers and listeners of measurable variations has been suggested by several studies. For example, in the work of Sundberg and colleagues (Sundberg, Friberg, and Frydén, 1989; Friberg *et al.*, 1991), the remarks of a professional musician on how to improve a synthesized performance without expressive variations were translated into concrete formulations of rules for the variations of tempo, timing, dynamics, and intonation. In consequent studies, the quality of the synthesized performances with variations was judged to be high compared to the quality of performances without variations (see Thompson *et al.*, 1989).

However, several studies have also shown that the perception of time intervals does not directly correspond to their physical properties. For example, Repp (1992b, 1998) demonstrated a dependency of the perceptual length of a timing perturbation on its position within the phrase structure of music, and Nakajima (1987) found a systematic overestimation of short, empty duration intervals by a constant interval. This means that the overestimation is relatively large for shorter durations.

An additional problem: it is unknown which of the many quantitative representations of tempo and dynamics comes closest to the perceptual representation. The variety of representations includes the representation of timing as either duration or tempo variation (Friberg and Sundberg, 1999); the use of different time scales such as the note level, beat level, bar level, or phrase level (Bengtsson and Gabrielsson, 1983); the use of normalization, which means that the variations are expressed relative to the mean (see, e.g., Gabrielsson, 1987, 1988) instead of in absolute values such as milliseconds or beats per minute (see Repp, 1992a, 1992b; Langner and Goebl, 2003); the use of derivatives of tempo and dynamics rather than absolute values in order to represent the control of tempo change (Kronman and Sundberg, 1987), or the perception of the ''change of change'' in dynamics (Gjerdingen, 1988). Exploration of some of these representations for tempo showed a considerable effect of the time scale and the representation unit on the characteristics of the measured data (Timmers and Honing, 2002).

The main aim of the reported study is to test how well measured data represent perceptually salient characteristics of performances and what data representation comes closest

[a)]Current affiliation: Department of Music, King's College, London, UK. Electronic mail: renee.timmers@kcl.ac.uk

to perception. In comparing representations, it focuses on the size of the time span of the representation (local versus global), the unit of the representation (absolute versus normalized), and the appropriate derivative for the representation (i.e., absolute, first, or second derivative). In addition, the relative salience of tempo and dynamics is tested, as well as the validity of a compound measure that consists of the interaction between tempo and loudness. Tempo times loudness can be seen as a measure of integrated energy (cf. Todd, 1992; Zanon and Widmer, 2003). No comparison is made between the validity of tempo and that of duration. Instead, only tempo is used.

Thus this study addresses the questions of (1) whether the perception of performance deals more with global or local features; (2) whether variations are perceived as changes in absolute values or in relative values (i.e., relative to the average); (3) whether listeners pay attention to absolute levels or to changes therein, or even to changes within the variations; and finally, it addresses (4) whether listeners perceive tempo and loudness as separate dimensions or integrate the two into one compound feature.

The validity of the different data representations is tested by comparing their ability to explain the subjective similarity between pairs of performances. This is done in two experiments that have the same aim and musical material, but differ in experimental procedure. Experiment 2 is a replication of experiment 1 with a stricter experimental procedure. In both experiments, 20 participants listen to pairs of performances of a Chopin prelude and a Mozart sonata and rate the similarity between the performances. These performances are fragments from CD recordings of famous pianists. The tempo and loudness of the performances are measured at the beat level from the audio recordings. The distance in tempo and loudness between pairs of performances is then calculated using the different representations of tempo and loudness. Finally, these distance measures are input to separate multiple regression analyses and to stepwise regression analyses in an attempt to explain the similarity ratings. The degree to which each representation accounts for the variance in the similarity ratings is interpreted as a measure of its ability to capture salient characteristics of the performances.

II. METHOD EXPERIMENT 1

A. Musical material

Five performances of Chopin's Prelude Op. 28, No. 17 are used, as well as six performances of the first movement of Mozart's Sonata KV281. The five performances of the Chopin prelude are by Argerich, Harasiewicz, Kissin, Pollini, and Rubinstein (to be referred to as p1, p2, p3, p4, and p5, respectively)[1] and the six performances of the Mozart sonata are by Barenboim, Batik, Gould, Pires, Schiff, and Uchida (to be referred to as p1, p2, p3, p4, p5, and p6, respectively).[2] The opening bars of the two pieces are used in the experiment (mm. 1–10 for Chopin, and mm. 1–4 for Mozart) as well as six bars from the development section of the Mozart sonata (mm. 22–27 with upbeat). To refer to these fragments, the abbreviations Ch, M1, and M2 are used. These three

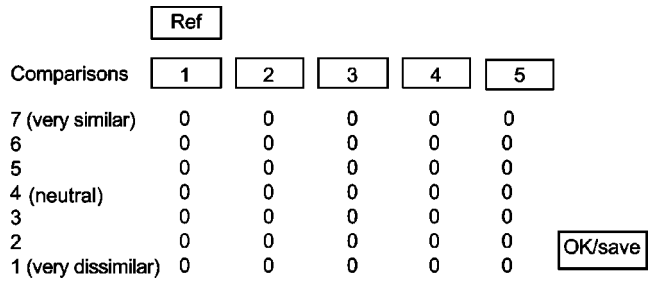


FIG. 1. User interface for the similarity rating in experiment 1.

fragments were chosen because they are expected to differ in the degree to which tempo and dynamics play a role. The importance of these parameters may be especially large in the Chopin prelude, which consists mainly of chords in a repeated eighth-note rhythm. It may be less important in the opening bars of the Mozart sonata, which contains other expressive features such as ornaments and arpeggios. It may again be important in the bars from the development section of the Mozart sonata, which contains leaps in sixteenth notes.

B. Participants

Seven women and 13 men participated in experiment 1. Their age varied from 26 to 45. Fifteen participants were experienced musicians who had had 10 or more years of musical training. Five participants were nonmusicians, who had no more than 3 years of instrumental lessons. Among the musicians were six pianists and nine nonpianists.

C. Procedure

The participants were tested on an individual basis. Half of the participants first listened to the Chopin performances and then to the Mozart performances, while the order was reversed for the other half. The order of the Mozart fragments was always M1 followed by M2.

The presentation of performance pairs was semirandom. To facilitate the similarity judgments, the presentation of performance pairs was grouped into blocks that contained one reference performance and four or five comparison performances. For the Chopin prelude, the participants made four comparisons per block, since the total number of performances was five. The fifth performance was the reference performance. For the Mozart sonata, the participants made five comparisons per block, since the total number of performances was six. Figure 1 shows the organization of one block for the Chopin prelude. In this way, the participant was confronted with all performances basically at once, which provided a stable frame of reference for the similarity ratings. In each subsequent block, a different performance became the reference. The order of these references was randomized over participants as well as the order of the comparison performances within a block.

The participants sat in front of a Macintosh iBook computer and saw the user interface depicted in Fig. 1 on the screen. The interface contained play buttons for the reference performance and the comparison performances. They listened alternatively to the reference performance and a comparison performance via headphones, and rated the similarity

between the two on a scale from 1 to 7 by pressing one of the radio buttons. One meant very dissimilar, while 7 meant very similar. They could listen to each performance as often as they wished and could correct the ratings until they pressed the ok/save button. This would bring up the following block of performances, which consisted of the same performances in a different order of comparison and with a different reference performance. The session ended when all performances had been the reference performance once. This resulted in 20 comparisons for the Chopin fragment and 30 comparisons for the Mozart fragments. Each of the 10 performance pairs of the Chopin fragment and 15 performance pairs of the Mozart fragments were rated twice: once with one of the performances as reference and the once with the other performance as reference.

After the experiment, the participants filled out a questionnaire about their rating strategy. They were asked to describe on what bases they made the similarity judgments and to what aspects of the performances they paid most attention. This last question was answered by giving a rating to a list of parameters on a scale from 0–3. Zero meant no attention, while 3 meant most attention. The parameters are listed in Table III. The total duration of the experiment was around 1 h.

D. Apparatus

The experimental data were collected using POCO (Honing, 1990), running on an Apple iBook under Macintosh OS 9.2. A special POCO module was designed containing a user interface (see Fig. 1), playback of audio files, and recording the responses into a log file. The sound files used in the judgments were CD-quality stereo audio files (sampled at 44.1 kHz). Sony's dynamic stereo professional headphones MDR-7506 were used.

E. Similarity predictions

Central to this study is the prediction of the subjective judgment of similarity between performances on the basis of different representations of their tempo and dynamics. The measurement of tempo and loudness from the audio recordings was performed using algorithms developed by members of the Music and AI group at the Austrian Research Institute for Artificial Intelligence. A beat-tracking algorithm was used to locate the beat within the audio file (see Dixon, 2001). The output of the beat-tracking procedure was hand-corrected using an interface especially designed for this purpose (Dixon and Goebl, 2001). The location of the beat was defined to coincide with the onset of the corresponding melody note. Local tempo in beats per minute was calculated for each interbeat interval. To get a measure of local loudness, the localized beats were used as well. The maximum amplitude level was selected around each beat, which spans from the halfway point of the previous interbeat interval to the halfway point of the following interbeat interval (see also Langner and Goebl, 2003). This level in dB was recalculated into sones, which is an approximation of the subjective perception of the loudness of tones (see Pampalk *et al.*, 2002).

As mentioned in the Introduction, the different representations of tempo and dynamics to be compared vary in time scale (local or global), unit (absolute or normalized), and derivation (absolute or derivative). To predict the subjective distance between two performances, the difference in tempo and loudness between two performances was calculated using different methods for different representations. Equations (1)–(9) give the calculation method for each representation. For brevity, the calculations are only presented for tempo. Similar calculations were applied to the loudness of each beat. Capital T is a vector that consists of a tempo indication for each interbeat interval ($T=[t_1,t_2,\ldots,t_n]$, in which t stands for local tempo at interbeat interval n).

In the equations below, the subscript of T refers to one of the two performances of a performance pair that are compared. The horizontal line above an expression indicates averaging. The vertical lines at both sides of an expression indicate that the absolute value is taken. The superscripts in change, and change of change, indicate the first and second derivative, respectively.

More specifically, the first distinction is between a global and a local representation. The global representation of tempo calculates the average tempo of a performance. The distance in tempo between two performances is then calculated by taking the absolute difference in average tempo of performances 1 and 2; see Eq. (1)

$$\text{Difference in global tempo} \quad |\overline{T_1}-\overline{T_2}|. \tag{1}$$

The distance in local tempo between two performances is calculated by calculating the absolute difference in local tempo for each beat of the two performances and taking the average of these absolute differences; see Eq. (2)

$$\text{Difference in local tempo} \quad |\overline{T_1-T_2}|. \tag{2}$$

The second distinction is between absolute tempo and loudness [such as in Eqs. (1) and (2)], and relative tempo and loudness. Relative tempo (or loudness) is calculated by dividing the absolute value by the average tempo (or loudness) of the performance. The difference in relative tempo variation is given by Eq. (3)

$$\text{Difference in relative tempo variation} \quad \left|\overline{\frac{T_1}{\overline{T_1}}-\frac{T_2}{\overline{T_2}}}\right|. \tag{3}$$

The third distinction is between absolute tempo and loudness and the changes within the tempo and loudness values. The derivative captures the change in tempo (or loudness) over time. In other words, it measures the amount of acceleration and deceleration, and the amount of crescendo and decrescendo. The second derivative captures the change within the change in tempo (or loudness) over time. This means that it measures changes in the direction and amount of acceleration/deceleration, and crescendo/decrescendo. The distance between two performances based on differences in tempo change and the change of tempo change is given by Eqs. (4) and (5), respectively

$$\text{Difference in tempo change} \quad |\overline{T_1'-T_2'}|, \tag{4}$$

Difference in change of tempo change $|\overline{T_1'' - T_2''}|$. (5)

The derivative is defined as the difference in local tempo (or loudness) between successive beats, so T' is $[t_2 - t_1, t_3 - t_2, \cdots, t_n - t_{n-1}]$, in which t stands for local tempo at the nth interbeat interval. In this way, the calculation is not strictly a derivative over time, but an event-based calculation of change.

In the fourth distinction, tempo and dynamics are either treated as separate variables as above in Eqs. (1)–(5), or treated as a compound variable as in Eq. (6) and Eq. (7). The compound variable integrates tempo and loudness per beat through multiplication. The difference in this measure between two performances is calculated on a global level using Eq. (6) and a local level using Eq. (7)

Difference in global tempo times loudness

$$|\overline{T_1^* L_1} - \overline{T_2^* L_2}|, \tag{6}$$

Difference in local tempo times loudness

$$|\overline{T_1^* L_1 - T_2^* L_2}|. \tag{7}$$

Finally, two measures that are regularly used in performance research are added to be more complete. The first is the standard deviation of the variation (see, e.g., Timmers *et al.*, 2000; Timmers, 2003; and Zanon and Widmer, 2003). The distance between two performances in tempo and dynamics is with this measure assumed to be the absolute difference between the standard deviation of local tempo or loudness of performance 1 and 2; see Eq. (8)

Difference in amount of tempo variation

$$|\mathrm{std}(T_1) - \mathrm{std}(T_2)|. \tag{8}$$

The final measure calculates the correlation between the local tempo (and loudness) profile of the two performances; see Eq. (9). This is the most frequently used method to assess the similarity between the timing and dynamics profiles of different performances (Clarke, 1993; Repp, 1994, 2000; Timmers *et al.*, 2000)

Correlation between tempo profiles $\mathrm{corr}(T_1, T_2)$. (9)

To compare the predictive power of the different representations, two methods are used. The first approach compares the predictive power of the nine measures by running separate multiple regression analyses for each measure. All of these models consist of one tempo and one loudness component, except for the models based on the compound measures as in Eq. (5) and Eq. (6), which have only one component. The regression models with two components have the format shown in Eq. (10). The regression model consists of two components (the difference measures D_t and D_l), an intercept (a) and weights (b and c). The regression models for the compound measures have only one component (D_{t*l}), an intercept (a), and one weight (b)

$$s = a + b*D_t + c*D_l. \tag{10}$$

The second approach takes all difference measures based on tempo, loudness, and the interaction between them as input of a stepwise regression analysis. The stepwise regression analysis adds components to the analysis in order of explained variance. It adds components as long as their addition to the explained variance is significant. An additional restriction is that only components are included for which the effect is in the predicted direction: an increase in difference (and a decrease in correlation) should lead to a decrease in similarity rating. In this way, the components are sorted in order of explained variance. These analyses were done using JMP 4.0.

III. RESULTS EXPERIMENT 1

The presentation of the results is divided into three parts: First, the results of the similarity ratings are presented, followed by the results of the prediction of the similarity ratings by the different models and the stepwise regression analysis. Third, the results of the questionnaire are presented and related to the results of the similarity rating study.

A. Similarity ratings

To test if the effect of performance pair is systematic over participants and presentations, a repeated measures ANOVA was run in SPSS10 with pair and order as independent within-subject variables and the similarity rating as dependent variable. A separate ANOVA was run for each fragment.

For M1 and Ch, the main effect of pair is the only significant effect. This is also the case when the analysis is corrected for violations of sphericity using the Greenhouse–Geisser correction [$F(9,11) = 25.7$, $p < 0.001$ for Ch; $F(14,6) = 19.4$, $p < 0.001$ for M1].[3] For M2, however, all effects are significant using the same correction for violations of sphericity. The effect of pair is the strongest effect, followed by the main effect of order [$F(14,6) = 21.7$, $p < 0.001$ for the main effect of pair, $F(1,19) = 10.8$, $p = 0.004$ for the main effect of order, and $F(14,6) = 3.2$, $p = 0.004$ for the interaction effect]. The main effect of order is hard to explain. Like the interaction between pair and order, it might suggest that for some similarity ratings of M2 the similarity rating depended on which of the performances was the reference performance. The blocking of stimuli may have caused this context effect. In experiment 2, this issue is resolved by presenting all performance pairs sequentially.

Although the interaction between pair and order was significant for M2, the size of the effect was rather small. In fact, the average difference in ratings of a pair in the two orders remained under 1.15 points (pair 4–5 of Ch). To get a robust similarity rating that is independent of context, the following analyses use the average of the ratings of pairs in the two orders. It therefore has 10 data points per participant for Ch, and 15 data points per participant for M1 and M2.

B. Prediction of similarity ratings

The first comparison between the explanatory power of the different representations is done by fitting the nine measures to the similarity ratings using separate multiple regres-

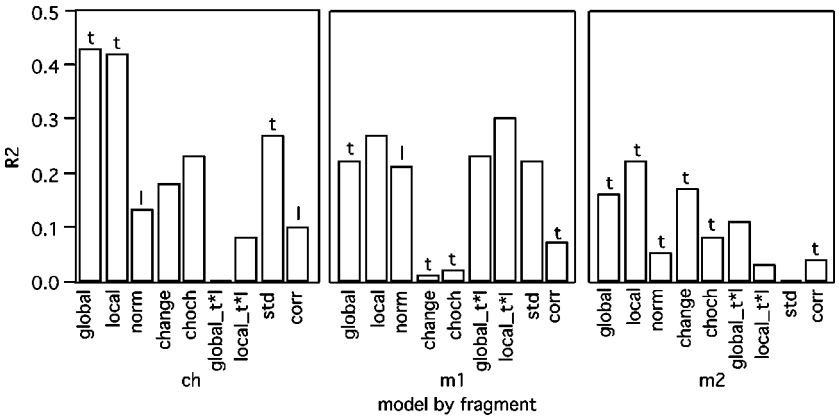


FIG. 2. Explained variance (R^2) of the fits of each model to the similarity ratings of all participants, separated per fragment for experiment 1. The models included are based on global, local, and normalized (norm) tempo and loudness; change and change of change (choch) of tempo and loudness; global and local tempo times loudness (t^*l), the standard deviation of (std) and the correlation (corr) between tempo and loudness profiles. Letters indicate the component that accounted for 2/3 or more of the explained variance (t for tempo and l for loudness).

sion models for each measure. All of these models consist of one tempo and one loudness component, except for the models based on the compound measures that have only one component. Figure 2 shows the explained variance for each of the regression models. The letter on top of each line indicates which parameter (t for tempo and l for loudness) contributes most to the explanation, which means that it explains at least 2/3 of the total explained variance. If no letter is indicated, both tempo and loudness contribute roughly equally to the explanation. Global and local tempo times loudness do not have a letter, because they always include both variables.

Figure 2 shows that the explained variance is higher for the Chopin fragment than for the Mozart fragments, and that tempo is the more important factor for the explanation. It further shows that (1) the regression analysis based on local tempo and loudness does better than based on global tempo and loudness for the Mozart fragments. (2) To take relative values instead of absolute values does not improve the explained variance. This is suggested by the low explained variance of the regression analysis that uses normalization (abbreviation norm). (3) To take the derivative of tempo and loudness is not an improvement, given the low explained variance for models based on change and change of change (abbreviation choch). (4) To treat tempo and loudness as separate values is better than to treat them as one compound variable for Ch and M2, but not for M1. (5) The regression analysis based on the standard deviation of tempo and loudness does quite well for Ch and M1, but is not best. (6) Correlation does not explain the similarity ratings well.

The second comparison between the relative strength of the different representations takes all separate difference measures based on tempo and loudness as potential input of a stepwise regression model. This includes the measures given by Eqs. (1)–(9) and the parallel measures for loudness. The components are added stepwise in order of explained variance as long as the addition in explained variance is significant ($p<0.05$) and the direction of the effect of the component is as predicted, which means that an increase in the measured difference between two performances leads to a decrease in the similarity rating.

By using stepwise regression, the components are sorted in order of explained variance. The benefit of this method is that there is no overlap in explained variance, and the focus is only on those variables that are most responsible for the explanation of variance. This is in contrast to the previous presentation of the results, in which part of the explained variance by one model might be due to the correlation with another model.

Tables I and II show the results of the stepwise regression analyses for each fragment for the musicians and nonmusicians separately. The results for the musicians and nonmusicians are highly similar: Local or global tempo is the strongest component for Ch and M2, and global or local tempo times loudness is second, although not for the musicians' ratings of Ch. The reverse is true for M1; the main component is local tempo times loudness, while the second component for the musicians is local tempo. The components are more often local than global. The explained variance is larger for the Chopin fragment than for the Mozart fragments. It is larger for the nonmusicians than for the musicians.

Note that some of the models that did quite well in the separate multiple regression analyses do not occur in the stepwise regression analyses; they did not account for vari-

TABLE I. Results of the stepwise regression analysis for musicians ($N=15$). Parameters in order of entrance of the stepwise regression model; the total explained variances for each step, and the F and p value for the full model.

Fragment	Parameters	R^2	F value	p value
Ch	Global t	0.45	$F(1,148)=120$	<0.0001
M1	Local t^*l	0.27	$F(2,222)=46.3$	<0.0001
	Local t	0.29		
M2	Local t	0.23	$F(3,221)=43.4$	<0.0001
	Global t^*l	0.35		
	Change l	0.37		

TABLE II. Results of the stepwise regression analysis for nonmusicians ($N=5$). Parameters in order of entrance of the stepwise regression model; the total explained variances for each step, and the F and p value for the full model.

Fragment	Parameters	R^2	F value	p value
Ch	Local t	0.52	$F(2,47)=30.0$	<0.0001
	Global t^*l	0.56		
M1	Local t^*l	0.40	$F(1,73)=47.8$	<0.0001
M2	Local t	0.22	$F(2,72)=17.3$	<0.0001
	Global t^*l	0.32		

TABLE III. Sum of attention ratings for the Chopin and Mozart fragments expressed as percentage of the maximal sum of the ratings.

Parameter	Chopin (% of max)	Mozart (% of max)
Tempo	71	78
Loudness	36	42
Rubato	73	73
Dynamics	62	64
Articulation	56	78
Pedal	40	33
Phrasing	78	71
Interpretation	76	76
Character	76	78
Emotion	49	44

ance in addition to the variance accounted for by the strongest components. Note as well that change in loudness makes a small contribution to the explained variance of the musicians' ratings of M2.

C. Interviews

After the similarity rating experiment, the participants were asked to indicate to what aspects of the performances they had paid most attention. This was first done by free choice and secondly by giving an attention rating to ten variables (see Table III).

The aspects that were mentioned and the number of participants mentioning them in response to the free-choice question are tempo (11), articulation (7), character and style/overall impression (7), interpretation (6), rubato (5), dynamics (4), the quality of the pianist/the smoothness of playing (4), loudness (2), arpeggios, ornaments (2), the sound of the recording (1), phrasing (1), and perception of movement (1). The attention ratings show a similar pattern (Table III): most attention is paid to global tempo and rubato, less attention is paid to dynamics, and little to overall loudness. Articulation is an important factor for the Mozart fragments, but less so for the Chopin fragment, which may have been due to the larger use of pedal in the Chopin fragment. The interpretation of the music is important as well as the character and style of the performance.

The importance of tempo and tempo variation and the lesser importance of dynamic variation and overall loudness agree with the results of the experiment. The importance of articulation for the Mozart fragments may account for the lower explained variance for the Mozart fragments than for the Chopin fragment. The tendency of several participants to listen to the overall impression of a performance may indicate that they did not listen very analytically, which possibly may relate to the large contribution of the tempo times loudness measures. Differences in interpretation of the music such as phrasing may have been accounted for indirectly, though probably taking the relationship with musical structure into account would have improved the variance explained.

IV. EXPERIMENT 2

A second experiment was run with the same purpose as the first experiment. The only difference with the first experiment is the experimental procedure, which was changed to be in accordance with the general procedure for similarity rating studies. Instead of blocking stimuli into groups of comparison performances with a reference performance, the performance pairs were presented one after another and the rating was done for each pair sequentially, to avoid possible dependencies between the ratings of different pairs. The results of experiment 2 are not expected to be different from the results of experiment 1. Instead, experiment 2 is a replication of experiment 1 with a stricter experimental procedure.

V. METHOD OF EXPERIMENT 2

A. Musical material

The musical material was the same as in experiment 1, with the exception that pairs of performances were in experiment 2 combined into one audio file with a 2200-ms interval between the end of the first performance and the start of the second performance. A practice trial was added that used four performances of the second movement of Mozart KV332 piano sonata by Argerich, Gould, Pires and Schiff. These performances were taken from the same CDs as used in experiment 1.

B. Participants

Eleven women and nine men participated in experiment 2. Their age varied between 21 and 48. Fifteen participants were experienced musicians who had had 10 or more years of musical training. Five participants were nonmusicians, who had had no more than 3 years of instrumental lessons. Among the musicians were nine pianists and six nonpianists.

C. Procedure

The participants were tested on an individual basis. They sat in front of a Macintosh iBook computer and saw a simplified interface on the screen. The interface had one play button, one set of seven vertically aligned radio buttons to make the similarity rating, and an ok/save button. The participants read the instructions from paper. The instructions described the task of the participants and the procedure of the experiment. They were asked to listen to a pair of performances that would sound by pressing the play button, and to indicate the similarity between the performances on a scale from 1–7. One means that the performances are very different, while 7 means that they are very similar. The values in between could be used for fine tuning. After the rating of similarity, they pressed the ok/save button to continue with the next pair of performances.

The entire experiment consisted of four blocks: a block to practice and three blocks to rate the similarity between pairs of performances for each of the three fragments. Each participant rated all pairs of performances of a fragment once. The order of presentation of fragments and performance pair was randomized over subjects. The order of the

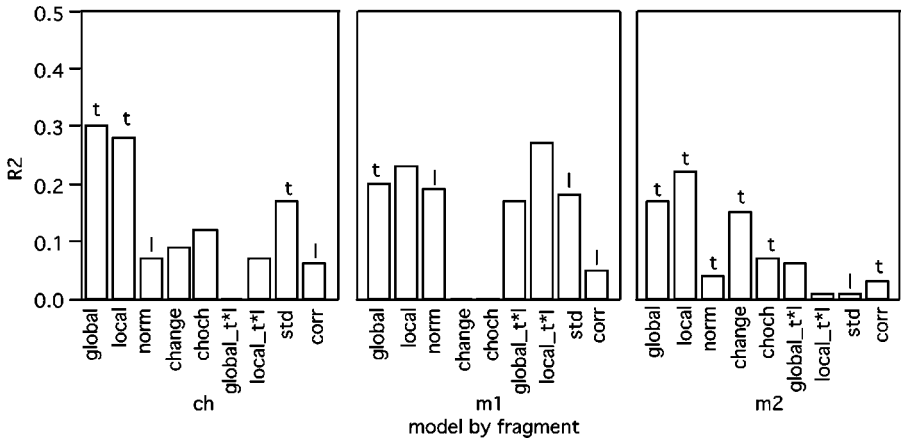


FIG. 3. Explained variance (R^2) of the fits of each model to the similarity ratings of all participants, separated per fragment for experiment 2. The models included are based on global, local, and normalized (norm) tempo and loudness; change and change of change (choch) of tempo and loudness; global and local tempo times loudness (t^*l), the standard deviation of (std) and the correlation (corr) between tempo and loudness profiles. Letters indicate the component that accounted for 2/3 or more of the explained variance (t for tempo and l for loudness).

performances within a pair was counterbalanced over participants. The total duration of the experiment was around half an hour.

D. Apparatus

The same apparatus was used as in experiment 1 with the exception of the interface, which was simplified. The interface included only one play button and one rating scale with seven radio buttons.

E. Similarity predictions

The same analyses were done with the same predictions for the similarity ratings as in experiment 1.

VI. RESULTS EXPERIMENT 2

A. Similarity ratings

To test if the effect of performance pair was systematic over participants and order, a repeated measures ANOVA was run in SPSS10 with pair as within-subject variable, order as between-subjects variable, and the similarity rating as dependent variable. A separate ANOVA was run for each fragment.

For all three fragments, the main effect of pair was the only significant effect. This was also the case when the analysis was corrected for violations of sphericity using the Greenhouse–Geisser correction [$F(9,11)=13.5$, $p<0.001$ for Ch; $F(14,6)=11.1$, $p<0.001$ for M1; and $F(14,6)=11.3$, $p<0.001$ for M2].

B. Prediction of similarity ratings

As for experiment 1, the first comparison between the explanatory power of the different representations was made by fitting the nine measures to the similarity ratings using separate multiple regression models for each measure. Figure 3 shows the explained variance for each of the regression models.

Although the explained variance is a bit lower than in experiment 1, the results are highly similar overall. The models that do well and the ones that do not do well are the same as in experiment 1. Indeed, the correlation between the explained variances of the models as found in the two experiments is above 0.97 for each of the three fragments.

The results of the stepwise regression analyses show a similar high agreement with the results of experiment 1, especially for the musicians (see Tables IV and V). Again, local or global tempo, and local or global tempo times loudness are the two strongest components. In addition, the change of change or the correlation in loudness explains a small part of the variance for M2.

Experiment 2 generally replicates the results of experiment 1, with the exception that the explained variance was a bit higher in experiment 1 than in experiment 2. This difference can be attributed to the procedure of experiment 1, which provided the participants with a frame of reference for the similarity ratings by presenting all performances within each rating block. Besides this effect in quantity, there was no qualitative effect of the difference in procedure or difference in participant pool, which strengthens the generality of the results.

TABLE IV. Results of the stepwise regression analysis for musicians ($N=15$) for experiment 2. Parameters in order of entrance of the stepwise regression model; the total explained variances for each step, and the F and p value for the full model.

Fragment	Parameters	R^2	F value	p value
Ch	Global t	0.42	$F(1,148)=106$	<0.0001
M1	Local t^*l	0.33	$F(2,222)=58.2$	<0.0001
	Local t	0.34		
M2	Local t	0.25	$F(3,221)=33.4$	<0.0001
	Global t^*l	0.30		
	Change l	0.32		

TABLE V. Results of the stepwise regression analysis for nonmusicians ($N=5$) for experiment 2. Parameters in order of entrance of the stepwise regression model; the total explained variances for each step, and the F and p value for the full model.

Fragment	Parameters	R^2	F value	p value
Ch	Local t^*l	0.11	$F(2,47)=4.8$	<0.02
	Global t	0.17		
M1	Global t^*l	0.20	$F(1,73)=18.4$	<0.0001
M2	Local t	0.14	$F(3,71)=11.5$	<0.0001
	Global t^*l	0.26		
	Corr l	0.33		

VII. GENERAL DISCUSSION

As outlined in the Introduction, the reported study addressed the questions of (1) whether the perception of performance deals more with the global or the local features; (2) whether variations are perceived as changes in absolute values or in relative values (i.e., relative to the average); (3) whether listeners pay attention to absolute levels or to changes therein, or even to changes within the variations; and (4) whether listeners perceive tempo and loudness as separate dimensions or integrate the two into one compound feature. The results of the two experiments show local models to be slightly stronger than global models, and models based on absolute values to be stronger than models based on normalized values. Models based on absolute values were also stronger than those based on derivatives. This strength of absolute representations concerned absolute tempo as well as the compound feature of tempo times loudness. Loudness was only sporadically significant as a separate feature, and when it was significant, it was in other representations than absolute loudness.

In addition, this study showed the limited ability of correlation to capture the similarity in tempo and dynamics of two performances as well as the medium strength of models based on the standard deviation of tempo and dynamics to do so.

The strength of the parameters changed with fragment in a similar way for the musicians and the nonmusicians and in the same way in the two experiments. This strongly suggests bottom-up processes driven by the specific characteristics of the musical stimuli.

Parts of these results are confirmations of previous literature, while other parts were less expected. For example, Repp (2000) also mentioned the limited ability of correlations to capture differences between patterns related to the means and the standard deviations of the variations. The current study stresses the importance of such aspects not covered by correlation as the extent of the variation and the absolute value of the measure. It demonstrated the salience of global tempo for the evaluation of performances, which is in line with studies on the emotion of music for which tempo is an important factor (e.g., Hevner, 1937) and studies on the reproduction of the absolute tempo of memorized music (Levitin and Cook, 1996). Nevertheless, the large role of local tempo in the similarity predictions opens an unexplored area of investigation. It implies that even in a comparison between performances, the absolute tempo of a time unit of one performer is compared to that of the other performer rather than the interpretation of the passage in terms of acceleration or deceleration. Similarly, the representation of performance variables as an integration of tempo and loudness has hardly been explored, although it is prominent in Todd's theory of expression (e.g., 1992). It seems important for future research to further investigate the relevance of these representations.

This study is not conclusive about the absolute strength of the representations described here. It only provides a relative ranking of the different measures. Probably the extraction of salient performance characteristics can still be improved upon and a more complete model might be defined to explain the distance between a pair of performances. The interview held with the participants highlighted some of the aspects that were missed by the models. For example, the models did not include articulation, timbre, fluency, or quality of the performance, and did not take the relation with the musical structure into account. In addition, it might have been that participants focused their attention on specific parts of the music rather than the entire fragment and that primacy and recency effects played a role.

Nevertheless, the measured differences in tempo and loudness were quite well able to predict the subjective distance between performances, and seem therefore reliable to represent a considerable part of the performance characteristics. The parameters most responsible for this explanation for both musicians and nonmusicians were local and global tempo, and local and global tempo times loudness. Local tempo and the interaction between tempo and loudness are not often used in performance research, and a shift in attention towards these representations of performances seems important for future research, also when comparing between different interpretations of music.

ACKNOWLEDGMENTS

This study was realized with financial support of the Mosart IHP-Network, HPRN-CT-2000-00115, the START program of the Austrian Federal Ministry for Education, Science and Culture (Grant No. Y99-INF), and a Talent stipendium of the Dutch Scientific Organization (NWO). In addition, the Austrian Research Institute for Artificial Intelligence acknowledges basic financial support by the Austrian Federal Ministry for Education, Science and Culture. I would like to thank Ric Ashley, Simon Dixon, Werner Goebl, Josef Linschinger, Elias Pampalk, Asmir Tobudic, and Gerhard Widmer for providing the performance data and the perfect environment to do this study and Henkjan Honing for his help with the interface and POCO. I also thank them and the reviewers for their helpful comments.

[1]The recordings of the Chopin Prelude were taken from the following CDs: Argerich, Philips Classics, 456 703-2 (Great Pianists), recording 10/1975, München Herkules-Saal (orig. Deutsche Grammophon); Harasiewicz, Philips Classics 442 268-2, recorded 1963; Kissin, BMG, 09026 63535 2, recorded 1999; Pollini, DGG 413 796-2, recorded 1975; Rubinstein, BMG, GD 60822, recorded 1946.

[2]The recordings of the Mozart Sonata were taken from the following CDs: Daniel Barenboim, Emi Classics, CDZ 7 67295 2, recorded 1984; Roland Batik, Gramola, 98701-705, recorded 1990; Glenn Gould, Sony Classical SM4K 52627, recorded 1967; Maria João Pires, DGG, 431 761-2, recorded 1991; András Schiff, ADD (Decca), 443 720-2, recorded 1980; Mitsuko Uchida, Philips Classics, 464 856-2, recorded 1987.

[3]The Greenhouse–Geisser correction for violations of sphericity was applied whenever appropriate. If the correction is applied, the uncorrected degrees of freedom are reported. The reported probability value is the probability following correction.

Bengtsson, I., and Gabrielsson, A. (**1983**). "Analysis and synthesis of musical rhythm," in *Studies of Music Performance*, edited by J. Sundberg (Royal Swedish Academy of Music, Stockholm), pp. 76–181.

Clarke, E. F. (**1993**). "Imitating and evaluating real and transformed musical performances," Music Percept. **10**, 317–341.

Dixon, S. (**2001**). "Automatic extraction of tempo and beat from expressive performances," J. New Music Res. **30**, 39–58.

Dixon, S., and Goebl, W. (**2001**). ''Pinpointing the Beat: Tapping to Expressive Performances,'' in *Proceedings of the 7th International Conference on Music Perception and Cognition* (ICMPC7, Sydney, Australia), pp. 617–620.

Friberg, A., and Sundberg, J. (**1999**). ''Does music performance allude to locomotion? A model of final *ritardandi* derived from measurements of stopping runners,'' J. Acoust. Soc. Am. **105**, 1469–1484.

Friberg, A., Frydén, L., Bodin, L. G., and Sundberg, J. (**1991**). ''Performance rules for computer-controlled contemporary keyboard music,'' Comput. Music J. **15**, 49–55.

Gabrielsson, A. (**1987**). ''Once again: The theme from Mozart's Piano Sonata in A major: A comparison of five performances,'' in *Action and Perception in Rhythm and Music*, edited by A. Gabrielsson (Royal Swedish Academy of Music, Stockholm), pp. 81–103.

Gabrielsson, A. (**1988**). ''Timing in music performance and its relation to music experience,'' in *Generative Processes in Music. The Psychology of Performance, Improvisation, and Composition*, edited by J. Sloboda (Clarendon, Oxford), pp. 27–51.

Gjerdingen, R. O. (**1988**). ''Shape and motion in the microstructure of song,'' Music Percept. **6**, 35–64.

Hevner, K. (**1937**). ''The affective value of pitch and tempo in music,'' Am. J. Psychol. **49**, 621–630.

Honing, H. (**1990**). ''POCO, An Environment for Analysing, Modifying and Generating Expression in Music,'' in *Proceedings of the 1990 International Computer Music Association* (CMA, San Francisco), pp. 364–368.

Kendall, R. A., and Carterette, E. C. (**1990**). ''The communication of musical expression,'' Music Percept. **8**, 129–164.

Kronman, U., and Sundberg, J. (**1987**). ''Is the musical ritard an allusion to physical motion?'' in *Action and Perception in Rhythm and Music*, edited by A. Gabrielsson (Royal Swedish Academy of Music, Stockholm).

Langner, J., and Goebl, W. (**2003**). ''Visualizing expressive performance in tempo-loudness space,'' Comput. Music J. **27**, 69–83.

Levitin, D. J., and Cook, P. R. (**1996**). ''Memory for musical tempo: Additional evidence that auditory memory is absolute,'' Percept. Psychophys. **58**, 927–935.

Nakajima, Y. (**1987**). ''A model of empty duration perception,'' Percept. **16**, 485–520.

Palmer, C. (**1989**). ''Mapping musical thought to musical performance,'' J. Exp. Psychol. Hum. Percept. Perform. **15**, 331–346.

Palmer, C. (**1996**). ''On the assignment of structure in music performance,'' Music Percept. **14**, 23–56.

Pampalk, E., Rauber, A., and Merkl, D. (**2002**). ''Content-based organization and visualization of music archives,'' in *Proceedings of the 10th ACM International Conference on Multimedia* (MM'02, Juan-les-Pins, France), pp. 570–579.

Repp, B. H. (**1990**). ''Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists,'' J. Acoust. Soc. Am. **88**, 622–641.

Repp, B. H. (**1992a**). ''Diversity and commonality in music performance—An analysis of timing microstructure in Schumann's Traumerei,'' J. Acoust. Soc. Am. **92**, 2546–2568.

Repp, B. H. (**1992b**). ''Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations,'' Cognition **44**, 241–281.

Repp, B. H. (**1994**). ''Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study,'' Psychol. Res. **56**, 269–284.

Repp, B. H. (**1998**). ''Obligatory 'expectations' of expressive timing induced by perception of musical structure,'' Psychol. Res. **61**, 33–43.

Repp, B. H. (**2000**). ''Pattern typicality and dimensional interactions in pianists' imitation of expressive timing and dynamics,'' Music Percept. **18**, 173–211.

Seashore, C. E. (**1938**). *Psychology of Music* (Dover, New York).

Sundberg, J., Friberg, A., and Frydén, L. (**1989**). ''Rules for automated performances of ensemble music,'' Contemp. Music Rev. **3**, 89–109.

Thompson, W. F., Sundberg, J., Friberg, A., and Frydén, L. (**1989**). ''The use of rules for expression in the performance of melodies,'' Psychol. Music **17**, 63–82.

Timmers, R. (**2003**). ''On the contextual appropriateness of expression,'' Music Percept. **20**, 225–240.

Timmers, R., and Honing, H. (**2002**). ''On music performance, theories, measurement and diversity,'' Cognitive Processing (International Quarterly of Cognitive Sciences) **1–2**, 1–19.

Timmers, R., Ashley, R., Desain, P., and Heijink, H. (**2000**). ''The influence of musical context on tempo rubato,'' J. New Music Res. **29**, 131–158.

Todd, N. P. (**1992**). ''The dynamics of dynamics: A model of musical expression,'' J. Acoust. Soc. Am. **91**, 3540–3550.

Zanon, P., and Widmer, G. (**2003**). ''Learning to recognize famous pianists with machine learning techniques,'' in *Proceedings of the Third Stockholm Music Acoustics Conference* (SMAC03, Stockholm, Sweden).

Primarily nonlinear effects observed in a driven asymmetrical vibrating wire[a)]

Roger J. Hanson,[b)] H. Kent Macomber,[c)] Andrew C. Morrison,[d)]
and Matthew A. Boucher[e)]
Department of Physics, University of Northern Iowa, Cedar Falls, Iowa 50614-0150

(Received 4 December 2003; revised 5 October 2004; accepted 10 October 2004)

The purpose of the work reported here is to further experimentally explore the wide variety of behaviors exhibited by driven vibrating wires, primarily in the nonlinear regime. When the wire is driven near a resonant frequency, it is found that most such behaviors are significantly affected by the splitting of the resonant frequency and by the existence of a ''characteristic'' axis associated with each split frequency. It is shown that frequency splitting decreases with increasing wire tension and can be altered by twisting. Two methods are described for determining the orientation of characteristic axes. Evidence is provided, with a possible explanation, that each axis has the same orientation everywhere along the wire. Frequency response data exhibiting nonlinear generation of transverse motion perpendicular to the driving direction, hysteresis, *linear* generation of perpendicular motion (sometimes tubular), and generation of motion at harmonics of the driving frequency are exhibited and discussed. Also reported under seemingly unchanging conditions are abrupt large changes in the harmonic content of the motion that sometimes involve large subharmonics and harmonics thereof. Slow transitions from one stable state of vibration to another and quasiperiodic motions are also exhibited. Possible musical significance is discussed. [DOI: 10.1121/1.1828511]

PACS numbers: 43.75.Gh, 43.40.Cw, 43.75.De, 43.75.Mn [NHF] Pages: 400–412

I. INTRODUCTION

Though a sinusoidally driven harpsichord wire could hardly be considered a musical instrument, a more complete understanding of a vibrating wire, especially its nonlinear behaviors and asymmetrical properties, is potentially significant for stringed musical instruments. Further knowledge and understanding can, for example, shed light on what musicians sometimes refer to as a ''false'' string. The effect of nonlinearity and asymmetry on the behavior of *plucked* strings, though not a specific subject of this paper, is a related topic and of definite musical significance.

A basic theoretical framework for the vibrations of a wire is provided in the tutorial accounts of Valette.[1,2] A more rigorous and comprehensive framework is provided by Watzky[3,4] and includes other treatments as special cases. If the motion of the wire is sufficiently small, its vibrations are quite well described by the classical linear theory of the vibrating string, modified if necessary to include stiffness. The theory shows that the wire is capable of certain simple free vibrations or ''normal modes,'' each with its own characteristic frequency and wire shape. Driving vibrations at a frequency in the vicinity of a normal-mode frequency or ''resonance'' produces the classic response curve of a driven harmonic oscillator.

When, however, the vibrations of the wire are sufficiently large, entirely new and complex behaviors are observed. A prominent example is the ''bending over'' of the frequency response curve in the direction of higher frequencies until the resonance peak is replaced by a discontinuous drop to a much lower level of vibration as the driving frequency is increased, with hysteresis then exhibited as the driving frequency is decreased. Generation of motion perpendicular to the driving direction is another example, though, as seen later in Sec. V B, such motion is also possible in the linear regime. A third example is the generation of vibrations at harmonics of the driving frequency. An especially interesting fourth example is the slow and, in some cases, periodic change in a complicated pattern of motion. A final example, though not observed in the work reported here, is chaotic motion. (For analysis and observations of such motion, see Molteno and Tufillaro.[5])

For the explanation of these behaviors, one looks naturally to violations of the assumptions made in developing the linear theory. Since the material of the wire typically operates in a linear elastic regime, the nonlinear behavior arises primarily from nonlinear relations that express the geometry of the wire in terms of its displacement. Important consequences of these are variations in the tension in the wire and coupling between motions of the wire in the longitudinal and two transverse directions. Through a careful development from first principles, Valette derives a set of approximate nonlinear equations[6] governing motion of the wire in the three directions. He then exploits the fact that usually in

[a)]Portions of this work were presented in ''Unusual motions of a nonlinear asymmetrical vibrating string,'' Proceedings of the Stockholm Music Acoustics Conference (SMAC 03), Stockholm, Sweden, August 2003, and ''Nonlinear effects in an asymmetrical vibrating string,'' Proceedings of The International Symposium on Musical Acoustics (ISMA 01), Perugia, Italy, September 2001.

[b)]Electronic mail: roger.hanson@cfu.net
[c)]Electronic mail: macomber@cfu.net
[d)]Now at Department of Physics, Northern Illinois University, DeKalb, Illinois 60115; electronic mail: morrison@physics.niu.edu
[e)]Now at School of Music, Indiana University, Bloomington, Indiana 47405.

musical strings and certainly in the work reported here, longitudinal waves travel much faster than transverse ones. This allows him to express the longitudinal motion approximately in terms of the transverse motions and reduce the governing equations to two approximate coupled nonlinear partial differential equations for the transverse motions.

Though the nonlinearity of these equations is, through approximation, of the lowest possible order, they are analytically intractable. Numerical computation of solutions is, of course, a possibility but suffers from serious limitations. First, numerical computation produces only highly specific solutions. But experience with much simpler nonlinear differential equations shows that they often have a very complex variety of qualitatively different solutions, including hysteretic and chaotic ones. A set of numerically computed solutions may thus largely fail to convey the richness of the motions embraced by the governing equations. (Of course, this is also true of a set of experimental investigations.) Furthermore, very delicate computational problems may well arise that if not adequately dealt with will severely compromise the solutions. Beyond these difficulties, there is the question of the adequacy of the governing equations themselves. Equations that contain only the lowest order of nonlinearity will fail at sufficiently large motion. Failure can also be expected if the motion involves significant bending and twisting that are not allowed for in the equations. The more comprehensive equations[7] of Watzky must then be used, perhaps generalized, if possible, to allow for nonuniformity of the cross section and material properties of the wire or nonlinear elastic behavior. Because of these difficulties, exploratory experimental measurements of the nonlinear vibrations of a wire appear to be useful in understanding the motion and guiding the theoretician. Nature can be relied on to ''use'' the correct equations and ''solve'' them correctly.

Earlier experimental measurements of nonlinear responses to a sinusoidal driving force have been reported,[8] together with a summary of the theoretical and experimental work on this subject prior to 1994. The responses measured include generation of transverse motion perpendicular to the driving force for small as well as large amplitudes and sudden amplitude jumps and hysteresis effects for large amplitudes. The phase relations between the components of transverse motion are such that as the driving frequency is increased, the motion changes from planar to tubular and finally collapses.

The work reported here is a sequel to the previous study. The emphasis is on additional factors affecting the nonlinear generation of perpendicular motion, the generation of higher harmonics, and on the rich variety of possible motions. Included in the latter are: (1) ''spontaneous'' transitions from one stable state to another with very different harmonic content, (2) generation of subharmonics, (3) slow oscillations from one quasistable state to another, and (4) quasiperiodic motions in which a pattern of motion is executed rapidly as it is driven with a period of a few hundredths of a second while the pattern itself changes slowly with a period of many seconds.

All of the above-described phenomena are greatly affected by a property of the wires of which there was no awareness in the earlier study: each natural frequency of free vibration is split into two closely spaced frequencies with separation Δf relative to the average natural frequency f_{av}, that is, $\Delta f/f_{\mathrm{av}}$, of the order of 0.1%–2%. Investigation shows that the splitting is not caused by asymmetric end connections. Furthermore, as discussed in Sec. III, there is a ''characteristic'' transverse axis for each of the two frequencies into which a natural frequency is split. The splitting and associated characteristic axes have a profound effect on the nonlinear generation of large motion perpendicular to the driving direction. Especially crucial in this regard are the amount of splitting and the direction of the driving force relative to the characteristic axes. Great care is necessary in the interpretation of results so as not to be misled by *linear* generation of perpendicular motion (sometimes tubular) when the driving force has components along both characteristic axes. Consequently, the interpretation of some of the data in the earlier report[8] is now questionable. Since the characteristic axes are in very different directions for different free-vibrational modes, their existence can also have important effects on the nonlinear generation of higher harmonics observed in different directions.

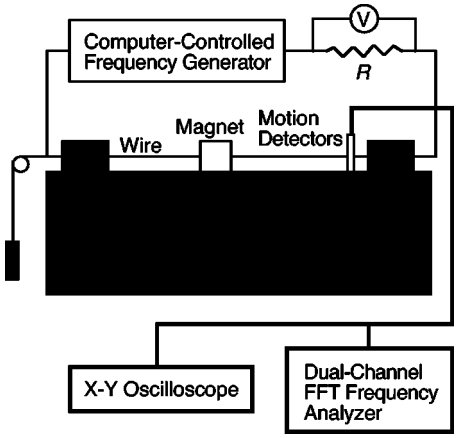


FIG. 1. Schematic diagram of the experimental apparatus for observing and measuring displacements of a point on a driven vibrating wire. Tension in the wire is established with a weight hanging from the pulley, after which the wire is clamped at both ends. Earlier data were obtained with a Rockland Model 9040 analyzer and more recent data with a Stanford Research Systems Model SR785 FFT dual-channel analyzer with a built-in precision frequency generator. An interfaced computer controls the frequency generator and records the analyzer output.

II. EXPERIMENTAL METHODS

The experimental arrangement shown in Fig. 1 is largely as described for the earlier studies.[8] A major improvement is the capability of rotating the magnet and adjusting its position with an uncertainty of 0.1°. A 70-cm-long brass harpsichord wire[9] is mounted rigidly at its ends on a steel I-beam, 14 cm wide, embedded in a concrete slab with a length of 115 cm, width of 46 cm, and depth of 14 cm. It is driven electromagnetically by an ac current in the wire passing through the magnetic field of a permanent magnet mounted several centimeters from the midpoint of the wire. The value of resistance R is 345 or 40 Ω depending upon the current range needed. The back emf induced in the wire when vi-

brating at the largest wire amplitude reported in this article is measured to be 1.5% of the total potential drop across the 40 Ω resistor and wire.

The motions in two orthogonal transverse directions of a very small segment of the wire are sensed with optical interrupter detectors (OI),[10] the voltage outputs of which are fed to a dual-channel fast-Fourier-transform (FFT) frequency analyzer and an X-Y oscilloscope. Since the linear region of the OI detectors is about 0.2 mm of total wire motion, the detectors are located within about 6–10 mm of a clamped end of the wire for large-amplitude motion. For extremely small-amplitude motion, the detectors are moved closer to an antinode.

The ends of the vibrating portion of the wire are held with collets which can be clamped to mounts on the steel beam. Each collet consists of a small sleeve with four equally spaced longitudinal slits and an external taper which tightens when pulled into an internally tapered fixed socket and thereby grips the wire where it runs through the narrow opening in the sleeve. These collets provide clamping with very nearly cylindrical symmetry. After the collets are clamped to the wire, a given tension is established by clamping the right-hand collet (in Fig. 1) to the beam, permitting the hanging weight to establish a nominal tension, and finally clamping the left-hand collet to the beam. Subsequently, the clamping is unaltered, but the tension unavoidably varies due to ambient temperature changes. For careful measurements near a resonance, special efforts are made to keep the ambient temperature constant to prevent resonant frequency shifts. A styrofoam enclosure is especially beneficial.

Since stiffness affects the shape of a normal mode near the collet, an assumption of sinusoidal shape is not valid for conversion of amplitude measurements at an OI detector position near a collet to corresponding amplitude values at an antinode. To obtain the relation between amplitudes at an OI detector position and those at an antinode, a photodiode array (PDA) detector[11] is used for measuring amplitudes of the order of 1–2 mm peak at an antinode. Simultaneous measurements at the OI and PDA detector positions made while driving at the resonant frequency of the mode being examined provide the conversion factors from OI data to antinode amplitudes for that mode. Limited determinations of the shapes of the normal modes utilizing simultaneous OI and PDA measurements at various longitudinal positions agree well with the calculated shapes based on a linear model for a brass wire with significant stiffness.

III. WIRE ASYMMETRY

The splitting of the frequency of each mode of free vibration into two closely spaced frequencies was not observed for the data reported in 1994.[8] Though the wire sample used is no longer available, it is likely that the splitting, if it existed, would have been less than 0.1 Hz out of 70 Hz for the conditions of that work. The resulting beats would have had a period of 10 s or more, at least as large as the decay time of the free vibrations. Hancock[12] observed splittings of some of the resonant peaks in measurements of the frequency response of musical strings subject to forced vibrations of very small amplitude. Those measurements were under conditions of high tension and very low amplitudes for which nonlinear effects probably were small. He hypothesized deviations from perfect circularity of the cross section of the wire as a possible cause and deliberately flattened some wires to explore this.

In the current work, long data periods for high-resolution FFT measurements of the free vibrations reveal the amount of splitting and precise values of each of the frequency components. Further, they reveal that for each of the two closely spaced frequencies the free-vibration signals from the two detectors are either in phase or 180° out of phase. As a consequence of this, the motion can be viewed as a superposition of a low-frequency vibration along a low-frequency ''characteristic'' axis and a high-frequency vibration along a high-frequency ''characteristic'' axis. Based solely on the observed phase relations, nothing can be said about the orientation of one axis relative to the other. A detector oriented to measure motion perpendicular to, say, the high-frequency axis would show only low-frequency vibration and vice versa. Should the characteristic axes be orthogonal, these two detector orientations would coincide with the axes. If, then, a wire is driven with a sinusoidal force, the direction of wire motion depends strongly on the driving frequency and the orientation of the characteristic axes. It is observed that the angle between the two characteristic axes is $90° \pm 0.1°$, except for some extreme cases discussed in Sec. IV C of an imposed twist of as much as 180° for which there can be a deviation up to 0.5° from 90°.

For each of the three modes beyond the first (approximately second, third, and fourth harmonics) there is also observed a splitting of the same order of magnitude as for the first mode, but bearing no simple numerical relation to it. As with the first mode there are observed orthogonal characteristic axes associated with each frequency of a closely spaced pair. No simple relationship is apparent for the orientation of the characteristic axes for one mode relative to another. For the wire samples used it is observed that the axis for the *higher* of the two frequencies for the second mode corresponds within about 10° to the axis for the *lower* of the two frequencies for the first mode. As discussed in Sec. VI, this could be relevant to a possible interpretation of the difference in the second harmonics in Fig. 9. The low axis for the third mode nearly coincides with that for the first mode and the high axes also nearly coincide.

A possible cause of frequency splitting could be asymmetric bending stiffness of the wire. For if it bends more easily in one direction than in another, the restoring force due to stiffness would be less in the first direction and greater in the second. Consequently, free vibration in the first direction would have a lower frequency than that in the second.

To explore this hypothesis for a given wire, the wire was first placed under a sufficiently high tension that any asymmetry in restoring force is masked by the fundamentally symmetric restoring force due to tension. (Experimental measurements of the effect of tension on frequency splitting are discussed later in Sec. IV C.) The resulting spectrum of free vibrations exhibits no clearly resolvable splitting and is measured through the first 15 modes. To fit the data, the linear theory of the normal modes of free vibration of a stiff

wire with clamped ends is employed. In the dimensionless form in which the theory is used, there are only two free parameters, one of which expresses the stiffness of the wire. By adjustment of the parameters, it is found possible to obtain a nearly perfect fit to the spectrum, thus validating the theory employed.

The tension in the wire was then reduced sufficiently that splitting, now resolved, was measured for the first 12 modes. A plot of frequency splitting versus mode number yields an oscillatory shape, rather than a monotonically increasing one. The splittings for modes 4, 6, 8, and 10 are much lower than those of the adjacent odd modes whereas for mode 2 the splitting is higher than any of the others. This means that the splitting cannot be explained as due to a low stiffness in one direction and a high stiffness in another direction, as might be produced by a simple asymmetry in the cross section of the wire. This is perhaps not surprising, particularly in view of the variation in the direction of characteristic axes, either low or high, as one goes from mode to mode. The underlying cause of the splitting thus remains unknown but needs to be illuminated since it is an important part of the basic physics of the wires and their vibrations.

IV. EXPERIMENTAL MEASUREMENTS OF FREQUENCY SPLITTINGS AND ORIENTATIONS OF CHARACTERISTIC AXES

A. Factors affecting frequency splitting

The following evidence indicates that frequency splitting and the existence of a characteristic axis for each frequency component are not due to lack of cylindrical symmetry of the collets that clamp the ends of the wire. Rather, they depend only upon properties of the wire itself and conditions imposed upon it, especially tension and to a lesser extent twist:

(1) If the collets are loosened, rotated about 30°, and retightened without rotating the wire, the orientations of the characteristic axes remain unchanged, and the amount of splitting is unchanged.
(2) As the tension is increased there is a *sharp decrease in absolute splitting* (see Sec. IV C). If the splitting were due to an asymmetric mount it would be expected that the *relative* splitting would be almost constant and that there would be an *increase in the absolute splitting* as the tension increases.
(3) Relative splittings differing sometimes by a factor greater than 2 are observed for different wire samples under the same tension, whereas the process of loosening, rotating, and retightening the collets on a given sample has little effect.
(4) Since, for a given tension, the natural frequency is inversely proportional to the wire length, a typical splitting of 0.4% would require a difference of length of 0.4% along two perpendicular longitudinal planes if the end mounts rather than the wire itself were the cause. For the 70-cm-long wire samples used, this would represent a difference of length of about 3 mm, which by inspection of the collets can be seen to be unrealistic.
(5) If the wire is given a twist by rotating the collet at one end, but with the collet remaining clamped to the wire, the amount of splitting can be changed significantly.

B. Methods of determining orientations of the characteristic axes

Two distinctly different methods are used to determine the orientation of each of the characteristic axes. The first is for a *freely vibrating* wire and the second for the wire driven by a sinusoidal force. The methods described in the following are for determining the *low-frequency* axis, but analogous procedures are used to determine the high-frequency axis. It should be noted that orthogonality of the high and low axes is not assumed, but the measurements do show essential orthogonality.

Method (1). For the freely vibrating wire following a pluck, a determination of the low-frequency axis orientation relative to the axis of one of the detectors is made from the low-frequency component amplitudes of the outputs of the horizontal and vertical OI detectors. These are obtained with the high-resolution FFT analyzer. The accuracy of this method is a few degrees. It has the advantage of being done quickly and can yield results at any particular longitudinal position of the OI detectors along the wire.

Method (2). The wire is driven with a relatively high driving force at the *low-frequency resonance in a direction nearly perpendicular to the low-frequency axis*. The initial trial choice for this driving direction is obtained from the result of Method (1). Even though the driving force is relatively high, the responses in both directions are relatively low so that the nonlinear effects are not significant. In the driving direction the response is low because, though the *driving direction* is essentially along the high-frequency axis, the *driving frequency* is in the small tail region of the high-frequency response function. In the direction nearly perpendicular to the driving direction the response is low, even though the *driving frequency* is at the peak of the low-frequency response function, because the *force component* is very small. Though the response is in an essentially linear region of wire motion the pattern of the detector outputs observed on the X-Y oscilloscope has an oval shape as expected from linear theory since the driving frequency is at the low resonance and much below the high resonance, thus yielding a 90° phase difference. The driving direction is then adjusted for a minimum (essentially zero) oscilloscope deflection in the direction perpendicular to the driving direction so that the pattern is a straight line indicating no force component in the direction of the low axis. The direction of the low-frequency axis is then determined as 90° from the driving direction. Thus, the low-frequency axis is experimentally defined as the direction perpendicular to the driving direction in which there is essentially no response at the low-resonance driving frequency. Though this method is somewhat tedious, it is capable of very high precision of $\pm 0.1°$. The determination that the driving force is perpendicular to a characteristic axis refers to the magnet position along the wire, normally kept fixed near the center of the wire, and so the method is restricted to a given longitudinal position.

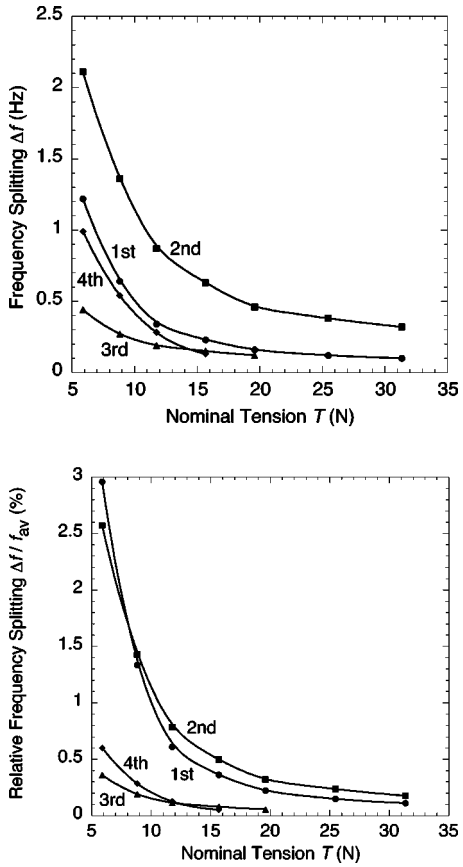


FIG. 2. (a) Frequency splitting Δf of each of the first four modes of free vibration following a pluck of the wire vs nominal tension T in the wire. (b) Frequency splitting Δf relative to the average resonant frequency f_{av} for each of the first four modes of free vibration following a pluck of the wire vs nominal tension T in the wire.

C. Dependence on wire tension and applied twist

An example of the dependence of frequency splitting on the wire tension is shown in Fig. 2(a) for the first four modes of free vibration. The modes are approximately but not exactly harmonically related. It should be noted that the separation Δf of the natural frequencies of free vibration decreases markedly with increased tension. Since the restoring force provided by tension is intrinsically symmetric, higher tension will diminish the relative importance of any asymmetric restoring force and should reduce the frequency splitting of the modes. It is illustrated in Fig. 2(b) that $\Delta f/f_{av}$ decreases even more sharply with increasing tension. Furthermore, it should be noted that there is no apparent simple relationship among the sizes of the relative splittings of the modes. This is probably because the asymmetry in the wire and the anisotropic behavior that it produces vary along the length of the wire. Since each mode will be most affected by the behavior in the vicinity of its antinodes and the location of these vary from mode to mode, it is expected that there will be different splittings and characteristic axes for different modes with no simple relation among them.

Measurements show a large dependence of splitting on an applied twist. The twist is applied by rotating one end of the wire by an amount referred to as the twist angle while the other end is held fixed. There is no apparent simple relationship among the splittings of the four modes, and the dependence on the sense of the twist is very asymmetric. For a particular sample, much larger splittings are generally observed for clockwise twists (as viewed from one of the ends) than for counterclockwise twists, except that for a clockwise twist of about 50° the splitting is essentially zero for the third and fourth modes. For this sample the effect of a clockwise twist is most pronounced for the second and third modes; the splitting for the second ranges from 0.2 Hz at a twist of 0° to 1.9 Hz at 180° and for the third from 0.1 to 1.4 Hz for the same change in twist. It is possible under certain conditions to apply a twist that eliminates an observable splitting of the fundamental. An interesting use of a vibrating wire as a detector, the associated problems due to splitting of the degeneracy of the normal modes, and the effect of twisting, all applied to detection of superfluid circulation in liquid helium II, are described by Vinen.[13] The nonuniformity in cross-sectional area or elastic properties of gut musical strings resulting in a ''false'' response and also the beneficial effects of twisting are discussed by Gough,[14] with reference to private communication with Benade.

The rotation of the low- and high-frequency axes produced by twisting is approximately one-half of the twist angle. This is intuitively reasonable since it is an average of the rotation of one end through the twist angle and the other end through zero. For twist angles varying from 0° to 180° it is found for a given sample that the angle between the two characteristic wire axes for the first mode is changed continuously from 89.9°±0.1° to 89.5°±0.1°. As explained in Sec. V B it is advantageous for obtaining the data presented in most of the following figures to have a characteristic axis aligned nearly perpendicular to one of the OI detector axes. The detector axes are essentially fixed, so it is difficult to obtain this condition closer than several degrees by only rotating both ends of the wire. The final fine adjustment to within about 0.3°–0.5° is accomplished by twisting one end of the wire relative to the other. For the data of Figs. 7(a), (b), and 9, significant twist is employed to get the large splittings desired for those data.

It should be emphasized that when a large twist is applied at one end, measurements show that the new characteristic axis orientations are independent of the *longitudinal position along the wire*. It is not found, as one could expect, that the rotation of the axes is greater at an approximate midpoint of the wire than at the end of the wire held fixed. This is graphically exhibited when axis orientations are first measured with no applied twist and then with a very large twist at the one end of about 100°. The resulting rotations of the characteristic axes are measured at different positions by Method (1) and Method (2) outlined in Sec. IV B. Though the precision is lower by Method (1), the large characteristic axes rotation of about 50° at the *end of the wire held fixed* during the twist process is in agreement with the amount of rotation as determined by Method (2) utilizing the magnetic driving force near the *middle of the wire*.

This somewhat surprising result could possibly be interpreted as follows: it may be imagined that after one end of the wire is twisted, but before vibrations are started, the orientation of any given characteristic axis varies from one end of the wire to the other. When vibrations are started, there

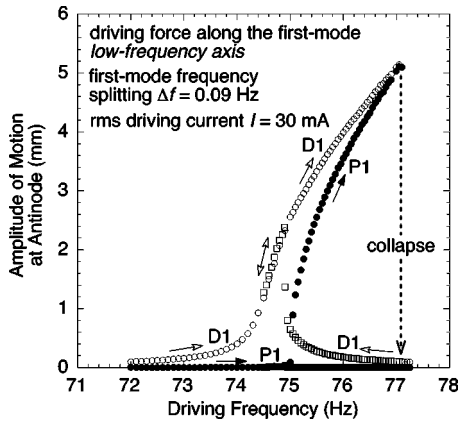


FIG. 3. Frequency responses near the first-mode resonance. The large driving force is in the direction of the first-mode *low-frequency* axis. Motion in the driving direction is denoted by D1 and in the perpendicular direction by P1. Arrows differentiate between data obtained with increasing driving frequency and those obtained with decreasing driving frequency. Also, circles denote increasing frequency and squares decreasing frequency (open for D1 and closed for P1). Note the rapid rise of P1 as frequency increases near 75 Hz. Note also that after collapse P1 remains negligible on the scale shown as the driving frequency is decreased (not indicated with an arrow).

will then be an initial tendency for motion at the frequency associated with the axis to take place in different directions along the wire. That this is happening will be communicated to all elements of the wire by waves, perhaps partly or wholly of a torsional nature. Since what drives the waves are differences in the direction of motion, it seems likely that their effect will be to reduce such differences until a sort of equilibrium is attained and all directions of motion are the same. Under conditions typical of musical strings and certainly typical of the work reported here, torsional waves travel much faster than transverse ones. Thus, it would be possible for them to maintain the special kind of equilibrium described even as transverse motion evolves. This would be analogous to the role played by longitudinal waves, which, as noted in Sec. I, also typically travel much faster than transverse waves and thus have an equilibrating effect on tension in the wire. By using the comprehensive framework of Watzky,[3,4] it might be possible to confirm or refute this explanation and better understand how characteristic axes, the same everywhere along the wire, are produced and how and why they vary from mode to mode. This would require, of course, that Watzky's equations be generalized to allow for nonuniformity in the wire. They could then perhaps be simplified by treating the nonuniformity as a perturbation, a reasonable step given that frequency splittings are at most a few percent.

V. GENERATION OF MOTION PERPENDICULAR TO DRIVING DIRECTION AND HYSTERESIS EFFECTS

A. Generation due to nonlinear processes

The frequency response of the wire near the first-mode resonance for a large driving force and a small splitting of 0.09 Hz is shown in Fig. 3. The characteristics of collapse and hysteresis are similar to those shown in Fig. 4 of Ref. 8, but the measurements are carried out with the driving force essentially in the *known* direction of the *low-frequency axis* of the fundamental vibrational mode. In these and succeeding graphs D1 refers to motion in the driving direction and P1 to motion in the perpendicular direction at the first harmonic of the driving frequency, that is, at the driving frequency. Similarly, in later figures D2 refers to motion in the driving direction at the second harmonic of the driving frequency, etc. As predicted by theory in which changing tension is included, there is no longer a classical resonance response for a large driving force but rather a region of large amplitude followed by collapse and hysteresis when the driving frequency is then lowered. This can be interpreted intuitively as follows: as the free-vibration frequency measured at low amplitude is approached, the amplitude of the wire motion for a large driving force increases. This results in a higher average tension which in turn raises the resonant frequency. In effect then, the resonant frequency is ''chased'' but never quite reached because it is always beyond the driving frequency. In the region before collapse, the perpendicular motion generated by nonlinear processes is nearly equal in magnitude to the motion in the driving direction. The motion is observed as tubular because of the 90° phase difference between the motion in the two directions. At nearly maximum delivered power (driving force in phase with velocity), the system becomes unstable and collapse occurs. After collapse the resonant frequency is much lower than the driving frequency so there is little wire motion as the driving frequency is reduced until there is a sudden upward jump to the same path followed as when the frequency was increasing. Though D1 then increases sharply, the amplitude does not become great enough to generate P1 motion perceptible on the scale shown unless the driving frequency is again increased.

Generation of perpendicular motion as shown in Fig. 3 is affected very strongly by the combined effects of the amount of frequency splitting (varied by tension change) and the strength of the driving force. These dependencies are illustrated in Figs. 4(a) and (b) for earlier data with slightly different wire conditions of twist and tension yielding the values of Δf. An important conclusion is that the degree of overlap of the low-frequency and high-frequency response functions in the driving and perpendicular directions appears to be a significant factor in determining the amount of nonlinear generation of perpendicular motion and the driving force required to produce it. Note, however, in the discussion of Figs. 6(a) and (b) that when driving along the *high-frequency* axis, the amount of time allowed for buildup of perpendicular motion for the driving frequency in the vicinity of the *low-frequency* resonance can also be an important factor.

Measurements for a driving force along the low-frequency axis under very carefully controlled conditions for a small splitting of 0.090 Hz are made near the critical regions of frequency and amplitude for significant generation of P1 motion. They indicate that the sharp increase in P1 due to nonlinearity occurs within essentially the same narrow driving-frequency range regardless of the amount of the driving force *provided that the force is large enough so that collapse does not occur before that driving-frequency range*

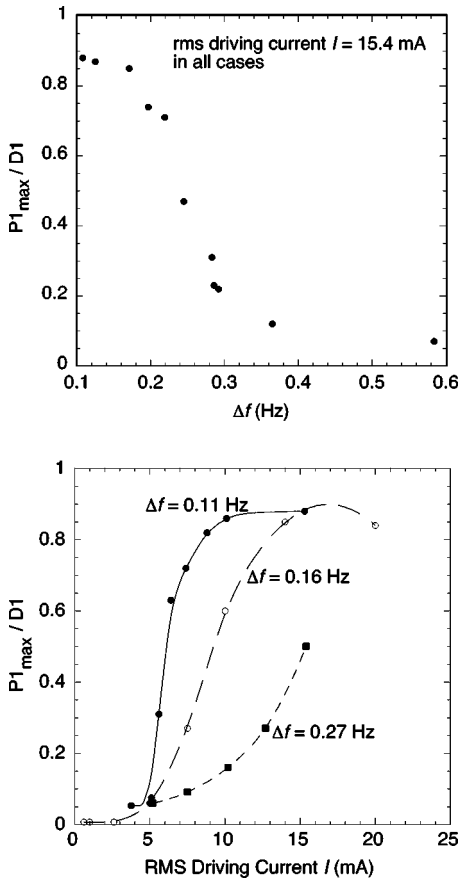


FIG. 4. (a) Ratio of maximum perpendicular motion $P1_{max}$ to motion D1 in the driving direction vs frequency splitting Δf. The large driving force is in the direction of the first-mode *low-frequency axis*. (b) Ratio of maximum perpendicular motion $P1_{max}$ to motion D1 in the driving direction vs driving current for three different frequency splittings. The driving force is in the direction of the first-mode *low-frequency axis*.

is reached. It is observed, however, that the range becomes narrower as the driving force is increased. For a rms driving current $I=7.42$ mA, collapse of D1 occurs just after a small but sharp rise in P1. For $I=7.60$ mA, the amount of the sharp increase in P1 before collapse occurs is about six times the increase at 7.42 mA. At higher currents, P1 increases sharply within a narrow frequency range and more gradually at higher frequencies. The width of the range within which there occurs a sharp rise of about a factor of 10 greater than the rise for 7.42 mA is 0.035 Hz for $I=7.78$ mA, 0.025 Hz for $I=8.84$ mA, and 0.01 Hz for $I=10.6$ mA.

In Fig. 4(b) it can be seen that for a splitting of 0.16 Hz and a rms driving current of 15 mA, the P1 max/D1 ratio is nearly 0.9 for driving along the *low-frequency axis*. A frequency response curve under those conditions is shown in Fig. 5, which exhibits a substantial second harmonic in both of the directions. Subsequent figures also show significant second-harmonic generation. An especially interesting example when driving in the direction of the *high-frequency axis* is shown in Fig. 6. In Fig. 6(a) where the driving frequency is increased at a normal rate of about 0.005 Hz/s there is very little generation of P1 motion, but there is a very large amount of perpendicular second-harmonic motion generated by nonlinear processes. In Fig. 6(b) the frequency is increased very, very slowly in the vicinity of the low resonance (even adjusted back and forth to follow small resonance shifts due to ambient temperature changes), resulting in a slowly growing large P1 response. Once P1 is ''nursed up'' to a large value the state is robust and P1 is relatively

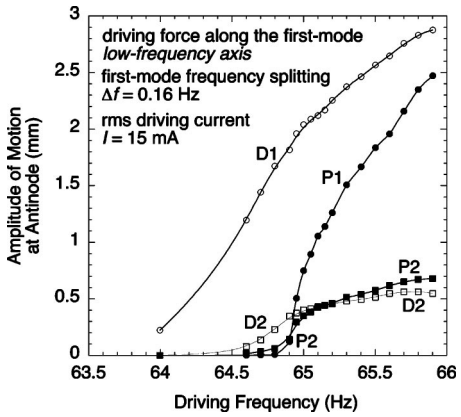


FIG. 5. Frequency responses near the first-mode resonance. The large driving force is in the direction of the first-mode *low-frequency axis*.

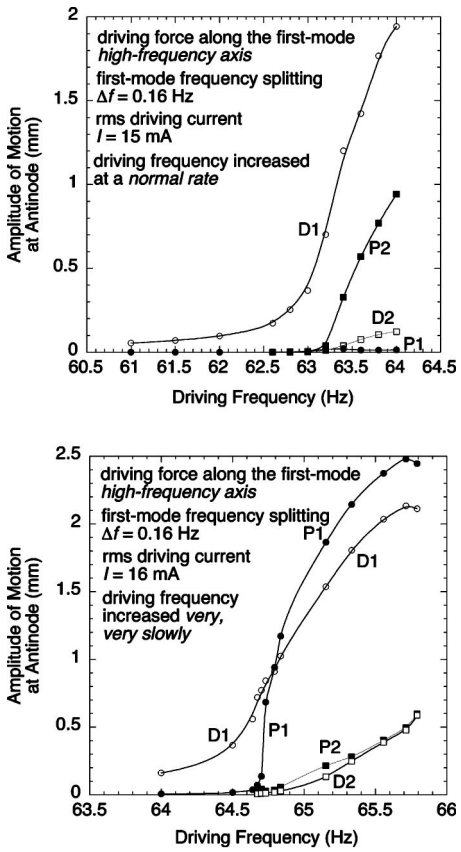


FIG. 6. (a) Frequency responses near the first-mode resonance. The large driving force is in the direction of the first-mode *high-frequency axis*. The driving frequency is increased at a normal rate of about 0.005 Hz/s. (b) Frequency responses near the first-mode resonance. The large driving force is in the direction of the first-mode *high-frequency axis*. The amplitude P1 is very much greater than in (a) because the driving frequency is increased extremely slowly and held almost constant in the vicinity of the low-frequency resonance for which the characteristic *low-frequency axis* is in the P1 direction. The driving frequency is even adjusted back and forth to follow slight changes in the resonant frequency of the wire due to ambient temperature changes, thereby allowing time for a very slow buildup of P1. The behavior of P2 also differs greatly from that in (a).

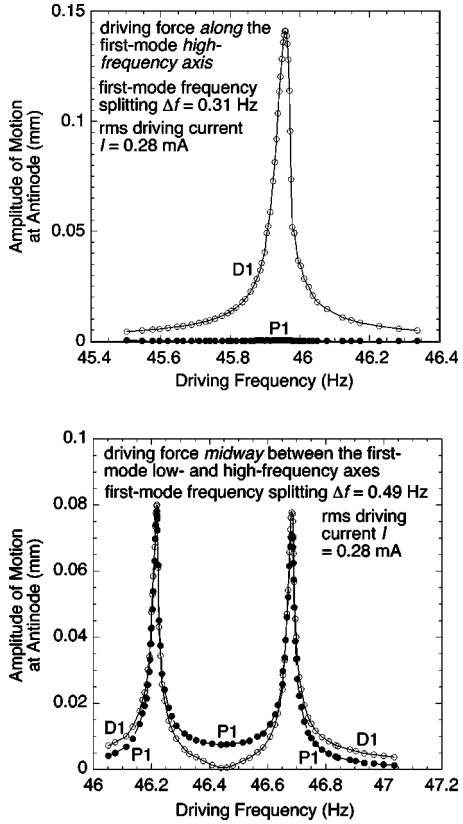


FIG. 7. (a) Frequency responses near the first-mode resonance. The weak driving force is in the direction of the first-mode *high-frequency axis*. The response P1 in the perpendicular direction is shown to be essentially zero. No response for either D1 or P1 is evident in the region of the low-frequency resonance. (b) Frequency responses near the first-mode resonance. The weak driving force is in a direction *midway between the two first-mode characteristic axes*. The splitting is large enough that the responses at the two resonant frequencies are clearly separated. The interpretation as a linear phenomenon is discussed in the text. The splitting $\Delta f = 0.49$ Hz was obtained by a FFT measurement of the free vibration of the wire.

insensitive to further increases (slow or fast) in driving frequency, but remains large. It should also be noted that the final value of P2 is smaller if the P1 value is ''nursed up'' to its high value as in Fig. 6(b). This reciprocal relation between P1 and P2 is observed in many trials and is also very evident in the oscilloscope display of the motion of a point on the wire.

B. Generation due to linear processes

Measurements to determine the generation of perpendicular motion by nonlinear processes can be very misleading if there is not great care in taking into account linear effects of wire asymmetry, easily ignored if the relationship of the characteristic axes to the driving direction is not considered. Since the effects of frequency splitting and characteristic axes are most pronounced at relatively low tensions, measurements are presented in this section under conditions of low tension and high splitting to illustrate the problem.

In Fig. 7(a) is shown the response of the wire when driven very weakly, so that linear processes should dominate, in the direction of the high-frequency axis. There is negligible P1 motion evident on the linear plot since there is

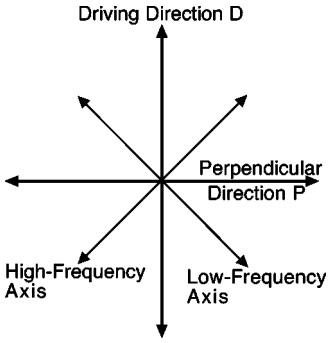


FIG. 8. Relation of the low- and high-frequency axes and the driving D and perpendicular P directions to illustrate the linear interpretation in the text of Fig. 7(b).

essentially no force component in that direction. In Fig. 7(b) the driving direction is at approximately a 45° angle with respect to each of the two characteristic axes. In this case a double peak is observed since the components of force along the two characteristic axes are essentially the same. Furthermore, there is P1 motion comparable to that of D1 at each of the two peaks. A careful mathematical treatment demonstrates that this generation of P1 motion in Fig. 7(b) can be interpreted solely in terms of *linear* processes. Also, for this low-amplitude motion the amplitude is found to be proportional to driving current, an important aspect of linearity. A qualitative interpretation of Fig. 7(b) can be made with reference to the diagram in Fig. 8. When driving at frequencies somewhat below or somewhat above the two resonances, the responses along the two characteristic axes will be comparable and in phase; thus the motion components will reinforce in the D direction and mostly cancel in the P direction. When driving at the low resonance, almost all of the response is along the low-frequency axis which has equal components along the D and P directions. Midway between the two resonances there will be a 180° phase difference[15] between the equal responses along the two characteristic wire axes; thus the components in the D direction will cancel and in the P direction will add, in agreement with the higher value of P1 in that region. When driving at a frequency near one of the resonances in a direction different from its characteristic axis the straight-line motion displayed on an X-Y oscilloscope is thus approximately along the *characteristic axis* for that resonance *rather than in the driving direction*. This occurs provided that the motion is low amplitude where nonlinear effects are insignificant, and that there is sufficient splitting of the response curves of the two frequencies. For somewhat smaller splitting, however, oval motion is observed, apparently due to linear effects.

The above-noted results call into question the interpretation of Figs. 6–8 of the earlier report[8] for very weak driving forces and small amplitudes. The P1 motion in those data, except for the upper decades of Fig. 8, was probably due to linear rather than nonlinear processes. The nearly constant ratio of the D and P amplitudes (labeled as Y and Z, respectively) for the lower three decades of amplitude in Fig. 8 suggests that only linear processes were involved in that region. The tubular motion in Fig. 7 of that report could have a linear interpretation. Though the same wire and experimen-

tal conditions are no longer available, a response curve similar to those in Fig. 6 of Ref. 8 are obtained for a small splitting and appropriate orientation of the characteristic axes relative to the driving force and the detectors.

In order to avoid misleading linear effects described above when investigating nonlinear aspects of the frequency response of wire motion, the *driving direction should be aligned with a characteristic axis*. If this were not done the motion would tend to be in the direction of a characteristic axis when driving near the corresponding frequency, and so there would be a component of motion perpendicular to the driving direction due to linear effects. Additionally, it is advantageous for the driving direction and collinear characteristic axis to be perpendicular to a detector axis; all motion measured by that detector is then perpendicular to the driving direction and can be attributed to nonlinear effects.

The procedures for obtaining the data of Figs. 7(a) and (b) well illustrate the severe experimental difficulties associated with ambient temperature changes which must be overcome to obtain some of the data for this paper. Calculations show that the (linear) normal-mode frequencies of the wire change by about -3.1% per °C temperature change, which means that small ambient temperature changes produce large effects. Applied to the D1 resonance shown in Fig. 7(a) this means that a temperature change of about 0.03 °C would shift the resonance by its full width at half maximum! Much smaller temperature changes produce apparent discontinuities in the data, especially in the region of sharp increase or decrease on each side of the resonance curve. The problems are overcome by enclosing the system in a styrofoam box and operating under stable conditions of ventilating system off, cool weather without sunshine, doors and windows tightly closed, etc. The very thick stone and masonry walls of the old laboratory building are also helpful in achieving temperature stability.

VI. GENERATION OF HIGHER HARMONICS

Generation by nonlinear processes in the wire of large second harmonics of the driving frequency, both D2 and P2, is exhibited in Figs. 5 and 6. In addition, significant higher harmonics are observed. In Fig. 9 are shown frequency response curves for a large driving force perpendicular to the low-frequency axis and hence approximately in the direction of the high-frequency axis. In addition to the commonly observed second harmonic there are significant third and fourth harmonics. In this example the splitting is very large, achieved with a low tension and significant twist. Because of this large splitting the value of P1 is very small in the high-amplitude region of D1, in contrast to some preceding figures for smaller splittings. There appears to be no obvious interpretation of the large dip in P1 shortly below the collapse frequency, but it is another example of the rich variety of responses in a nonlinear asymmetric system.

Relatively small harmonics observed in the output voltage of an OI detector have two possible sources: actual wire motion and nonlinearity in the response of the OI detector to wire motion. The latter becomes increasingly important as the detector is exposed to larger and larger wire motion and can mislead the observer about the motion of the wire. The

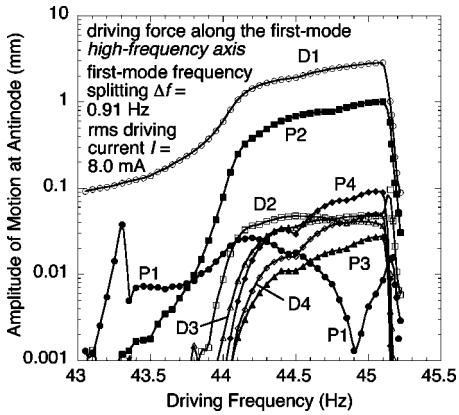


FIG. 9. Generation of the first four harmonics in the driving and perpendicular directions for a moderately large driving force in the direction of the first-mode *high-frequency axis*. Because of the large frequency splitting of 0.91 Hz, the measured P1 motion is very small except for the peak in the region of the low-frequency resonance, due perhaps to a small component of driving force along the low-frequency axis.

importance of minimizing contributions due to nonlinear detector response is especially important when measuring small amplitudes of the order of a few percent of the fundamental for the third and fourth harmonics. The procedure finally adopted after much study and many investigations is to first locate the detector very near the end of the wire and adjust the driving current to such a low value that no third or fourth harmonics are evident in the output of the detector. With the same driving current the detector is then moved farther and farther from the end and hence to regions of greater and greater wire motion until third or fourth harmonics, likely due to detector nonlinearity, are observed. From such measurements an acceptable operating region of the OI detector is selected. By locating the detector only 0.6 cm from the end of the wire for the data of Fig. 9 a satisfactory range of wire motions can be measured while restricting the artificial generation by the OI detector to less than 1/10 of the observed maximum amplitudes for the third harmonic and much less for the fourth. Thus the small amplitudes of third and fourth harmonics are concluded to be valid indications of actual wire motion at those harmonics due to nonlinear processes. Similarly, third and fourth harmonics are also observed when driving essentially along the low-frequency axis.

It is noteworthy in Fig. 9 that P2 is much larger than D2, which can possibly be interpreted as follows, with careful attention to the fact that the driving force is along the *high-frequency axis* of the first mode: When D1 and P2 are large, the driving frequency is in the neighborhood of the high resonant frequency of the first mode. In that neighborhood, the second harmonic of the driving frequency matches more closely with the high resonant frequency of the second mode than with the low resonant frequency of that mode. Thus, nonlinear generation of a second harmonic might be expected to cause considerable motion in the direction of the high-frequency axis of the second mode. For the wire sample represented in Fig. 9, that axis is within a few degrees of being *perpendicular* to the high-frequency axis of the first mode. It then follows that the second-harmonic response in

TABLE I. Example of a sharp transition between a state with little D3 and D4 to one with large D3 and D4 and reduced D2. The driving force is along the low-frequency axis at a current of 1.00 mA rms, which is large enough to bring the system into a nonlinear region but much smaller than for the response curves in Figs. 3, 5, 6, and 9. All measurements are made just before collapse.

Time	D1 (at OI)	D2/D1	D3/D1	D4/D1
Day 1	198 mV	15%	0.1%	0.1%
Day 2	95 mV	4%	59%	7%
Day 3	107 mV	1.5%	42%	15%
Day 4	110 mV	2.4%	43%	18%
Day 5	99 mV	9%	42%	26%
Day 5[a]	178 mV	15%	0.2%	0.1%

[a]After abrupt state change with unchanged driving conditions.

the perpendicular direction would be greater than in the driving direction.

Also noteworthy in Fig. 9 is that P2 is much larger than P3 whereas the usual understanding of coupling between direct and perpendicular motion would lead one to expect the reverse. Since the tension in the wire changes with a period that is half that of the transverse motion, it and any wave velocity dependent on it vary at twice the driving frequency. In most nonlinear theories of vibrating wires, the square of wave velocity is multiplied by a vibratory displacement that varies at the driving frequency. By the usual arguments employed in cases of weak nonlinearity, this leads to the expectation that among the higher harmonics present in the vibration, the third harmonic will be the most prominent. But if the nonlinearity is strong, this may no longer be the case. Instead, perhaps the second harmonic will be dominant, as the experimental results shown in Fig. 9 indicate.

VII. RICHNESS OF POSSIBILITIES OF A NONLINEAR SYSTEM

As pointed out in Sec. I, the many possibilities for solutions of the coupled nonlinear equations of a driven vibrating wire made more complex by wire asymmetry suggest that experimental measurements should yield a rich array of very different types of behavior. This has indeed been observed, and some especially interesting observed motions are described in this section.

An example of dramatic changes of state involving different harmonic content under seemingly unchanged driving conditions is presented in Table I. The data for Day 1 and the second entry for Day 5 could be considered ''usual'' reproducible behavior of the system studied. The data for the intervening period between Day 1 and the second entry for Day 5 are for some ''unusual'' behaviors which occurred with seemingly identical driving and wire conditions except for small tension variations due to temperature. It should be noted that the maximum value of D1 before collapse is much smaller for the unusual behavior but that the sizes of the third and fourth harmonics relative to the first are much greater. The abrupt change from unusual to usual behavior on Day 5 occurred with no apparent cause, and the motion continued as usual for the remainder of the day. The 10% difference in D1 amplitudes just before collapse for Day 1 and Day 5[a] is not especially significant since the rate of increase of driving frequency, external disturbances, etc., can contribute to triggering of the collapse.

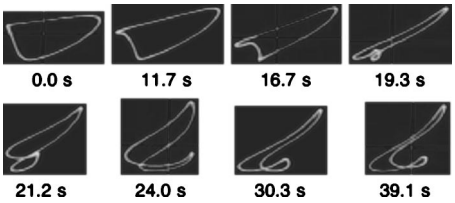


FIG. 10. Frames captured from a video recording showing changes in the oscilloscope pattern of transverse motion of a point on the wire when the wire is driven along the first-mode *high-frequency axis* (vertical in the pictures) with a large driving force ($I=25$ mA). The driving frequency is held constant at about 65.5 Hz (near the first-mode resonance) while the changes occur.

In Figs. 10 and 11 some dramatic transitions are illustrated in a series of frames taken from a video recording of the trace on the X-Y oscilloscope to which the inputs are signals from the two OI detectors. The trace shows the path followed by a point on the wire, the transverse motion of which is sensed by the (orthogonal) detectors. Each series is obtained during a pause in the slow increase of the driving frequency through the resonance regions. The driving force is vertical, essentially in the direction of the first-mode high-frequency axis.

It should be emphasized that prior to the time labeled 0.0 s in Fig. 10 the pattern had been relatively stable, changing only slightly as the frequency had been slowly increased. Then when it was obvious that a transition was starting to occur, the driving frequency was held constant during the entire 39 s period shown. The pattern shown in the last frame of Fig. 10 remained relatively stable, indicating that the transition to an entirely different state had occurred. The amplitudes of the first four harmonics of the driving frequency in both the D and P directions were recorded before and after the sequence and are consistent with the pattern changes shown. That is, initially P1 (horizontal) is larger than D1 and then gradually decreases, accompanied by significant increases in the higher harmonics. The region where P1 is significantly larger than D1 is in the region of the low-frequency resonance, the axis for which is in the P1 direction, thereby apparently enhancing the nonlinear generation of perpendicular motion. The transition which occurs at this frequency is then an important example of the profound ef-

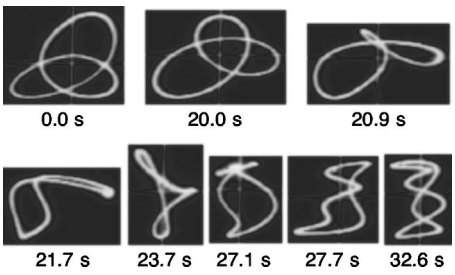


FIG. 11. Frames captured from a video recording showing changes in the oscilloscope pattern of transverse motion of a point on the wire when the driving frequency is held constant at a value approximately 0.5 Hz higher than in Fig. 10.

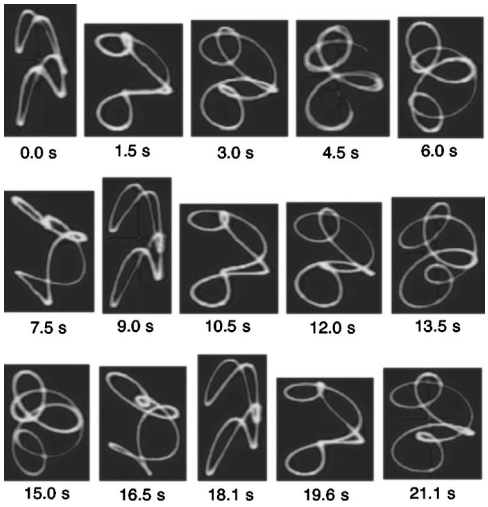


FIG. 12. Frames captured from a video recording showing oscilloscope patterns at 1.5 s intervals of transverse motion of a point on the wire being driven near the first-mode resonant frequency of 50 Hz with a moderately large driving force ($I=12$ mA) in the direction of the high-frequency axis. As can be seen in the patterns at 0.0, 9.0, and 18.1 s *or* the patterns at 3.0, 12.0, 21.1 s, the changing pattern repeats itself with a period of about 9 s. The same repeating pattern continued for many minutes beyond the 21 s.

fect which the splitting of the resonant frequency has on the wire behavior.

After the stable state shown in the last frame of Fig. 10 had been established, the driving frequency was again slowly increased, accompanied by an evolution of the pattern to that shown in the first frame of Fig. 11, for which the driving frequency is about 0.5 Hz higher than in Fig. 10. The change in pattern from the end of the series in Fig. 10 to the beginning of the series in Fig. 11 occurs gradually and continuously during the slow change of driving frequency.

At the time of the first frame of Fig. 11 another slow ''spontaneous'' transition began to occur and the driving frequency was then held constant during the entire sequence. The large D2 and P2 both drop by almost a factor of 10, accompanied by a large drop in D4, while P4 remains nearly constant and very significant, as is readily seen in the pattern of the last frame. The pattern in the last frame of Fig. 11 remained essentially stable as long as the driving frequency was held constant. Upon further increase of the driving frequency beyond Fig. 11, the pattern slowly evolved to one in which P4 was much reduced.

The series shown in Fig. 12 illustrates a very different phenomenon which, when observed on the oscilloscope and in the recording of it, appears quasiperiodic with a reproducing pattern period of about 9 s while being driven with a period of 0.02 s. The driving frequency is unchanged during the exhibited sequence, and the pattern sequences actually continue for many minutes with no obvious change in periodicity.

In Fig. 13 the driving frequency is that of the third mode and exhibited is a rather sudden transition (while the driving frequency remains constant) to a state where there are large peaks at several harmonics of *1/3 of the driving frequency*.

An interesting video example, not displayed here, shows that driving near the second-mode high-frequency resonance (horizontal direction) generates a subharmonic response at

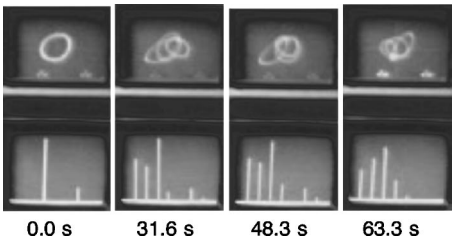


FIG. 13. The wire is driven with a large force at the third-mode resonant frequency indicated by the prominent line in the first FFT spectrum. As the driving frequency is increased slightly, several harmonics of 1/3 of the driving frequency are suddenly generated with an accompanying change of oscilloscope pattern. It can be seen from the linear vertical scale of the FFT spectrum that the generated amplitudes at 1/3 and 2/3 of the driving frequency are more than one-half as large as the driving-frequency amplitude. These phenomena were observed over ten years ago, at which time the observers were unaware of the significance of frequency splitting and characteristic axes.

one-half the driving frequency. This response is in the perpendicular direction (vertical) and is a factor of about eight times greater than the response in the driving direction at the driving frequency. The resulting oscilloscope pattern is a very slender figure eight and is stable as long as the driving frequency is at 88.2 Hz. If the driving frequency is then increased by 0.2 Hz and held constant, a slowly varying pattern is observed in which the ratio of the horizontal amplitude at the driving frequency to the vertical amplitude at one-half that frequency oscillates over a period of seconds from about 1:2 to 12:2. Varying phase relations also add to the complexity of the changing pattern. Video recordings including this phenomenon and the real-time patterns of Figs. 10–13, all on a CD, can be obtained at no charge upon request to the University of Northern Iowa Physics Department (319-273-2420 or becky.adams@uni.edu.) Limited clips of some of these video recordings can be viewed at http://www.physics.uni.edu/hanson.shtml.

It is noteworthy that for the phenomenon described in the preceding paragraph and the phenomena exhibited in Figs. 10–12 the driving force is always in the direction of the *high-frequency axis*. That is also the case for the data in Figs. 6(a) and (b). Many other examples of unusual behavior have been observed when driving along the high-frequency axis. If the driving force is along the high-frequency axis it might be expected that there would be little response until the high-frequency resonance region is reached. But the low-frequency resonance is inevitably encountered as the driving frequency is increased, sometimes yielding unusual and interesting behavior. Driving along the low-frequency axis is less likely to yield this kind of behavior because the high-frequency resonance is always ''pushed ahead'' and never reached. An exception is the extreme behavior of abrupt state changes shown in Table I which occurred while driving along the low-frequency axis.

VIII. MUSICAL SIGNIFICANCE

The investigations emphasize the fundamental properties of vibrating wires rather than their musical significance. Several observations of musical significance can, however, be made. As seen in Fig. 2(b) the relative amount of splitting

depends very strongly on wire tension in the region of low tension. Edward Kottick, a maker and historian of harpsichords and related instruments, points out[16] that, although conventional wisdom has always been that a string sounds best when tuned to within two or three semitones of the breaking point, there are building traditions in which much lower tensions are used. Musicians sometime describe an unsatisfactory string as ''false,'' a term which probably refers to the situation where the partial structure starts to get so far from harmonic that it becomes difficult to identify the pitch. Kottick finds that a yellow brass wire of the same diameter (0.55 mm) from the same manufacturer as for these measurements is no longer satisfactory when the tension is below the one corresponding to the point at 25.5 N in Fig. 2(b), where $\Delta f/f_{\mathrm{av}}$ is observed to be 0.15% for the first mode and 0.24% for the second one. For a different sample of the same type of wire at a lower tension of about 10 N, $\Delta f/f_{\mathrm{av}}$ is observed to be 1.0% and 1.6% for the first and second modes, respectively. In this case the difference between the high-frequency component of the second mode and twice the low frequency component of the first mode is about 1.3%, a significant part of a semitone (5.9%). The comparison of these two components is meaningful since the high-frequency axis for the second mode nearly coincides with the low-frequency axis for the first mode. This could well contribute to musicians' characterization of a string as ''false.'' It should also be noted that for this example the low-frequency component of the second mode is *lower* than the second harmonic of the low-frequency component of the first mode, and the high-frequency component of the second mode is *higher* than the second harmonic of the high-frequency component of the first mode. In other words, there is a ''flattening'' effect for the low-frequency components and a ''sharpening'' effect for the high-frequency components. Thus the usual ''sharpening'' effect for a linear stiff string is ambiguous in this situation.

Since nonlinear and asymmetric characteristics of a vibrating string also have a counterpart in plucked string behavior, that subject, though not presented in this paper, has musical significance. The precession of the motion of a plucked string as a consequence of nonlinearity and possible significance in a plucked or struck string are treated by Elliott[17] and Gough.[18] The effect of asymmetry of gut strings is referred to by Gough.[14]

IX. CONCLUSIONS

Nonlinear effects resulting from varying wire tension and geometric factors, and also wire asymmetry play a big role in the behavior of a sinusoidally driven brass harpsichord wire (string). The wire asymmetry produces a splitting of the frequency of each normal mode of free vibration into two closely spaced frequencies. A characteristic transverse wire axis is associated with each of these frequency components. Nonlinear generation of transverse motion perpendicular to the driving force direction is strongly dependent upon the amount of splitting and the direction of the driving force with respect to the characteristic wire axes. Generation of perpendicular motion by *linear processes* can, under some conditions, be misinterpreted as *nonlinear generation*. This probably invalidates the interpretation of Figs. 6–8 of the earlier report[8] for very weak driving forces and small amplitudes. Nonlinear generation of very large second harmonics of the same order of magnitude as the fundamental are observed in many situations. For large driving currents significant third and fourth harmonics can be generated. Under seemingly unchanging conditions, there can occur abrupt large changes in the harmonic content of the motion. These may involve the appearance of large subharmonics and harmonics thereof. Such changes can also be induced by changing the driving frequency moderately rapidly. As the driving frequency is slowly increased, a transition from a stable state of motion can be initiated. With the driving frequency then held fixed, the motion can slowly evolve to a different stable state. Another possibility is quasiperiodic behavior in which the pattern of motion rapidly executed at the driving frequency changes very slowly in an apparently periodic manner. Some of the most interesting and unusual phenomena occur when driving along the *high-frequency axis*. Further measurements and theoretical modeling of the dependence on the driving frequency of the sudden onset of perpendicular motion at the driving frequency for different driving forces near the minimum required force would be instructive. Investigation of frequency splitting and its dependence on twist and tension for a greater variety of wire samples is warranted. Also, further investigations of the significance of nonlinear and asymmetric characteristics for plucked and struck stringed instruments would be useful.

ACKNOWLEDGMENTS

Undergraduate students Cary Pint, Mike Kassakatis, and Cheryl Claassen provided invaluable assistance in data acquisition and analysis. Takashi Nakamura, also an undergraduate student, provided the initial suggestion for and preliminary versions of the PDA optical detector. Physics Technician Larry Dirkes refined the PDA system and constructed other critical equipment. David Behroozi assisted in video recording. Michael Asbury made some helpful contributions. Charles Sengstock captured still frames from the movie of unusual motions produced by Joseph Marchesani and Dennis Reimer from experimental video footage. Edward Kottick provided the harpsichord wires studied as well as valuable consultation about their musical use. Consultation with Colin Gough has been very helpful. The University of Northern Iowa Graduate College provided essential financial support for student assistance.

[1] C. Valette, ''The mechanics of vibrating strings,'' in *Mechanics of Musical Instruments*, edited by A. Hirschberg, J. Kergomard, and G. Weinreich (Springer, Wien, 1995), pp. 115–183.
[2] C. Valette and C. Cuesta, *Mécanique de la Corde Vibrante (Mechanics of the vibrating string)* (Hermès, Paris, 1993).
[3] A. Watzky, ''Non-linear three-dimensional large-amplitude damped free vibration of a stiff elastic stretched string,'' J. Sound Vib. **153**, 125–142 (1992).
[4] A. Watzky, ''Sur la vibration non linéaire des fils précontraints (cordes, verges tendes)'' [''On the nonlinear vibrations of prestressed wires (strings, stretched rods)''], thèse de l'Université P. et M. Curie, Paris, 1992 (Ph.D. dissertation, University P. and M. Curie, Paris, 1992).
[5] T. C. Molteno and N. B. Tufillaro, ''An experimental investigation into the dynamics of a string,'' Am. J. Phys. **72**, 1157–1169 (2004).

[6] Reference 1, p. 160, Eqs. 151, 152, and 153 for which bending, twisting, damping, asymmetry, and external driving force are not included.
[7] Reference 3, p. 133, Eqs. 4.13a, 4.13b, 4.13c, and 4.14 for which damping, asymmetry, and external driving force are not included.
[8] R. J. Hanson, J. M. Anderson, and H. K. Macomber, ''Measurements of nonlinear effects in a driven vibrating wire,'' J. Acoust. Soc. Am. **96**, 1549–1556 (1994).
[9] Yellow brass wire 0.55 mm in diameter manufactured by Zuckerman Harpsichord, Stonington, CT.
[10] OI refers to the General Electric Photon Coupled Interrupter Module H21B1. See R. J. Hanson, ''Optoelectronic detection of string vibration,'' Phys. Teach. **25**, 165–166 (1987). Experimentation with different values of resistors is needed for optimum performance of different module samples.
[11] A photodiode array (PDA) of 128 photodiodes arranged linearly over a range of 2–3 mm is used to sense the horizontal transverse position of the wire. A light source above the wire casts a shadow on the detector mounted below, and a digital algorithm is used to determine the center of the shadow. The PDA is EG&G RETICON Catalog No. RL0128GAG connected to a RC0301LNN Amplifier Board. The algorithm and associated electronics were developed locally.
[12] M. Hancock, ''The dynamics of musical strings,'' Catgut Acoustical Society Journal **1**, 8 (Series II), 23–35 (1991).
[13] W. F. Vinen, ''The detection of single quanta circulation in liquid helium II,'' Proc. R. Soc. London, Ser. A **260**, 218–236 (1961).
[14] C. E. Gough, ''The theory of string resonances on musical instruments,'' Acustica **49**, 124–141 (1981). See especially p. 139.
[15] The driving frequency is above the low-frequency resonance so the motion due to the low resonance should be out of phase with the driving force. On the other hand, the driving frequency is below the high-frequency resonance so the motion due to the high resonance should be in phase with the driving force. Thus, the two motions differ in phase by 180° along the two characteristic axes.
[16] Private communication.
[17] J. A. Elliott, ''Intrinsic nonlinear effects in vibrating strings,'' Am. J. Phys. **48**, 478–480 (1980).
[18] C. E. Gough, ''The nonlinear free vibration of a damped elastic string,'' J. Acoust. Soc. Am. **75**, 1770–1776 (1984).

Identifying ultrasonic scattering sites from three-dimensional impedance maps

Jonathan Mamou and Michael L. Oelze
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

William D. O'Brien, Jr.[a)]
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801and Department of Bioengineering, University of Illinois, 1406 West Green Street, Urbana, Illinois 61801

James F. Zachary
Department of Bioengineering, University of Illinois, 1406 West Green Street, Urbana, Illinois 61801 and Department of Veterinary Pathobiology, University of Illinois, 2001 South Lincoln, Urbana, Illinois 61802

(Received 4 March 2004; revised 30 August 2004; accepted 4 September 2004)

Ultrasonic backscattered signals contain frequency-dependent information that is usually discarded to produce conventional *B*-mode images. It is hypothesized that parametrization of the quantitative ultrasound frequency-dependent information (i.e., estimating scatterer size and acoustic concentration) may be related to discrete scattering anatomic structures in tissues. Thus, an estimation technique is proposed to extract scatterer size and acoustic concentration from the power spectrum derived from a three-dimensional impedance map (3DZM) of a tissue volume. The 3DZM can be viewed as a computational phantom and is produced from a 3D histologic data set. The 3D histologic data set is constructed from tissue sections that have been appropriately stained to highlight specific tissue features. These tissue features are assigned acoustic impedance values to yield a 3DZM. From the power spectrum, scatterer size and acoustic concentration estimates were obtained by optimization. The 3DZM technique was validated by simulations that showed relative errors of less than 3% for all estimated parameters. Estimates using the 3DZM technique were obtained and compared against published ultrasonically derived estimates for two mammary tumors, a rat fibroadenoma and a 4T1 mouse mammary carcinoma. For both tumors, the relative difference between ultrasonic and 3DZM estimates was less than 10% for the average scatterer size. [DOI: 10.1121/1.1810191]

PACS numbers: 43.80.Qf, 43.20.Fn, 43.80.Vj [FD] Pages: 413–423

I. INTRODUCTION

Conventional *B*-mode images are derived from backscattered radio frequency (rf) echo signals. The rf echoes are created by reflections from interfaces between acoustically different regions (macrostructure) and by coherent and incoherent scattering from tissue microstructures. Those echo signals contain frequency-dependent information about the smaller scale tissue structures (<wavelength). *B*-mode image processing removes the frequency-dependent information available in the rf echo signals. *B*-mode images are good at displaying larger scale tissue structures (>wavelength) but, to display and quantify smaller scale structures, the frequency-dependent information must be utilized.

The rf echo signals backscattered from biological tissues contain information about the size, shape, number, and relative impedance (ratio of the acoustic impedance difference between the background and the scatterers to that of the background, the acoustic impedance is defined as the product of the density and the speed of propagation of sound) of the scattering objects. The backscattered signal is a superposition of wavelets scattered from numerous small structures confined within the volume of ensonified tissue. The frequency-dependent backscattered signal is dependent on the average tissue properties (size, shape, number, compressibility, density) of the scatterers within the ensonified region relative to the compressibility and density of the medium surrounding the scatterers.[1] The backscattered signal is, therefore, modeled as a statistical distribution of scatterers.

It is hypothesized that enhancing existing *B*-mode images with quantified physical properties of the average tissue microstructures can improve diagnosis of diseased tissue. Quantitative ultrasound (QUS) images, images enhanced by scatterer parameters like the average scatterer size and acoustic concentration (product of the number density and the square of the relative acoustic impedance of the scatterers), have successfully been used to characterize different aspects of tissue microstructures. Noteworthy are the pioneering works that have demonstrated theoretically and experimentally the ability to ultrasonically quantify ocular, liver, prostate, renal, and cardiac tissues.[2–5] Research revealed[6] that the effective scatterer size in ocular tumors was a strong indicator of cancer. Further, QUS results have provided greater diagnostic accuracy in prostate cancer detection and lesion

[a)]Author to whom correspondence should be addressed; electronic mail: wdo@uiuc.edu

localization than all other noninvasive techniques combined.[7,8] Anisotropy measurements of renal tissues showed that changes in the scatterer strength (acoustic concentration) were responsible for the anisotropy of backscatter rather than changes in scatterer size.[5] QUS imaging techniques were shown to be capable of differentiating among conditions that caused increased cortical echogenicity and structural changes like glomerular hypertrophy.[9–12] Furthermore, QUS measurements agreed well with measurements of those structures in biopsy samples.

Recently, we examined two different types of tumors to determine if it were feasible to distinguish between the tumors and normal tissues based on scatterer property estimates.[13–15] In the first study, QUS images of spontaneous rat mammary tumors [fibroadenomas (Harlan, Indianapolis, IN)] and surrounding tissues were constructed and compared. In the second study, a carcinoma cell line [4T1 mouse mammary carcinoma, CRL-2539 American Type Culture Collection (ATCC, Manassas, VA)] was cultured and implanted into mice. Estimates of scatterer size and acoustic concentration were made for the fibroadenoma and were compared with the carcinoma estimates. A statistically significant difference between scatterer diameter estimates inside and outside the fibroadenomas was observed for five of the eight rats [analysis of variance (ANOVA), $p<0.05$].[16] A statistically significant difference between acoustic concentration estimates inside and outside the fibroadenomas was seen for all but one rat (ANOVA, $p<0.05$).[16] A clear distinction was seen between the two tumor types and between the surrounding normal tissues using a multivariate classification scheme.[15]

Even though the QUS images (scatter size and acoustic concentration) have shown their effectiveness in differentiating masses in animal mammary tumor models, what is missing to more fully exploit the QUS approach is an understanding of the anatomical scattering sources. This understanding is hypothesized to lead to the identification of the actual anatomical scattering sites in tissues. Identification of scatterering sites will lead to improved scattering models. These improved models will yield more accurate and precise parameter estimation techniques from which it is also hypothesized to lead to QUS capabilities for diagnosing disease. Identification of the scattering sites is also important for estimating the optimal frequency range of interrogation for different kinds of tissues. A very similar approach has been shown to be successful in the case of trabecular bones,[17] where a binary 3DZM was used to describe bone microstructures.

To accomplish these challenges, this contribution introduces a novel approach to identify the anatomical scattering sources. The approach uses volume sections (3D histologic maps) corresponding to actual scanned tumor volumes to generate a 3D impedance map (3DZM). The 3DZMs are created by manually aligning serial photomicrographs of tumor sections. Impedance values are assigned to the different stained structures by the use of look-up tables of acoustical properties of the different tissue elements. The Fourier transform of the 3D spatial autocorrelation function (SAF) of the 3DZM is then calculated and used to estimate scatterer properties.

II. THEORY

A. Weak scattering in an inhomogeneous medium

The theory of scattering of a propagating acoustic wave in a heterogeneous medium is reviewed.[18] Weak scattering is the case where the inhomogeneities that cause scattering have tissue property values (density, ρ, and compressibility, κ) very close to those of the rest of the medium. For an incident plane wave of unit amplitude, in the case of the Born approximation (multiple scattering is ignored), the backscattered pressure wave far from the interrogated (or scattering) volume is a spherical wave:

$$p_{\mathrm{bs}}=\frac{e^{-ikr}}{r}\Phi(2k), \tag{1}$$

where bs denotes backscattered, k is the propagation constant ($k=\omega/c$ where ω is the angular frequency and c is the propagation speed) and the angle distribution function, $\Phi(2k)$ is

$$\Phi(2k)=\frac{k^2}{4\pi}\int\int\int_{V_0}\gamma(r_0)e^{-2ikr_0}dv_0. \tag{2}$$

V_0 represents the scattering volume and the function $\gamma(r)$ is

$$\gamma(r)=\gamma_\kappa(r)-\gamma_\rho(r), \tag{3}$$

where

$$\gamma_\kappa(r)=\frac{\kappa(r)-\kappa_0}{\kappa_0} \tag{4}$$

and

$$\gamma_\rho(r)=\frac{\rho(r)-\rho_0}{\rho(r)}, \tag{5}$$

where $\rho(r)$ and $\kappa(r)$ are the density and compressibility of the scattering sites, and ρ_0 and κ_0 are the density and the compressibility of the material surrounding the scattering sites, respectively. $\gamma_\kappa(r)$ and $\gamma_\rho(r)$ are therefore the relative changes in compressibility and density in the scattering volume V_0, respectively.

The backscattered intensity is hence

$$I_{\mathrm{bs}}=Ak^4|\Phi(2k)|^2, \tag{6}$$

where A is a proportionality constant.

In the case of weak scattering, that is, when γ_κ and γ_ρ are very small (for example, <0.1 for this first-order Taylor series approximation to be accurate), the function $\gamma(r)$ approximates to[19]

$$\gamma(r)=-2\frac{z(r)-z_0}{z(r)}=-2\frac{\Delta z}{z}, \tag{7}$$

where z is the plane wave impedance $z=\rho c=\sqrt{\rho/\kappa}$, $z(r)$ is the acoustical impedance value at location r, and z_0 is the background impedance, that is, the impedance of the medium with no scatterers. Thus, the function $\gamma(r)$ is proportional to the relative change in acoustic impedance.

Combining Eqs. (2), (6), and (7) yields

$$I_{bs} = A' k^4 S(2k), \tag{8}$$

where

$$S(2k) = \left| \frac{\iiint_{V_0} (\Delta z/z)\, e^{-2ikr_0} dv_0}{\iiint_{V_0} (\Delta z/z)\, dv_0} \right|^2 = \frac{S'(2k)}{S'(0)} \tag{9}$$

and where A' is a new proportionality constant, and $S(2k)$ is normalized such that $S(0)=1$

The backscattered intensity divided by k^4 is proportional to the magnitude squared of the Fourier transform of the relative impedance function, which is equal to the Fourier transform of the SAF of the relative impedance function of the scattering volume.[20] $S'(2k)$ and $S(2k)$, defined by Eq. (9), are hence termed power spectrum and normalized power spectrum of the medium, respectively.

B. Form factor

Form factors (FFs) are functions that describe the amplitude of the backscattered intensity due to a single scattering structure as a function of frequency, or more commonly as a function of the wave vector amplitude k. The FFs are based on 3D spatial correlation models by assuming some form or shape for the scattering tissue structures. Usually simple scattering shapes are assumed and in most cases they have a spherical symmetry. Mathematically, FFs are related to the shape of the scatterer through the Fourier transform. Specifically, the FF is the Fourier transform of the 3D SAF of a 3D medium containing a single scatterer, that is, the magnitude squared of the Fourier transform of the single scatterer's shape. FFs are normalized to a value of 1 when $k=0$ and their derivative usually vanishes when $k=0$.[19] They are therefore readily comparable to the normalized power spectrum $S(2k)$ [Eq. (9)].

Commonly used FFs include the Gaussian, the fluid sphere, and the spherical shell. They all describe spherically symmetric scatterers and depend only on the radius a of the scattering structure. For the fluid sphere FF the scatterer is assumed to be a homogeneous sphere filled with a fluid. For the spherical shell FF the scatterer is assumed to be a solid shell and therefore no wave is propagated inside the sphere. The Gaussian FF is a continuous distribution with spherical symmetry of relative impedance between the scatterer and the surrounding medium. The common FF definitions are[1,19]

$$F_a^1(2k) = e^{-0.827k^2 a_{\text{eff}}^2} \quad \text{(Gaussian)}, \tag{10}$$

$$F_a^2(2k) = \left[\frac{j_1(2ka)}{(2/3)\, ka} \right]^2 \quad \text{(fluid sphere)}, \tag{11}$$

$$F_a^3(2k) = [j_0(2ka)]^2 \quad \text{(spherical shell)}. \tag{12}$$

The subscript eff in Eq. (10) denotes effective because the radius does not correspond to a sharp discontinuity for a continuous distribution; a_{eff} is related to the correlation distance, d, by $d=(3\sqrt{\pi/2})^{1/3} a_{\text{eff}} = 1.55 a_{\text{eff}}$.[19] The subscript eff is omitted in the remainder of the paper. Also, the term ''scatterer size'' will always refer to the diameter of the scatterer (i.e., $2a$ or $2a_{\text{eff}}$ for the Gaussian model). j_0 and j_1 are the spherical Bessel function of the first kind of order 0 and 1, respectively.

The Gaussian FF [Eq. (10)] has often been used to model scattering from soft tissues.[16,21,3] The Gaussian FF allows for efficient estimation schemes to be utilized when estimating scatterer properties, i.e., the scatterer size and acoustic concentration. Goodness-of-fit has been used to determine the best FF for describing scattering; however, the anatomical scattering sources are still undetermined. There may exist undiscovered FFs that better model scattering than those FFs commonly used. Faran has derived closed-form expressions for the scattering from cylinders and spheres[22] and Insana has compared common FFs [Eqs. (10)–(12)] with the results of the Faran theory for spheres.[1] Their results have shown variable agreement between closed-form results and the common FFs for different types of spherical scatterers. The variability is extremely dependent upon the acoustical properties of the considered spherical scattering structure being considered. Form factor models do not take into account the presence of shear waves in the scatterers, whereas the theory of Faran incorporates shear. Therefore, the FF models are expected to work well for soft tissue scattering that does not support significant shear wave.

III. METHODS

A. 3D impedance map

A 3D impedance map can be viewed as a computational phantom. It is essentially a 3D matrix that describes some of the acoustical properties of a tissue volume; the elements are the values of the acoustic impedance of the medium.

3DZMs are derived from a 3D histologic data set, and are independent of ultrasonically acquired and/or processed data. To produce a 3DZM, tissue is fixed in 10% neutral-buffered formalin, embedded in paraffin, sectioned, mounted on glass slides, and stained with H&E (Hematoxylin and Eosin stain). Hematoxylin stains negatively charged nucleic acids (chromatin in nuclei and ribosomes) blue. Eosin stains proteins pink such as cell cytoplasm, connective tissue, muscle, etc. Therefore, all cell structures with proteins stain pink and the more protein the darker the pink color.

Each stained section is photographed with a light microscope [Nikon (Nikon Corporation, Tokyo, Japan) Optiphot-2 optical microscope], the photographs digitized with a Sony (Sony Corporation, Tokyo, Japan) CCD-Iris/RGB color video camera as a bitmap image, and each bitmap image aligned (registered) to form the 3D histologic map. Structures of approximately 0.5 μm or larger can be resolved, therefore individual proteins and chromatin molecules cannot be resolved, but the organization of these molecules into tissue structure such as nuclei, chromatids, connective tissue, or muscle fibers allows them to be seen. Then, each bitmap image pixel value in the 3D histologic map is converted to an appropriate acoustic impedance value to form the 3DZM.

To convert each pixel value to an appropriate acoustic impedance value a color-threshold algorithm was used on every H&E stained bitmap image. The range of colors of the different structures in the bitmap images were recorded. Each structure was assigned an impedance value (Table I).

TABLE I. Impedance values used for the 3D impedance maps.

Rat fibroadenoma (Mrayl)	4T1 Mouse mammary carcinoma (Mrayl)
1.58—epithelial cells	1.45—fat
1.54—mammary duct	1.58—cytoplasm
	1.60—nuclei
	1.60—red blood cells
1.80—connective tissue (z_0)	1.80—connective tissue
	1.55—background (z_0)

The bitmap images represent a two-dimensional (2D) matrix with colors represented by three 8 bit values corresponding to red, green, and blue. The ranges of color values corresponding to the different structures were used to project the bitmap into a 2D matrix where certain color ranges, corresponding to identified structures, were given the value corresponding to its assigned impedance value from the table. Sometimes pixels in the bitmap images were of colors that did not correspond to identified color ranges. In the case of pixels outside the identified color ranges, nearest neighbor values were used to assign impedance values to the unidentified pixels.

Tissue proteins stained with eosin have shades of color ranging from very light pink to dark pink, depending on the concentration of protein within the structure. This differential staining allowed for the recognition of unique microstructures in the tissue. A range of colors corresponding to an identified microstructure in the stained bitmap images was then assigned a single color. The new single color was thus representative of the acoustic impedance for that tissue microstructure.

For each pixel of the same color in the color-thresholded bitmap image the same acoustic impedance value was assigned. Figure 1 shows an example of an original H&E stained photomicrograph of a tissue section and the color-thresholded bitmap from a rat fibroadenoma. The stained sections were manually registered. The number of tissue sections was kept at a reasonable number (66 maximum). However, increasing the number of tissue sections would result in the need for a registration algorithm to automatically align the sections because manually aligning the sections would become too labor intensive.

Two different mammary tumors were used in our analysis. The impedance values used for the different tissue microstructures contained in the tissue were somewhat empirical, but also based on a large body of measured data.[23–25] For most soft tissues, the propagation speeds and densities fall within a relatively narrow range (speeds: 1520–1580 m/s; densities: 980–1010 kg/m^3; impedances: 1.5–1.6 Mrayl), with whole blood (speed: 1580 m/s; density: 1040 kg/m^3; impedance: 1.6 Mrayl) at the high end of most soft tissues. For lipid-based tissues, both the speed and density are somewhat lower (speeds: 1400–1500 m/s; densities: 920–970 kg/m^3; impedances: 1.3–1.5 Mrayl). For the collagen-based tissues, both the speed and density are somewhat greater than most soft tissues (speeds: 1600–1700 m/s; densities: 1020–1100 kg/m^3; impedances: 1.6–1.9 Mrayl).

The first tumor was a rat fibroadenoma. Each of the 39 10-μm-thick tissue sections measured laterally 640 μm by 480 μm. Each bitmap image measured 800 pixels by 600 pixels. Thus, each pixel represents a square of size 0.8 μm by 0.8 μm. Figure 2(a) shows the 3D H&E stained histologic map. For this tumor, three impedance values were used (left column of Table I): 1.58 Mrayl [for the epithelial cells (in green in Fig. 2(a)], 1.54 Mrayl for the mammary duct (in red), and 1.8 Mrayl for the connective tissue (background impedance, z_0 in blue). The resulting 3DZM is displayed in Fig. 2(b).

The second tumor was a 4T1 mouse mammary carcinoma. Each of the 66 5-μm-thick tissue sections measured laterally 160 μm by 120 μm. Each bitmap image measured 800 pixels by 600 pixels. Thus, each pixel represents a square of size 0.2 μm by 0.2 μm. Figure 3(a) shows the 3D H&E stained histologic map.

For this tumor four distinct impedance values were used (right column of Table I): 1.45 Mrayl for the fat [white in Fig. 3(a)], 1.58 Mrayl for the cytoplasm (gray), 1.60 Mrayl for the nuclei (blue), 1.60 Mrayl for the red blood cells (red), and 1.8 Mrayl for the connective tissue (green). The resulting 3DZM is displayed in Fig. 3(b). Also, for this tumor the background impedance (z_0) was assumed to be 1.55 Mrayl.

All the 3D renderings [Figs. 2(a), (b), 3(a) and (b)] were obtained using the software package ETDIPS (www.cc.nih.gov/cip/software/etdips/; a multidimensional volume visualization and analysis software, co-developed by the US National Institutes of Health and the National University of Singapore).

B. Estimation technique

This section presents the methodology used to obtain estimates from the power spectrum of a 3DZM. First, the 3DZM is divided into smaller volumes called regions of interest (ROIs). By dividing the 3DZM into smaller volumes,

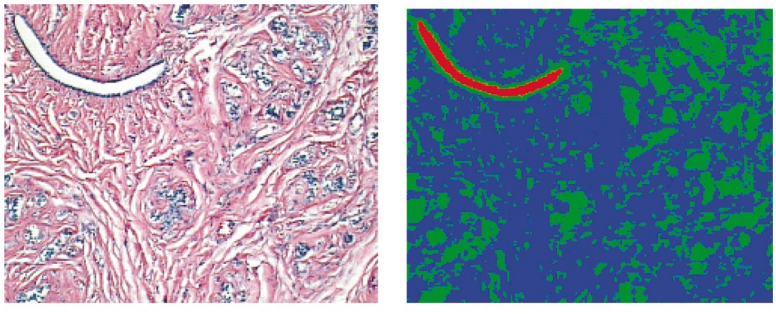


FIG. 1. The left image is a light microscopy illustration of the H&E-stained histology of a tissue section. The right image is the impedance map derived from this histologic section. Both images are 480 μm by 640 μm.

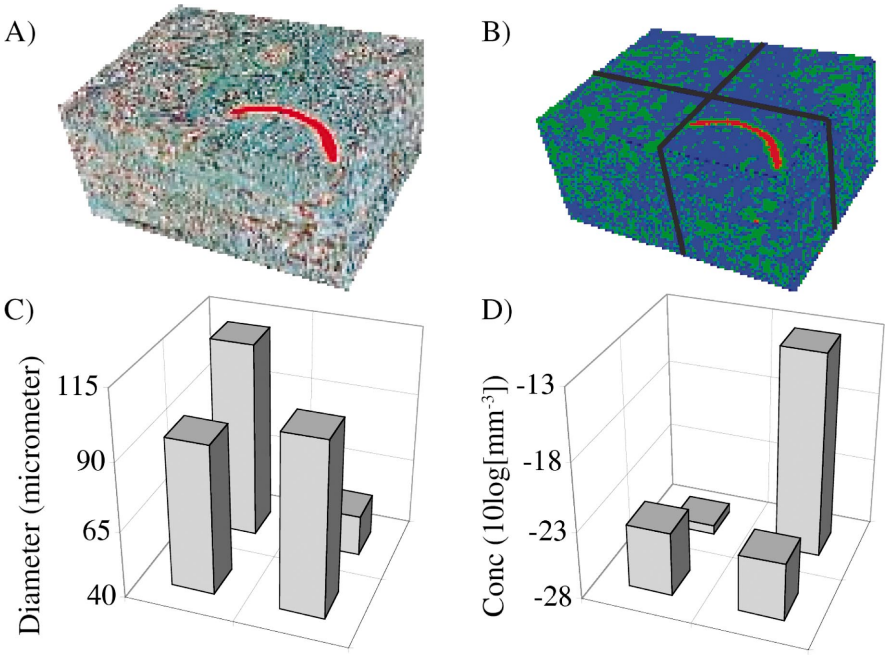


FIG. 2. (a) 3D H&E-stained histology of a fibroadenoma. (b) Derived 3D impedance map, black lines show the four ROIs. (c) Estimated scatterer diameter. (d) Estimated acoustic concentration. The volumes shown in (a) and (b) are of size 640 μm×480 μm×390 μm.

statistics about the estimates can be calculated from the different ROIs to show the precision of the estimation scheme. The scatterer size and acoustic concentration are then estimated for each ROI by an estimation routine that fits the Gaussian FF to the power spectrum. The Gaussian FF was used so that results obtained herein could be compared with published experimental ultrasound estimates[16] that used the Gaussian FF. However, the estimation technique could be used the same way with any FF. Estimates are obtained by minimizing the mean squared error between the log of the normalized power spectrum $[S(2k)$; Eq. (9)] and the log of the Gaussian FF [Eq. (10)] over the radius a, that is,

$$a^* = \operatorname{argmin}\left[\int_{k_{\min}}^{k_{\max}} [\log(S(2k)) - \log(F_a^1(2k))]^2 dk\right]. \tag{13}$$

The value of a that gives the smallest mean squared error is the scatterer radius estimate (a^*). Because the log of the Gaussian FF is $-0.827(ka)^2$, that is, a linear function of $(ka)^2$, the complexity of the minimization routine is reduced and allows for an efficient least squares solution.

However, it is not always possible to accurately obtain the normalized power spectrum $[S(2k)$; Eq. (9)] directly from the power spectrum $[S'(2k)$; Eq. (9)], because the

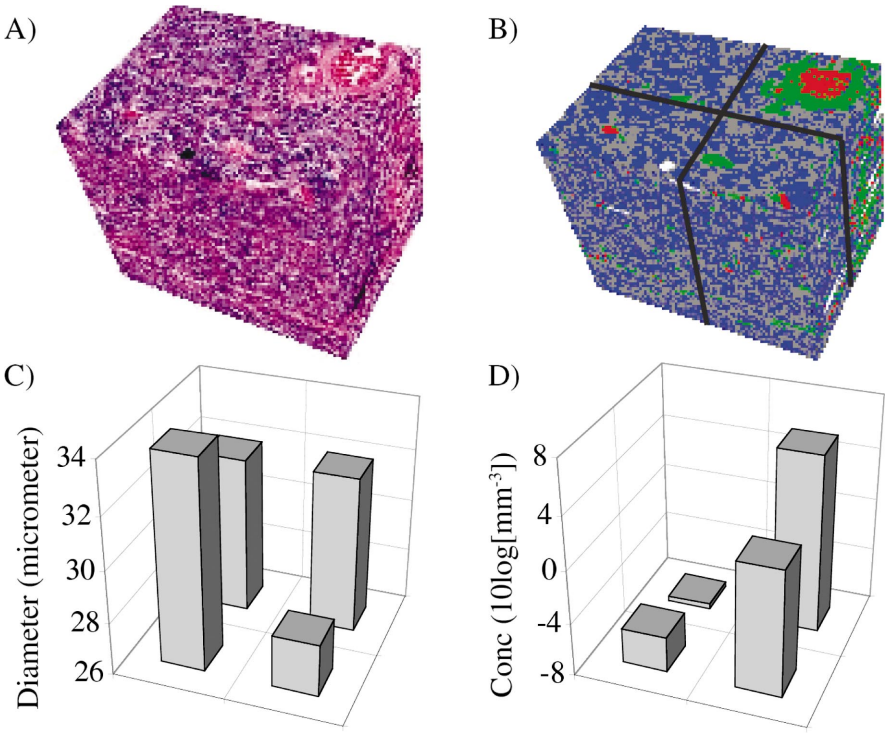


FIG. 3. (a) 3D H&E-stained histology of a 4T1 mouse mammary carcinoma. (b) Derived 3D impedance map, black lines show the four ROIs. (c) Estimated scatterer diameter. (d) Estimated acoustic concentration. The volumes shown in (a) and (b) are of size 160 μm×120 μm×330 μm.

computed $S'(0)$ might be very inaccurate. When $k=0$ the phase terms due to the random locations of scatterers add up instead of canceling each other as is the case when $k>0$. This fact makes the computed value of $S'(0)$ very different from its theoretical value [Eq. (9)]. Thus, Eq. (13) is slightly modified to include a gain parameter G_a, so that $S'(2k)$ can be used directly instead of $S(2k)$. The gain parameter G_a is obtained from

$$\log(G_a) = \sqrt{\frac{1}{k_{\max}-k_{\min}}\int_{k_{\min}}^{k_{\max}} |\log(S'(2k))-\log(F_a^1(2k))|^2 dk}. \tag{14}$$

Equation (14) is one of the choices to estimate a coefficient of proportionality between two functions. In particular in the idealistic case where S' and F_a^1 are proportional ($S'=\Gamma F_a^1$, where Γ is a proportionality constant), Eq. (14) yields $G_a=\Gamma$.

The gain parameter G_a is dependent upon a and can also be related to the acoustic concentration,

$$C=n\left[\frac{z-z_0}{z_0}\right]^2, \tag{15}$$

where n is the number density of scatterer, $(z-z_0)/z_0$ is the relative impedance mismatch between the scatterers and the background impedance, and z_0 is the background impedance. G_a and C are then related by

$$C=G_a/V_s^2, \tag{16}$$

where V_s is the volume of a single scatterer.[19] V_s is equal to $\frac{4}{3}\pi a^3$ for the fluid sphere and the spherical shell. V_s is also equal to $\frac{4}{3}\pi a^3$ for the Gaussian sphere.[1,19]

Incorporating the gain parameter G_a, Eq. (13) becomes

$$a^*=\text{argmin}\left[\int_{k_{\min}}^{k_{\max}}[\log(S'(2k))-\log(G_aF_a^1(2k))]^2 dk\right], \tag{17}$$

in which $S(2k)$ has also been replaced with $S'(2k)$. Equation (17) is now easier to solve because $S'(2k)$ is simple to compute from the 3DZM. $S'(2k)$ is the magnitude squared of the Fourier transform of the 3D impedance map. Also from a^*, we can compute G_{a^*} and therefore estimate the acoustic concentration C^* from Eq. (16).

In order to solve Eq. (17), it is necessary to properly select the optimization range, $k_{\min}$ and $k_{\max}$. An optimal ka^* range has been defined to be between 0.5 and 1.2 for the case of glass beads.[1] However, for the case of smoother FF (like the Gaussian FF) only the lower limit is the most important. For this work the ka range was chosen to be 0.5–2.0. For $ka^*<0.5$, FFs are usually flat (their derivative vanishes as ka^* approaches 0), which leads to large inaccuracies in the estimates: estimates are found to be close to the lower bound of the estimation range (usually below 3–5 μm). For $ka^*>2.0$, FF amplitudes are usually very low (e.g., -15 dB for the Gaussian FF) and noise can be dominant (backscattered signal amplitudes are already around -60 dB or less[1]). Table II shows the frequency ranges for scatterer of radii 10, 50, and 100 μm deduced from $ka=0.5$ and $ka=2.0$. A speed of sound of 1540 m/s was assumed.

TABLE II. Frequency ranges deduced from $ka=0.5$ and $ka=2.0$ ($c=1540$ m/s).

Scatterer radius (μm)	Frequency range (MHz)
10	12.3–49.0
50	2.45–9.80
100	1.23–4.90

The challenge for defining the optimal ka^* range is that it depends upon the estimated radius that is unknown. One approach could be to chose an average radius and then deduce from the optimal ka range values for $k_{\min}$ and $k_{\max}$. However, computational trials using this approach have shown a bias in the estimates: all estimates were found to be within a few percent of the average radius chosen.

An alternate approach was used to define $k_{\min}$ and $k_{\max}$. The goal of this approach was to obtain estimates that were not biased by the selected frequency range. Two new parameters k_{start} and Δk, were defined where $k_{\text{start}}=k_{\min}$ and $\Delta k=k_{\max}-k_{\min}$. An error function for the radius a is computed from

$$\text{Err}(a)=\text{Max}\{E(a,k_{\text{start}},\Delta k), \quad \text{for every } k_{\text{start}} \text{ and } \Delta k\}, \tag{18}$$

where

$$E(a,k_{\text{start}},\Delta k)=\frac{1}{\Delta k}\left[\int_{k_{\text{start}}}^{k_{\text{start}}+\Delta k}[\log(S'(2k)) - \log(G_aF_a^1(2k))]^2 dk\right], \tag{19}$$

and where the Err function is defined by the maximum of the mean squared errors over every frequency range. The optimal ka range is hence enforced by setting $E(a,k_{\text{start}},\Delta k)=0$ if $k_{\text{start}}a<0.5$ or $(k_{\text{start}}+\Delta k)a>2.0$.

Finally, a^* was defined as the argument of the absolute minimum of Err. This scheme has no built-in bias, because it evaluates every scatterer radius size over every frequency range while enforcing the optimal ka range. In particular, no specific frequency range has to be enforced. Thus, the strategy may possess the ability to find different populations of scatterers (different size, shape, impedance distribution, etc). Populations of scatterers may be revealed by determining the different local minima of the Err function.

An alternate choice of the Err function could be

$$\text{Err}'(a)=\text{Min}\{E(a,k_{\text{start}},\Delta k), \quad \text{for every } k_{\text{start}} \text{ and } \Delta k\}, \tag{20}$$

as replacement for Eq. (18) [also, enforcing optimal ka range would be accomplished by requiring $E(a,k_{\text{start}},\Delta k)\to\infty$ if $k_{\text{start}}a<0.5$ or $(k_{\text{start}}+\Delta k)a>2.0$]. However, applying Eq. (18) leads to better results because the Err$'$ functions curves are usually very flat and their minima are impossible to differentiate from noise. Err functions curves usually show steep minima that are clearly distinct from noise.

TABLE III. Simulation estimates.

	Medium A	Medium B
Diameter (% error)	41.1±2.9 μm (2.8%)	81.1±6.1 μm (1.4%)
Concentration (% error)	−14.28±1.1 dB (mm^{-3}) (2%)	−13.88±1.2 dB (mm^{-3}) (0.8%)

IV. RESULTS

Estimates for scatterer size and acoustic concentration were obtained for simulated 3DZMs, and from a rat fibroadenoma and a 4T1 mouse mammary carcinoma using the above-described estimation scheme. The simulated 3DZM studies are intended to assess the feasibility of the 3DZM approach for obtaining size and acoustic concentration estimates. In the simulated 3DZMs the quality of the estimates can be quantitatively assessed because the actual values are known.

A. Simulated media

1. Single population simulations

The first two simulated media were 3DZMs of the same size (256 μm by 256 μm by 256 μm) containing 15 fluid sphere-like scatterers of the same diameter. The scatterer diameters were 40 μm for medium A and 80 μm for medium B. The background had an impedance of 1.50 Mrayl and the spheres had an impedance of 1.51 Mrayl. The acoustic concentration was the same for both media (same number density of scatterer and same impedance difference between scatterer and background), that is,

$$C=\frac{15}{(256\times10^{-6})^3}\left[\frac{1.51-1.50}{1.50}\right]^2=0.0397\ \ \mathrm{mm}^{-3}.$$

Acoustic concentration values may have a large dynamic range from one ROI to another. Therefore, acoustic concentration is expressed as: $C_{\mathrm{dB}}=10\log(C)$ [and not $20\log(C)$, because C is directly related to the amplitude of the power spectrum and is hence a second-order quantity]. For the two simulated media $C_{\mathrm{dB}}=-14.01\,\mathrm{dB\,mm}^{-3}$. For the simulations, in Eq. (17), F_a^1 was replaced by the FF for a fluid sphere case F_a^2, and each voxel in the 3DZM was represented by a cube of side length 1 μm. Fluid spheres were used because the Gaussian model was more computationally intensive due to the infinite ''size'' of the Gaussian ''spheres.''

For both simulated media, the power spectra from 10 random realizations of the 3DZMs were averaged to limit the noise due to the spatial variations of the locations of the spheres. Fifty 10-realization spectra were computed for each media. From the 50 spectra, 50 size (diameter) and acoustic concentration estimates were obtained (Table III). For both media the average estimates were close to the actual values (relative errors less than 3% for all 4 average estimates). Log(Err) plots as a function of the scatterer diameter for realizations of media A and B showed that absolute minima were located very close to the actual scatterer diameters (Fig. 4), that is, 40.8 μm for medium A and 81.0 μm for medium B. The values of Err at the minima were also very close for media A and B curves [log(Err) values were around −1.8 for both curves]. Thus, the single population simulation studies validated the 3DZM methodology as a means to obtain accurate estimates.

Figure 4 shows that diameters between 20 and 160 μm were tested. Because the ka range chosen for optimization was 0.5–2.0, one can deduce that the optimization scheme evaluated any frequency range included in the interval (1.53–49.0 MHz).

Both curves (Fig. 4) showed sawtooth-like behavior for large scatterer diameters. Equation (18) should yield a smooth curve if an infinite number of frequency ranges were evaluated. However, because a finite number of frequency ranges were evaluated, the sawtooth-like behavior resulted. Also, the sawtooth-like behavior was more likely to occur for large scatterer diameters because fewer frequency ranges were available from which to select, and because enforcing the upper bound of the ka range limited the number of frequency ranges available for large a. This phenomenon could be removed artificially by low-pass filtering the error curves or by increasing the zero padding when computing the spectra so that more frequency ranges were available.

2. Two populations simulations

The next simulations evaluated the feasibility of identifying two population sizes of scatterers buried in a single medium. Seven media were simulated and the impedance mismatches were the same for both populations: 1.51 Mrayl for the spheres and 1.50 Mrayl for the background (Fig. 5). The acoustic concentration of the 40 μm scatterers was constant at $-14.01\,\mathrm{dB\,mm}^{-3}$ (same value as for media A and B). The acoustic concentrations of the 80 μm scatterers were (from top to bottom at the horizontal axis location of about 40 μm): −9.01, −14.01, −17.01, −20.01, −23.01, −26.01, and $-29.01\,\mathrm{dB\,mm}^{-3}$. These 7 media were the same size as

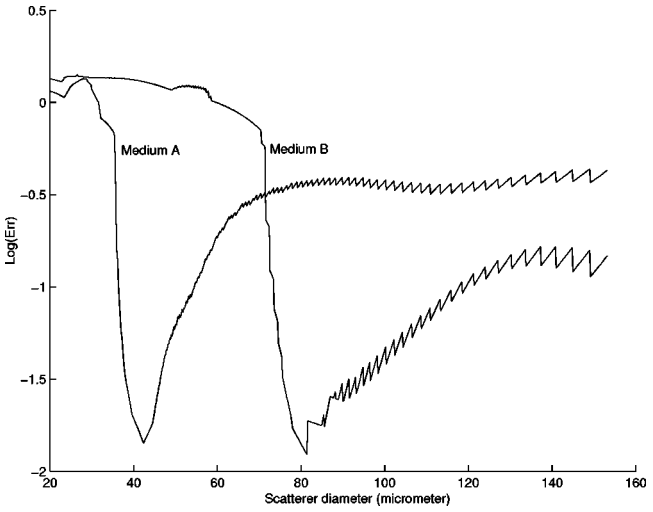


FIG. 4. Log(Err) vs scatterer diameter for realizations of media A and B.

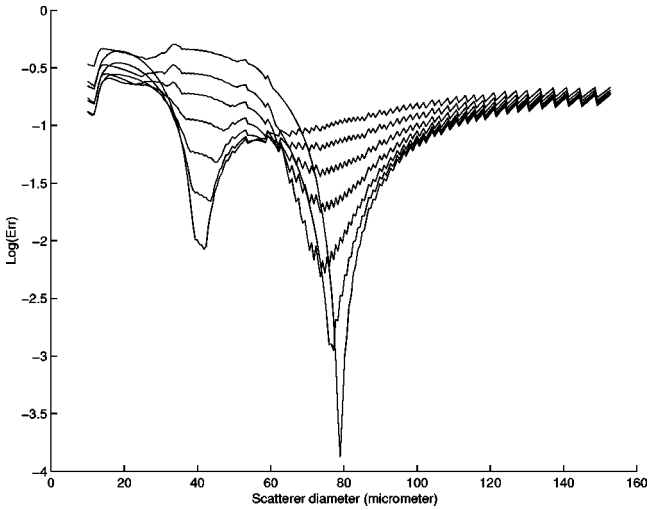


FIG. 5. Log(Err) vs scatterer diameter for realizations of two-population media. The acoustic concentration of the 40 μm scatterers was constant at -14.01 dB mm^{-3}. The acoustic concentrations of the 80 μm scatterers were (from top to bottom at the horizontal axis location of about 40 μm): -9.01, -14.01, -17.01, -20.01, -23.01, -26.01, and -29.01 dB mm^{-3}.

media A and B (cube with side length of 256 μm) and each voxel was represented by a cube of side length 1 μm.

As the acoustic concentration of the 80 μm scatterers increased, the topologies of the curves could be classified in three categories: (1) Single clear minimum around 40 μm (for 80 μm scatterer acoustic concentration of -29.01 and -26.01 dB mm^{-3}), (2) two clear minima around 40 μm and in the range 75–80 μm (for 80 μm scatterer acoustic concentration of -23.01 and -20.01 dB mm^{-3}), (3) single clear minimum near 80 μm (for 80 μm scatterer acoustic concentration of -17.01, -14.01, and -9.01 dB mm^{-3}). Hence, for the small scatterer population to be the only one detected [i.e., category (1)] a much higher acoustic concentration of small scatterers than large scatterers was necessary (a small scatterer acoustic concentration at least 12 dB mm^{-3} higher than that of the large scatterers). However, the large scatterer population is the only one detected [i.e., category (3)] when the small population acoustic concentration is at most 3 dB mm^{-3} higher than that of the large population. Finally, in between [i.e., category (2)], the curves tend to show two minima, one for each population.

A similar two-population study was also conducted where this time the number density was kept the same for the 40 and 80 μm populations. However, the impedance values of the 80 μm were varied. The impedance values chosen were such that the 80 μm acoustic concentration had the same seven values as in the previous simulations (-9.01, -14.01, -17.01, -20.01, -23.01, -26.01, and -29.01 dB mm^{-3}). Like before, the 40 μm acoustic concentration was kept at -14.01 dB mm^{-3}. The error curves obtained were extremely similar to those of Fig. 5.

These simulations showed that with a factor of 2 in size and under certain conditions (in particular concentration), it was possible to resolve the two populations using the 3DZM optimization scheme. However, it seems difficult to assess the resolution of the scheme based on this first set of two-population simulations.

The set of simulations showed that the 3DZM approach has the potential to resolve populations of different sized scatterers. Resolving different populations of scatterer size was possible only when the smaller sized scatterers' acoustic concentration was greater than that of the larger sized scatterers. The difficulty of resolving two scatterer size populations arises from the relative contribution of each population to the overall power spectrum. The power spectrum of a single population is proportional to Ca,[6] where a and C are the scatterer radius and acoustic concentration, respectively. Therefore, if the acoustic concentrations are equal, the contribution of the smaller sized scatterers to the overall magnitude of the power spectrum will be much less than the contribution of the larger sized scatterers because the spectrum is proportional to the radius to the sixth power.

These two 2-population studies demonstrated that the power spectrum is sensitive to the acoustic concentration, but not individually on the number density or the impedance values of the different scatterers. This is in accordance with ultrasonic scattering theory and was clearly pointed out by Insana.[1] Therefore, in order to resolve smaller scatterers in a two-population medium, either the number density of smaller scatterers must be much larger than that of the larger scatterers or the impedance mismatch for the smaller scatterers must be much larger than that for the larger scatterers.

B. Rat fibroadenoma results

The 3DZM of the rat fibroadenoma was divided into 4 ROIs, each of size 320 μm by 240 μm by 390 μm [Fig. 2(b)]. Figure 2(c) shows the scatterer size estimates and Fig. 2(d) shows the acoustic concentration estimates for the four ROIs. The mammary duct [in red in Fig. 2(b)] is present in two of the four ROIs, and appears in red in the bottom-right ROI. Also, the top-right ROI contains the mammary duct; however, it cannot be seen in Fig. 2(b) because it is not contained in the top section. The estimates obtained are very similar for the two ROIs that do not contain the mammary duct as well as the bottom-right ROI. However, the top-right size estimate is the smallest and the top-right concentration is the greatest. This might be due to the presence of the mammary duct. However, a similar effect was not observed in the estimates of the other ROI containing the mammary duct.

The error function (Fig. 6) contained a few distinct local minima (generally between one and three depending on the ROI). The estimates were obtained by selecting a minimum that was in the frequency range that was used for the same tumor when evaluated experimentally with ultrasound (-6 dB bandwidth 5–12 MHz).[16]

C. 4T1 mouse mammary carcinoma results

The 3DZM of the 4T1 mouse mammary carcinoma was divided into four ROIs, each of size 80 μm by 60 μm by 330 μm [Fig. 3(b)]. Figure 3(c) shows the scatterer size estimates and Fig. 3(d) shows the acoustic concentration estimates. The estimates are consistent among the four ROIs. However, the top-right ROI has the greatest acoustic concen-

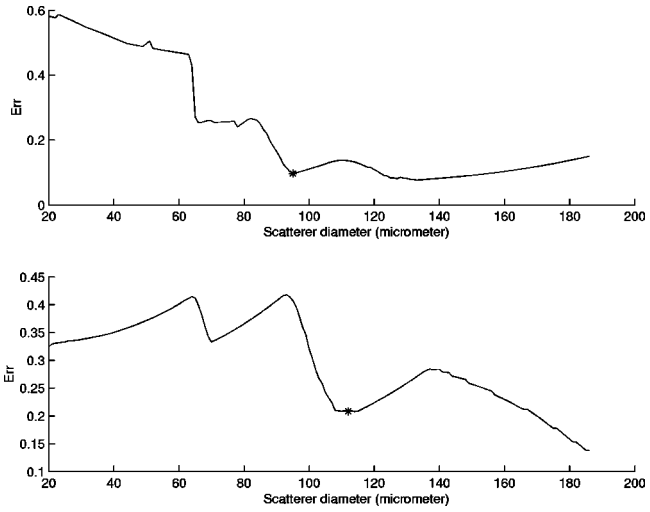


FIG. 6. Err function vs scatterer diameter obtained from two different ROIs of the rat fibroadenoma. Top curve was obtained from the top-left ROI of Fig. 2(b) and bottom curve was obtained from the top-right ROI of Fig. 2(b). The symbol "*" shows the selected minimum for each ROI.

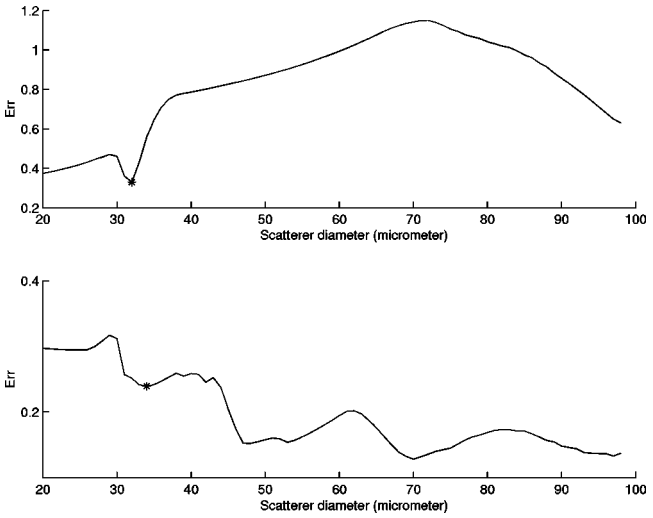


FIG. 7. Err function vs scatterer diameter obtained from two different ROIs of the 4T1 mouse mammary carcinoma. Top curve was obtained from the top-left ROI of Fig. 3(b) and the bottom curve was obtained from the bottom-left ROI of Fig. 3(b). The symbol "*" shows the selected minimum for each ROI.

tration (even though it does not have the smallest size estimate). This might be due to the presence of the red blood cells [in red in Fig. 3(b)] in this ROI.

Here also, the error function (Fig. 7) contained a few distinct local minima. The estimates were obtained by selecting a minimum that was in the frequency range that was used for the same tumor when evaluated experimentally with ultrasound (-6 dB bandwidth 10–25 MHz).[16] The top plot shows only one minimum, but the bottom plot shows four clear minima. However, for these two ROIs the selected size estimates were 32 and 34 μm, respectively.

D. Comparison with ultrasonic results

The estimates obtained using the 3DZM methodology for the two tumor types were compared to the published values obtained ultrasonically[14–16] (Table IV). The experimentally derived estimates represent a completely independent measure from that of the 3DZM estimates. The average scatterer size and acoustic concentration estimates (Table IV) were obtained by computing the mean and standard deviation of the estimates over the four ROIs for each of the two tumors. Results show a good agreement (difference less than 10%) for size estimates of both tumors. However, the acoustic concentration values are significantly different (difference greater than 6 dB mm^{-3}).

The 3DZM technique estimates of acoustic concentration were significantly smaller than those obtained ultrasonically (6.3 dB for the fibroadenoma and 11 dB for the 4T1 mouse mammary carcinoma). The difference in estimates might be due to several factors. First, more accurate impedance values may need to be assigned to a wider range of anatomic structures. In particular, a change in the background impedance value (z_0) will significantly modify the acoustic concentration estimates. Second, the internal distribution of acoustic properties (impedance) within the scattering structures will affect the estimates. Third, a small error in the estimate of the scatterer size can lead to a much larger error in the estimate of the acoustic concentration.[16] For example, a relative error of $+5\%$ of the scatterer size estimate will lead to a relative error of -27% of the acoustic concentration estimate.

It is encouraging that the scatterer size estimates obtained with the two independent techniques were directly comparable, because the scatterer size depends only on the variation of the power spectrum with frequency. However, the acoustic concentration depends upon the estimated scatterer size and the absolute amplitude values of the power spectrum.[16] The published acoustic concentration estimates determined ultrasonically were used only in a relative comparison. Absolute values for the acoustic concentration estimates were inaccurate because physical reflection losses (skin layer) were uncompensated. The physical reflection losses mostly change the absolute values of the power spectrum but not its variation with frequency. Therefore, acoustic concentration estimates found using the 3DZM may not be comparable to estimates obtained ultrasonically. The discrepancies between the ultrasonic and 3DZM acoustic

TABLE IV. 3D impedance map and ultrasound estimates.

	Rat fibroadenoma		4T1 mouse mammary carcinoma	
	Diameter (μm)	Concentration (dB mm^{-3})	Diameter (μm)	Concentration (dB mm^{-3})
Impedance map	91±25	−21.9±6.1	31.5±2.5	−1.4±6.1
Ultrasound	105±25	−15.6±5	30.0±9.6	10.6±6.9

concentration estimates need to be further investigated.

Some of the Err curves (Figs. 6 and 7) showed multiple local minima. The way the scatterer size estimate was selected among the multiple local minima was somewhat empirical. The minimum was selected in the frequency range that was used for the same tumor when evaluated ultrasonically and it is in part why the sizes agreed very well between the 3DZM and ultrasonic estimates. However, these Err curves (Figs. 6 and 7) all tended to have a local minima near the value found ultrasonically demonstrating agreement between the two independent techniques.

To assist with the understanding of the multiple local minima in the 3DZM-acquired curves, a next step would be to ultrasonically scan tissues over a wider range of frequencies, possibly with multiple transducers. Specifically, multiple minima may correspond to populations of scatterers with different sizes.

V. DISCUSSION

Identifying ultrasonic scattering sites would be a very beneficial adjunct to already existing QUS techniques. The work presented herein tackles this problem by using novel 3D representations of tissue microstructures (the 3DZM). In the present work, 3DZMs were used to obtain estimates of tissue microstructure parameters, namely the scatterer size and the acoustic concentration. The 3DZM estimation technique was shown to be very accurate for simulated 3DZMs (Table III). Estimates were also obtained for two mammary tumors and results were compared against published estimates that were obtained ultrasonically.[16] Average scatterer estimates for both techniques were within 10%. The results presented in this paper demonstrate that the 3DZMs may be accurate and useful representations of tissue microstructure. 3DZM approaches to ultrasonic scattering may hence have the potential to improve QUS techniques.

The methodologies presented herein characterize tissue microstructure from frequency-dependent information. The use of frequency-dependent scattering information from structures smaller than a wavelength has also been used to characterize different media using different energy sources. Laser scattering has been used to retrieve information about glass and polymer structures.[26] Neutron scattering has been used to measure molecular bond lengths.[27] Also, in acoustics, but in the low-frequency range (below 1 kHz), frequency-dependent scattering has been used to characterize turbules in the atmosphere.[28] Thus, ultrasonic scattering has been used[16,21,29] to quantitatively assess tissue microstructure. The 3DZM approach was developed to complement the ultrasonic estimation techniques, and further provide a computational tool for evaluating the anatomical scattering sites and sources.

The 3DZM approach is a powerful tool to obtain statistical descriptions of tissue microstructure. The 3DZM approach is a tissue-based capability to compute the power spectrum (and the spatial autocorrelation function, SAF) of the medium in a very straightforward way through the Fourier transform (see Sec. II A). To improve the power spectrum obtained from ultrasound, the ultrasonic techniques need to compensate for the experimental setting and the attenuation. Normalization removes dependence upon the experimental equipment[30] and attenuation compensation is used to mitigate the effects of attenuation.[31] If attenuation is not accounted for, estimates might become unreliable because attenuation is frequency dependent.

The power spectrum derived from ultrasonic measurements is only valid over the bandwidth of the transducer. Once the power spectrum is obtained, the optimization step in the ultrasonic technique is very similar to Eq. (13), except that $k_{\min}$ and $k_{\max}$ are fixed by the bandwidth of the transducer. The 3DZM approach has the advantage over the ultrasonic technique to obtain a power spectrum valid for a broader range of frequencies. The only limitation of the 3DZM approach is the absolute size of the tissue volume and the voxel size.

A couple of ways exist to improve the 3DZM technique. First, resolution could be improved. In particular, the section thickness could be reduced in order to have a similar pixel resolution in all three dimensions. Reducing the section thickness would also increase the similarities between two consecutive sections. Smaller slice thickness would have the advantage of being easier to align, or register, the sections.

Another improvement could be obtained by increasing the accuracy and the number of impedance values assigned to the different tissue structures. In some cases, it is difficult to measure acoustic properties of tissue microstructure, thus making the assignment of impedance values to different structures more challenging. However, an alternate technique might be used to deduce acoustic values by iteration. 3DZMs could be used as computational phantoms to simulate backscattered signals received by a given transducer. Then, comparison between simulation and ultrasonic experiments could be used to optimize the impedance values. At each step, the 3DZM would be updated until the simulated backscattered signals matched the signals obtained by ultrasonic measurements.

In the present work, the 3DZM methodologies were only conducted on tissues with lesions (i.e., tumors). Thus, the 3DZM methodology should also be tested on normal tissues to see whether the statistics of the 3DZM estimates would be different between healthy and diseased tissues. Estimates had different statistics on healthy and diseased tissues in the case of the ultrasonic techniques.[16]

The FF is, by definition, the normalized power spectrum of the 3DZM. Therefore, the 3DZMs can be used as a means to extract realistic FF through the Fourier transform. More interestingly, the 3DZM can also be used to help in identifying the anatomic scattering structures in tissues. From the inverse Fourier transform of the 3DZM-obtained FF, a 3D acoustic model of the SAF of a scattering structure can be determined. Then, it might be possible to identify the scattering sites by comparison of the SAF characteristics with the histology. Also, the assumed FF has great influence on the estimates. For example, the spherical shell FF has a steeper slope than that of the Gaussian FF.[1] Therefore, using the spherical shell FF would always lead to smaller estimates for the size and larger estimates for the acoustic concentration when compared to the estimates obtained with the Gaussian FF.

The spherical scattering assumption (or any nonspherical assumption) is still difficult to fully justify, because the scattering sites are unknown. It is our hope that the work provided in the manuscript can eventually lead to an understanding of the specific anatomic scattering sites which will then lead to a firm basis for the selection of a FF. Identifying the ultrasonic scattering sites would lead to great improvement of the diagnostic capabilities of quantitative ultrasound. Parameters could then be chosen that actually describe histology of tissue microstructures. Then, clinicians might be able to provide a diagnosis that is based on the ultrasonically obtained measurements of physical properties of tissue microstructure. Therefore, identifying the ultrasonic scattering sites may result in the possibility to develop novel and noninvasive diagnosis capabilities that are based on quantitative ultrasound approach.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the University of Illinois Research Board.

[1] M. F. Insana, R. F. Wagner, and D. G. Brown, ''Describing small-scale structure in random media using pulse-echo ultrasound,'' J. Acoust. Soc. Am. **87**, 179–192 (1990).
[2] J. E. Perez, J. G. Miller, B. Barzilai, S. Wickline, G. A. Mohr, K. Wear, Z. Vered, and B. E. Sobel, ''Quantitative characterization of myocardium with ultrasonic imaging,'' J. Nucl. Med. Allied Sci. **32**, 149–157 (1988).
[3] F. L. Lizzi, M. Greenebaum, E. J. Feleppa, M. Elbaum, and D. J. Coleman, ''Theoretical framework for spectrum analysis in ultrasonic tissue characterization,'' J. Acoust. Soc. Am. **73**, 1366–1373 (1983).
[4] F. L. Lizzi, M. Ostromogilsky, E. J. Felepa, M. C. Rorke, and M. M. Yaremko, ''Relationship of ultrasonic spectral parameters to features of tissue microstructure,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 319–329 (1986).
[5] M. F. Insana, J. G. Wood, and T. J. Hall, ''Identifying acoustic scattering sources in normal renal parenchyma in vivo by varying arterial and ureteral pressures,'' Ultrasound Med. Biol. **17**, 613–626 (1991).
[6] E. J. Feleppa, F. L. Lizzi, D. J. Coleman, and M. M. Yaremko, ''Diagnostics spectrum analysis in ophthalmology: A physical perspective,'' Ultrasound Med. Biol. **12**, 623–631 (1986).
[7] K. C. Balaji, W. R. Fair, E. J. Feleppa, C. R. Porter, H. Tsai, T. Liu, A. Kalisz, S. Urban, and J. Gillespie, ''Role of advanced 2 and 3-dimensional ultrasound for detecting prostate cancer,'' J. Urol. (Baltimore) **168**, 2422–2425 (2002).
[8] E. Feleppa, A. Kalisz, J. Sokil-Melgar, F. L. Lizzi, L. Tian, A. L. Rosado, M. C. Shao, W. R. Fair, W. Yu, M. S. Cookson, V. E. Reuter, and W. D. W. Heston, ''Typing prostate tissue by ultrasonic spectrum analysis,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 609–619 (1996).
[9] B. S. Garra, M. F. Insana, I. A. Sesterhenn, T. J. Hall, R. F. Wagner, C. Rotellar, J. Winchester, and R. K. Zeman, ''Quantitative ultrasonic detection of parenchymal structural change in diffuse renal disease,'' Invest. Radiol. **29**, 134–140 (1994).
[10] M. F. Insana and T. J. Hall, ''Parametric ultrasound imaging from backscatter coefficient measurements: Image formation and interpretation,'' Ultrason. Imaging **12**, 245–267 (1990).
[11] T. J. Hall, M. F. Insana, L. A. Harrison, and G. G. Cox, ''Ultrasonic measurement of glomerular diameters in normal adult humans,'' Ultrasound Med. Biol. **22**, 987–997 (1996).
[12] M. F. Insana, T. J. Hall, J. G. Wood, and Z. Y. Yan, ''Renal ultrasound using parametric imaging technique to detect changes in microstructure and function,'' Invest. Radiol. **28**, 720–725 (1993).
[13] M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., ''Parametric imaging of solid tumors using average scatterer size and acoustic concentration,'' 2003 World Congress on Ultrasound.
[14] M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., ''Differentiation of tumors types in vivo by scatterer property estimates and parametric images using ultrasound backscatter,'' Proceedings of the 2003 IEEE Ultrasonics Symposium, pp. 1022–1025.
[15] M. L. Oelze, W. D. O'Brien, Jr., and J. F. Zachary, ''Differentiation and characterization of rat mammary fibroadenomas and 4t1 mouse carcinomas using quantitative ultrasound imaging,'' IEEE Trans. Med. Imaging **23**, 764–771 (2004).
[16] M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., ''Parametric imaging of rat mammary tumors in vivo for the purposes of tissue characterization,'' J. Ultrasound Med. **21**, 1201–1210 (2002).
[17] F. Padilla, F. Peyrin, and P. Laugier, ''Prediction of backscatter coefficient in trabecular bones using a numerical model of three-dimensional microstructure,'' J. Acoust. Soc. Am. **113**, 1122–1129 (2003).
[18] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[19] K. K. Shung and G. A. Thieme, *Ultrasonic Scattering in Biological Tissues* (CRC Press, Boca Raton, FL, 1993).
[20] A. V. Oppemheim, R. W. Schaffer, and J. R. Buck, *Discrete-time Signal Processing* (Prentice Hall, Upper Saddle River, NJ, 1999).
[21] M. L. Oelze, J. F. Zachary, and W. D. O'Brien, Jr., ''Characterization of tissue microstructure using ultrasonic backscatter: Theory and technique for optimization using a Gaussian form factor,'' J. Acoust. Soc. Am. **112**, 1202–1211 (2002).
[22] J. J. Faran, ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**, 405–418 (1951).
[23] S. A. Goss, R. L. Johnston, and F. Dunn, ''Comprehensive compilation of empirical ultrasonic properties of mammalian tissues,'' J. Acoust. Soc. Am. **64**, 423–457 (1978).
[24] S. A. Goss, R. L. Johnston, and F. Dunn, ''Comprehensive compilation of empirical ultrasonic properties of mammalian tissues ii,'' J. Acoust. Soc. Am. **68**, 93–108 (1980).
[25] F. A. Duck, *Physical Properties of Tissue. A Comprehensive Reference Book* (Academic, New York, 1990).
[26] Y. Miyazaky, ''Light scattering of laser beams by random microinhomogeneities in glasses and polymer,'' Jpn. J. Appl. Phys. **13**, 1238–1248 (1974).
[27] P. A. Egelstaff, C. Gubbins, and K. C. Mo, ''Theory of inelastic neutron scattering from molecular fluids,'' J. Stat. Phys. **13**, 315–330 (1975).
[28] D. K. Wilson, J. G. Brasseur, and K. E. Gilbert, ''Acoustic scattering and the spectrum of atmospheric turbulence,'' J. Acoust. Soc. Am. **105**, 30–34 (1999).
[29] K. A. Topp, J. F. Zachary, and W. D. O'Brien, Jr., ''Quantifying b-mode images of in vivo rat mammary tumor with frequency dependence of backscatter,'' J. Ultrasound Med. **20**, 605–612 (2001).
[30] E. L. Madsen, M. F. Insana, and J. A. Zagzebski, ''Method of data reduction for accurate determination of acoustic backscatter coefficients,'' J. Acoust. Soc. Am. **76**, 913–923 (1984).
[31] M. L. Oelze and W. D. O'Brien, Jr., ''Frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media,'' J. Acoust. Soc. Am. **111**, 2308–2319 (2002).

Controlled ultrasound tissue erosion: The role of dynamic interaction between insonation and microbubble activity

Zhen Xu
Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

J. Brian Fowlkes
Department of Radiology and Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

Edward D. Rothman
Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109

Albert M. Levin
Department of Epidemiology, University of Michigan, Ann Arbor, Michigan 48109

Charles A. Cain
Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

(Received 25 June 2004; revised 3 October 2004; accepted 9 October 2004)

Previous studies showed that ultrasound can mechanically remove tissue in a localized, controlled manner. Moreover, enhanced acoustic backscatter is highly correlated with the erosion process. ''Initiation'' and ''extinction'' of this highly backscattering environment were studied in this paper. The relationship between initiation and erosion, variability of initiation and extinction, and effects of pulse intensity and gas saturation on time to initiation (initiation delay time) were investigated. A 788-kHz single-element transducer was used. Multiple pulses at a 3-cycle pulse duration and a 20-kHz pulse repetition frequency were applied. I_{SPPA} values between 1000 and 9000 W/cm^2 and gas saturation ranges of 24%–28%, 39%–49%, and 77%–81% were tested. Results show the following: (1) without initiation, erosion was never observed; (2) initiation and extinction of the highly backscattering environment were stochastic in nature and dependent on acoustic parameters; (3) initiation delay times were shorter with higher intensity and higher gas saturation (e.g., the mean initiation delay time was 66.9 s at I_{SPPA} of 4000 W/cm^2 and 3.6 ms at I_{SPPA} of 9000 W/cm^2); and (4) once initiated by high-intensity pulses, the highly backscattering environment and erosion can be sustained using a significantly lower intensity than that required to initiate the process. [DOI: 10.1121/1.1828551]

PACS numbers: 43.80.Gx, 43.80.Sh. [FD] Pages: 424–435

I. INTRODUCTION

Numerous studies have demonstrated the ability of producing a variety of biology effects using high-intensity pulsed ultrasound, e.g., creating cavities in the ventricular wall (Smith and Hynynen, 1998), lung hemorrhage (Child *et al.* 1990; Frizzell *et al.* 1994), intestinal hemorrhage (Dalecki *et al.* 1995), etc. Our previous studies showed that short, intense pulses (less than ten acoustic cycles) delivered at certain pulse repetition frequencies could achieve the most energy-efficient erosion and sharp, clearly defined perforations (Xu *et al.* 2004). This approach, which we call the controlled ultrasound tissue erosion (CUTE) technique, offers a potentially effective mechanism for the treatment of HLHS and other clinical problems where precise tissue removal is required.

This research was motivated by a need to noninvasively perforate the atrial septum of neonates with hypoplastic left heart syndrome (HLHS). HLHS is a congenital heart disease characterized by a dysfunctional left ventricle. It is usually fatal within 2 weeks after birth if not treated. A flow channel between the two atria (i.e., perforation of the atrial septum) must be created to extend patients' survival until reconstructive surgery is performed. Currently, cardiac catheterization and balloon atrial septostomy is the only option, which carries a combined morbidity and mortality rate as high as 50%.

This problem differs from most applications of therapeutic ultrasound in that ultrasonic exposures producing eventual tissue necrosis is not the desired outcome. Tissue must be physically removed during treatment so as to produce a perforation in the atrial septum sufficient for a flow channel between two atria. The immediate tissue removal required in this application cannot be accomplished by the production of thermal coagulation necrosis. This suggested the exploration of acoustic cavitation as a mechanism for the necessary controlled tissue removal.

It is highly likely that transient cavitation is the primary mechanism of the erosion process based on our previous results (Xu *et al.* 2004). Transient cavitation is characterized by the initially explosive growth of a cavity from a small, stabilized bubble or gas pocket, followed by an appropriately energetic collapse in response to the pressure variation or stress (Leighton, 1994). Such collapses are high-energy events and can generate a wide range of potentially destructive effects [e.g., erosion (Lush and Angell 1984; Belahadji *et al.*, 1991; Tomlinson and Matthews, 1994; Phillip and Ohl, 1995; Phillip and Lauterborn, 1998; Smith and Hynynen

1998), cell disruption (Jarman and Taylor, 1965; Dunn and Fry, 1971), and sonoluminescence (Ohl *et al.*, 1998; Ohl, 2000; Chakravarty and Walton, 2001)].

Acoustic backscatter of the insonating sound field is often used to monitor cavitation. Researchers have successfully demonstrated the sudden increase of backscattering signals from cavitating bubbles after insonifying with high-amplitude ultrasound pulses (Fairbank and Scully, 1977; Roy *et al.*, 1985; Atchley *et al.*, 1988; Holland and Apfel, 1990; Roy *et al.*, 1990a; 1990b; Madanshetty *et al.*, 1991; Bouakaz *et al.*, 1999; Shi *et al.*, 2000; Chen *et al.*, 2002). Variability of acoustic backscatter has also been used as a signature to detect cavitation (Kripfgans *et al.*, 2000). Further efforts have been made to quantify inertial cavitaton (IC) dose through acoustic backscatter (Calabrese, 1996; Everbach *et al.*, 1997; Poliachik *et al.*, 1999) and relate IC dose to cavitational bioeffects (Everbach *et al.*, 1997). In our experiments, we observed that an enhanced and rapidly changing acoustic backscatter was highly correlated with the erosion process. Accordingly, initiation and extinction of this highly backscattering environment were then investigated.

We hypothesize that the initiation of the highly backscattering environment we have seen is most likely due to generation of an ensemble of microbubbles (''bubble cloud'') which can be maintained by a train of appropriately designed acoustic pulses. Whenever subsequent pulses fail to maintain this bubble cloud, ''extinction'' occurs. This bubble cloud generated by acoustic pulses likely consists of a broad distribution of bubble sizes and temporal persistence. With an appropriately designed set of acoustic parameters, each pulse will affect the bubble population to produce an ensemble of effective cavitation nuclei for subsequent pulses and lead to more effective tissue erosion.

Matsumoto *et al.* have conducted research on modeling bubble cloud dynamics (Matsumoto *et al.*, 2002). Among others, Jöchle *et al.* have used optical scattering and high-speed camera to monitor the bubble cloud generated by a lithotripter in the water (Jochle *et al.*, 1996).

In this paper we describe our observations about the initiation and extinction of cavitation activity (possibly arising from a bubble cloud) when a tissue sample is irradiated with high-intensity, short-pulse acoustic fields. The tests detailed show how the occurrence of cavitation initiation, extinction, and the resulting cavitation-induced tissue erosion, depend on the acoustic field intensity, and the concentration of dissolved gas in the host water. We infer cavitation initiation and extinction based on the sudden appearance and disappearance of the enhanced, rapidly fluctuating, backscattered echoes received by a passive 5-MHz transducer. We have observed that without cavitation (as evidenced by the backscatter criteria for initiation) no erosion takes place.

We believe that initiation and maintenance of an appropriate set of cavitation nuclei (bubble clouds) are key to controlling the erosion process. As shown later in our results, without initiation, erosion was never observed. However, initiation is highly stochastic in nature (as is maintenance and extinction). We believe that all three processes depend strongly on acoustic parameters (e.g., pulse intensity, pulse duration, and pulse repetition frequency), and that the process can be optimized by simply changing acoustic pulse parameters. In this process, each pulse has at least two functions: (1) production of some fraction of the desired therapeutic effect, and (2) maintenance of the dynamic ensemble of caviation nuclei to optimally predispose the tissue for subsequent pulses. This approach can make the erosion process both controllable and predictable. By simple adjustment of acoustic parameters, the process might be optimized, even with feedback in real time, in order to achieve the most efficient erosion with the least possible propagated energy.

(Note that erosion, as referred to in this paper, is defined as obvious tissue removal that can be distinguished from the surrounding tissue.)

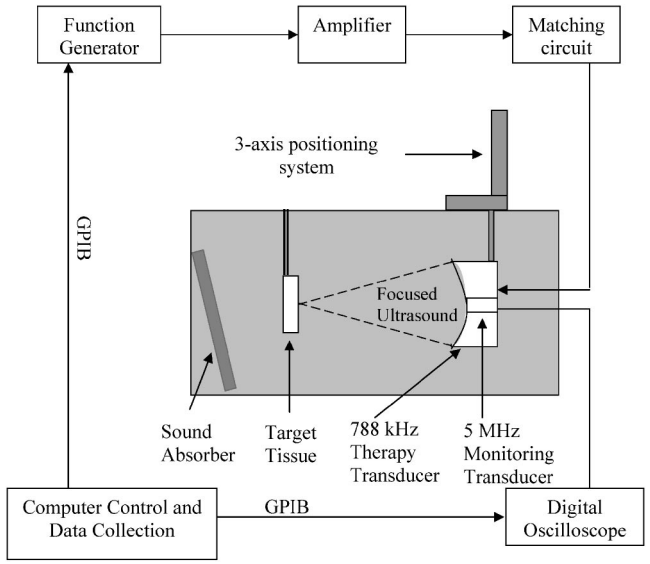


FIG. 1. Experimental setup. Ultrasound pulses were delivered by the 788-kHz therapy transducer. Porcine atrial wall sample was positioned at the focus of the therapy transducer. Acoustic backscatter was received by the 5-MHz monitoring transducer.

II. METHODS

A. Tissue samples

In vitro experiments were conducted on 33 porcine atrial wall samples. Porcine atrial wall was used because it is similar to the neonatal atrial septum and has a larger size. Fresh samples were obtained from a local slaughter house and used within 72 h of harvesting. All tissue specimens were preserved in a 0.9% sodium chloride solution at 4 °C prior to experimentation. A tissue holder was designed and built to

TABLE I. Exposure parameters (pulse duration=3 cycles).

I_{SPPA} (W/cm^2)	Peak compressional pressure (MPa)	Peak rarefractional pressure (MPa)
1000	7.8	5.2
2000	11.7	6.6
3000	15.2	7.5
3500	16.7	7.9
4000	18.3	8.3
5000	21.4	9.0
7000	27.3	10.1
9000	36	11.6

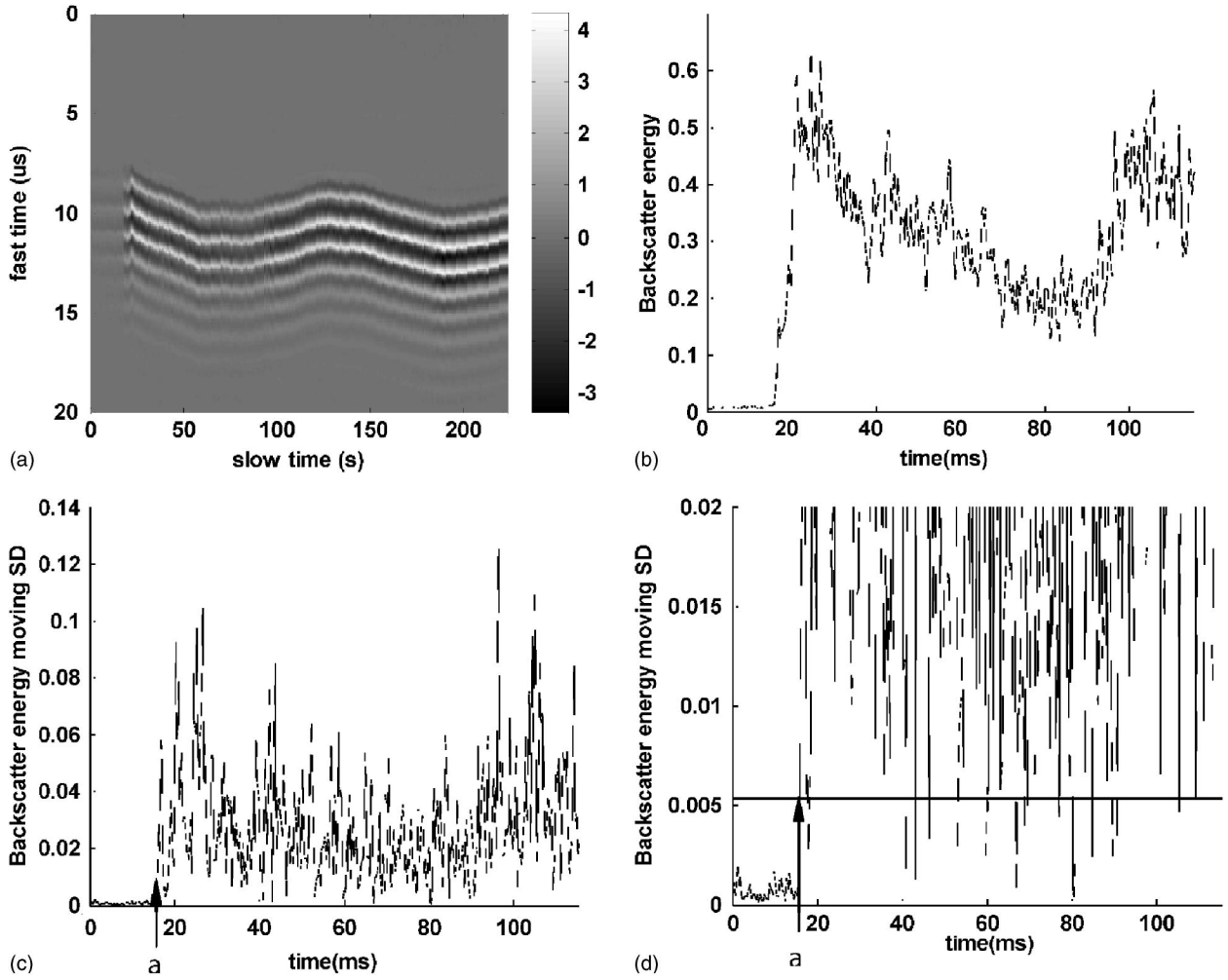


FIG. 2. Process to detect initiation of the highly backscattering environment. Panels (A), (B), (C), and (D) show the steps of initiation detection in sequence. Panel (A) shows the acoustic backscatter in fast time and slow time mode. Each vertical line shows backscatter recorded as the output voltage of the monitoring transducer in fast time. Panel (B) shows the backscatter energy versus time. Panel (C) shows the moving SD of backscatter energy versus time. Panel (D) is an expanded view of panel (C). The line is the initiation threshold, set by four times the SD estimation of uninitiated backscatter energy. In Panel (C) and (D), detected by the criteria defined for initiation, the highly backscattering environment was initiated at "a." Ultrasound pulses with a pulse duration (PD) of 3 cycles, a pulse repetition frequency (PRF) of 20 kHz, an I_{SPPA} of 5000 W/cm^2, and gas saturation of 46% were applied.

fix the tissue in the exposure water tank. The atrial septum of an HLHS patient with an intact atrial septum is usually stretched due to the pressure difference between the two atria (10 mmHg and above). The tissue was stretched under tension to mimic the *in vivo* condition.

B. Experimental setup

All ultrasound exposures occurred in a 61-cm-long×28-cm-wide×30.5-cm-high polycarbonate tank containing water degassed to a desired level prior to the experiment. The temperature in the tank was ~22 °C during the trials. A 788-kHz focused single-element therapy transducer (f number=1, Etalon Inc., Lebanon, IN) was mounted on a three-axis positioning system. The 788-kHz therapy transducer has an 8.8-cm aperture, an 8.8-cm focal length, and a 3.7-cm inner diameter hole for a monitoring transducer.

A focused single-element monitoring transducer with 5-MHz center frequency (f=4, Valpey Fisher Corp., Hopkinton, MA) was mounted coaxially with the therapy transducer to receive the backscatter from the erosion zone. The 5-MHz transducer has a 2.5-cm aperture and a 10-cm focal length. The 5-MHz passive transducer is used because (1) its focal length is 10 cm and it is smaller (2.54-cm diameter) than the inner center hole (3.7-cm diameter) of the primary transducer so that we can conveniently align it confocally with the primary transducer by fixing it in the center hole; and (2) it has a wide bandwidth (−6-dB bandwidth of 4 MHz) that can detect the higher harmonic components as well as the primary frequency component.

The tissue was placed at the focus of the therapy transducer. The focus of the transducer was first found in calibration by 3D scanning and maximizing the output of the membrane hydrophone. During the experiments, the focus of the therapy transducer was then determined based on the time delay of the tissue echo received by the monitoring transducer matching the focal length of the therapy transducer obtained from the previous calibration. An angled sound absorber (40 Durometer, Sorbothane, Inc., OH) was placed behind the tissue to decrease the sound reflection from the side of the tank, thereby reducing the formation of standing waves. Figure 1 shows the experimental setup.

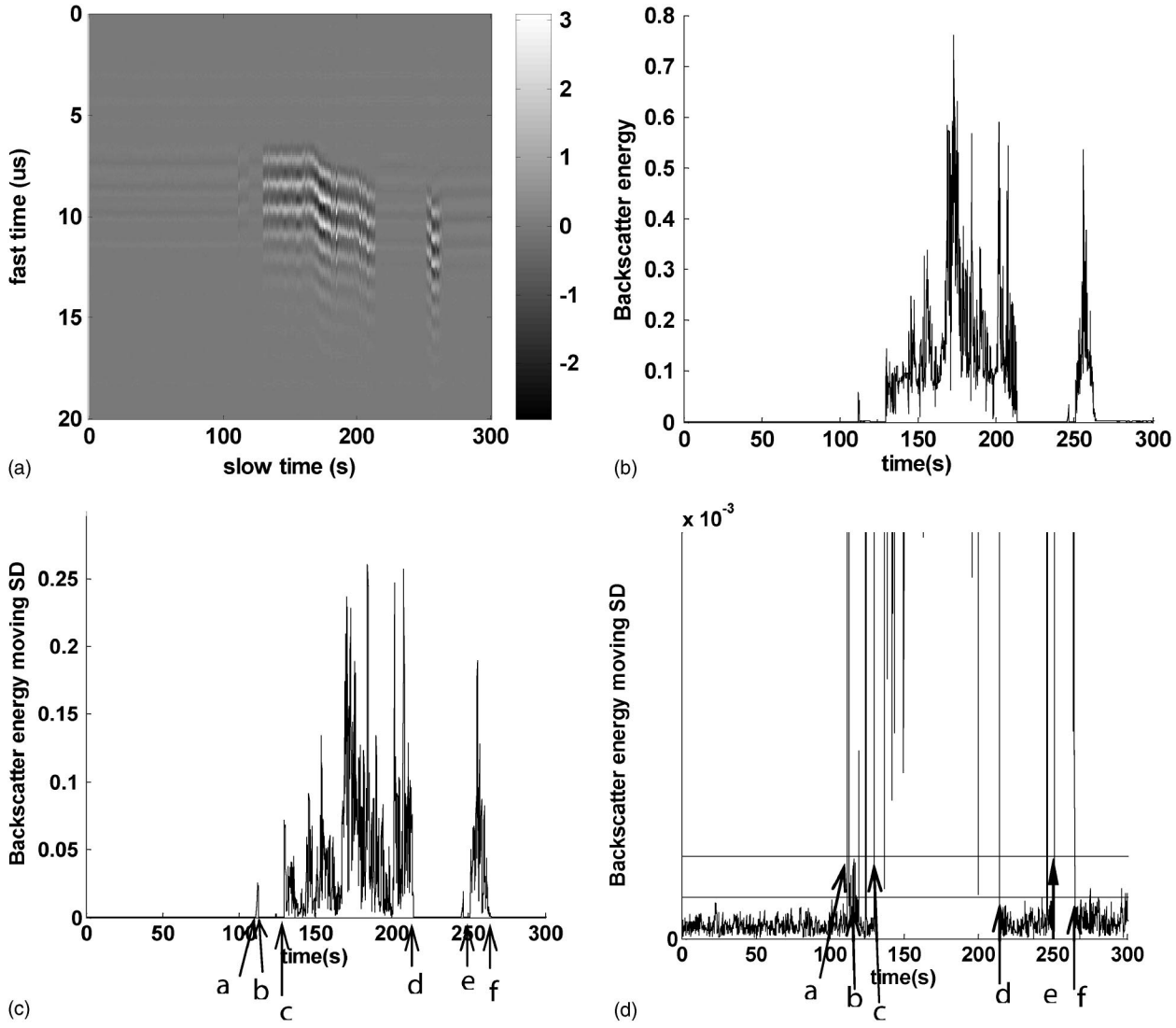


FIG. 3. Process to detect initiation and extinction. Panels (A), (B), (C), and (D) show the steps of initiation and extinction detection in sequence. Panel (A) shows the acoustic backscatter in fast time and slow time mode. Each vertical line shows backscatter recorded as the output voltage of the monitoring transducer in fast time. Panel (B) shows the backscatter energy versus time. Panel (C) shows the moving SD of backscatter energy versus time. Panel (D) is an expanded view of panel (C). The line above is the initiation threshold, set by four times the SD estimation of uninitiated backscatter energy. And, the line below is the extinction threshold, set by two times SD estimation of uninitiated backscatter energy. In panels (C) and (D), detected by the criteria defined for initiation and extinction, the highly backscattering environment was initiated at "a," extinguished at "b," spontaneously reinitiated at "c," extinguished again at "d," reinitiated again at "e," and tissue was finally perforated at "f." Ultrasound pulses with a PD of 3 cycles, a PRF of 20 kHz, an I_{SPPA} of 4000 W/cm^2, and gas saturation of 40% were applied.

C. Calibration

The spatial distribution of intensity of the 788-kHz transducer was measured in degassed water with a calibrated PVDF bilaminar shielded membrane hydrophone (GEC-Marconi Research Center, Chemlsford, UK). From transverse beam plots, a focal zone of 2.58-mm diameter is defined by the full width at half-maximum pressure points. The I_{SPPA} stated in this paper is the free-field spatial peak pulse average value calculated from the pulse integral and pulse duration as prescribed by the AIUM (AIUM, 1998). Peak rarefractional pressures and peak compressional pressures corresponding to I_{SPPA}'s between 1000–9000 W/cm^2 are listed in Table I.

D. Data collection

A total of 2000 20-μs-long waveforms was recorded using the sequence mode and single trigger of the digital oscilloscope (LeCroy 9354TM, Chestnut Ridge, NY). The oscilloscope trigger was synchronized with the therapy pulses, and the trigger time delay was adjusted such that the 20-μs-long backscattering signal was from the erosion zone. The interval between consecutive waveform recordings was set such that the whole initiation process could be recorded within the time span of the multiplication of the interval between consecutive recordings and 2000 (number of backscatter waveform collected). For example, with therapy pulses of PD 3 cycles at a PRF of 20 kHz, 2000 waveforms were recorded with a 200-μs interval between waveforms. The detected signals were digitized by the oscilloscope at a resolution of 40–100 ns. The recorded waveforms were then transferred to a computer through GPIB and processed by a MATLAB program (Mathworks, Natick, MA) to detect initiation based on criteria to be defined later. Then, the same procedures were repeated for extinction detection. But, the

interval between consecutive recordings was adjusted to 240 ms so that backscatter during the 8-min ultrasound treatment could be recorded.

E. Statistical approach to detect "initiation" and "extinction"

The statistical properties of cavitation have been investigated. In particular, the probabilistic nature of cavitation using different acoustic parameters has been studied (Messino *et al.*, 1963; Calabrese, 1996). Based on our experimental observations, the onset of initiation is accompanied by alterations in the backscatter signal, presumably associated with the onset of cavitation. One such change is a sudden increase in the backscatter amplitude at initiation. Further, this amplitude increase is followed by a chaotic fluctuation in the backscattering signal. Together, these two changes indicate an overall change in the variability of the signal as the transition is made between the uninitiated and initiated states. We developed a statistical method for the detection of initiation and extinction of the highly backscattering environment based on this change in variability.

To identify points of initiation and extinction based on variability in the backscattering signal, we applied a common technique from the area of statistical quality control of industrial processes, the Shewhart chart (Wetheril and Brown, 1991). Depending on the data, different Shewhart charts are used to identify changes in a time series process. For our particular situation, we used the *s* chart, where the sample standard deviation (SD) of the backscatter energy at point *i* in the time series is used as the measure of variability. However, for our data, a single measurement of the backscatter energy was made at each time point in a given experiment. For such "one-at-a-time" data, the SD at a single point cannot be directly estimated, and a moving SD approach is often employed.

In our situation, we used a moving window size of 3 to estimate the SD at each point *i* in the time series, SD_i. For example, the estimate of SD_i was calculated based on the backscatter energy at point *i* and the two points preceding it, $i-1$ and $i-2$. We define initiation to have occurred when five consecutive SD_i's exceed a threshold of four times the estimated SD of the uninitiated backscatter energy. We define extinction to have occurred when five consecutive SD_i's fall below a threshold of two times the SD of the uninitiated backscatter energy. The SD of the uninitiated backscatter energy was estimated from the first *n* frames of backscatter recorded prior to any high degree of variation in the signal potentially indicating initiation. The initiation threshold coefficient of 4 and extinction threshold of 2 are used as starting values and need further investigation.

The backscattering signal was the output voltage of the monitoring transducer. Backscatter energy was calculated by integrating the square of this voltage over each line in fast time

$$\text{Backscatter energy} = \frac{1}{N}\sum_{i=1}^{N} V^2(i), \tag{1}$$

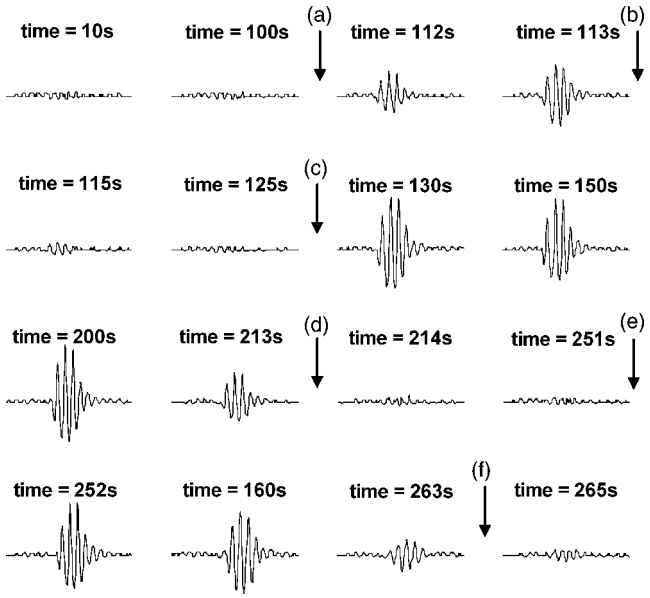


FIG. 4. Waveforms of acoustic backscatter corresponding to the data in Fig. 3. All the backscatter waveforms are 20-μs-long range gated from the erosion zone. "a"–"f" are the initiation and extinction points shown in Figs. 3(C) and (D). A highly backscattering environment was initiated at "a," extinguished at "b," spontaneously reinitiated at "c," extinguished again at "d," reinitiated again at "e," and tissue was finally perforated at "f." Ultrasound pulses with a PD of 3 cycles, a PRF of 20 kHz, an I_{SPPA} of 4000 W/cm^2, and gas saturation of 40% were applied.

where *N* is the number of points in one line of backscatter signal, and $V(i)$ is the voltage value of the *i*th point within this line of backscatter signal.

The statistical procedure for identifying initiation and extinction consists of the following steps:

(1) Step 1: the first $n(10 \leqslant n \leqslant 100)$ frames of backscatter prior to initiation were collected. Then, the SD of backscatter energy while uninitiated could be estimated based on the first uninitiated *n* points using the Shewhart charts [Eq. (2)] (Wetheril and Brown, 1991)

$$\text{Estimated standard deviation} = \frac{1}{1.102}\sum_{i=1}^{n-2} \text{range}(x_i, x_{i+1}, x_{i+2}), \tag{2}$$

$$\text{Range}(x_i, x_{i+1}, x_{i+2}) = \text{maximum}(x_i, x_{i+1}, x_{i+2}) - \text{minimum}(x_i, x_{i+1}, x_{i+2}). \tag{3}$$

The purpose of choosing a moving range of 3 points to estimate the SD of backscatter energy while uninitiated is to be consistent with the window size used to calculate the moving SD.

Step 2: the moving SD of backscatter energy is calculated. Then, initiation and extinction can be detected based on the two previously described criteria. Both criteria are programed in MATLAB (Mathworks, Natick, MA), so initiation and extinction can be detected automatically.

Figure 2 demonstrates the process of detecting initiation of a highly backscattering environment. Detection of extinction of a highly backscattering environment is demonstrated

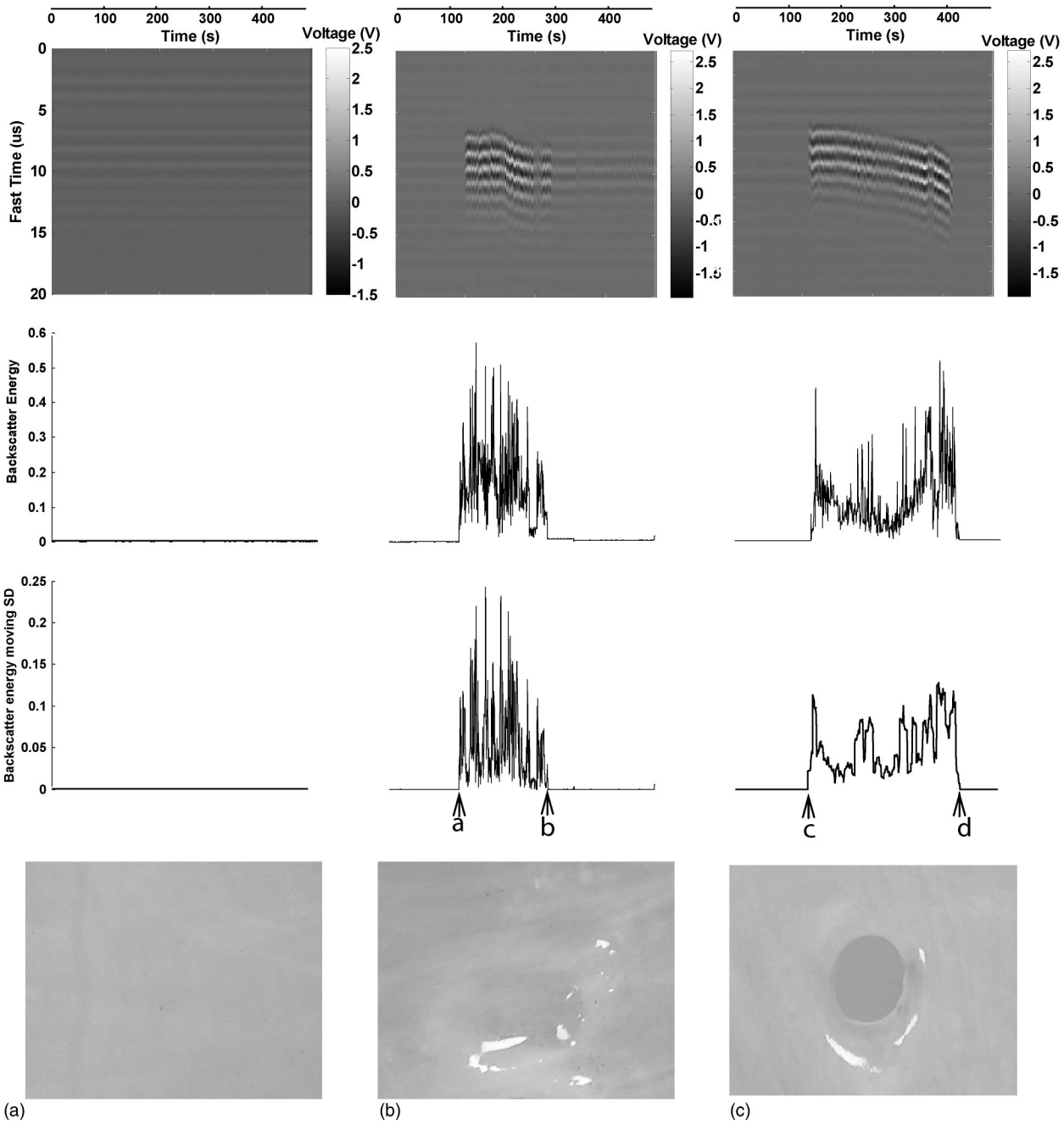


FIG. 5. The first row shows the acoustic backscatter in fast time and slow time mode. The second row shows the backscatter energy versus time. The third row shows the moving SD of backscatter energy versus time. The *x* axis (time) for each column is the same and shown above the column. The *y* axis for each row is the same and shown on the left side of each row. The fourth row depicts the tissue effects on porcine atrial wall tissue samples generated by the corresponding treatments. All the tissue samples were treated by a total of 8-min ultrasound pulses at an I_{SPPA} of 3500 W/cm^2, a PD of 3 cycles, a PRF of 20 kHz, and gas saturation of 40%–45%. In panel (A), neither initiation nor erosion was observed. In panel (B), initiation (''a'') and extinction (''b'') were detected and erosion was observed, but tissue was not perforated. In panel (C), initiation (''c'') was detected and erosion was observed, and tissue was perforated (''d'').

in figure 3. Figure 4 shows the actual waveforms of the acoustic backscatter before and after initiation and extinction. Figure 5 depicts the initiation and extinction phenomena and corresponding tissue effects generated.

(Please note that when tissue is perforated, the backscatter variability is greatly reduced and is detected as extinction based on the above criteria.)

F. Experiment design

The initiation and extinction processes and the relationship of initiation to erosion were studied through observation of the acoustic backscatter and the tissue effects generated by the corresponding ultrasound pulses. Moreover, the effects of intensity and gas saturation on initiation delay time were investigated. Initiation delay time here is defined as the interval between the onset of acoustic pulses and the first initiation (as previously defined) of a highly backscattering environment. Initiation delay time reported in this paper only includes the cases when an initiation was detected. The initiation and extinction processes were monitored by the acoustic backscattering signal received by the 5-MHz transducer and detected by the methods described earlier.

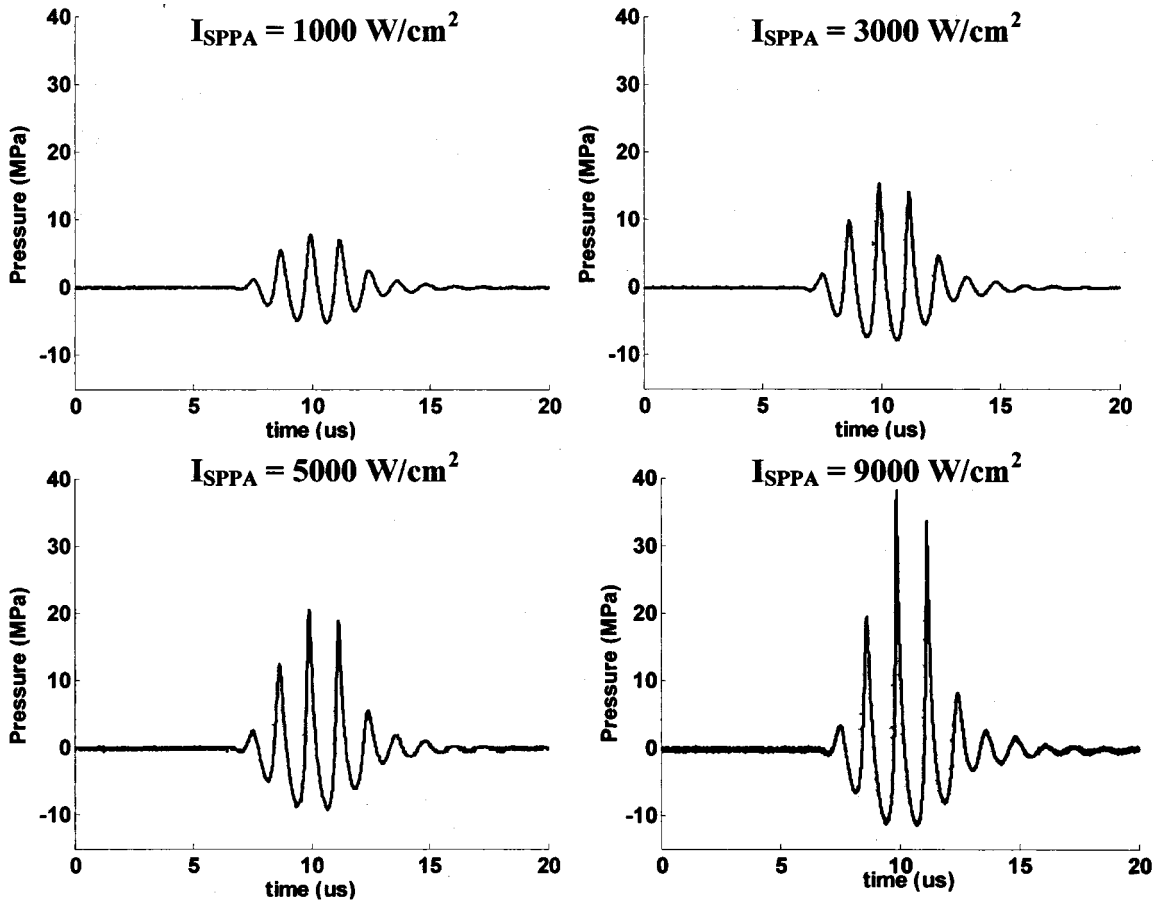


FIG. 6. The waveform of the therapeutic ultrasound signal with a PD of 3 cycles and I_{SPPA}'s of 1000, 3000, 5000, and 9000 W/cm^2 delivered by the 788-kHz therapy transducer as recorded by a membrane hydrophone. PD used in this paper is given as the number of cycles in the waveform at output of the function generator. However, the calculation of I_{SPPA} uses the appropriate definition of the PD.

A pulse duration (PD) of 3 cycles and a pulse repetition frequency (PRF) of 20 kHz were used in all the ultrasound exposures. This set of parameters was chosen because it achieved the fastest erosion in our previous studies (Xu *et al.*, 2004). Note that, except for the calculation of I_{SPPA}, PD is defined as the number of cycles in the waveform at the output of the function generator (Fig. 6). The parameters used in the following experiments were randomized.

For studies of the effects of intensity on the initiation delay time, gas saturation was set to 39%–49%. I_{SPPA}'s of 1000, 2000, 3000, 4000, 5000, 7000, and 9000 W/cm^2 were tested. For studies of the effects of gas saturation on initiation delay time, I_{SPPA} was kept constant at 5000 W/cm^2. Gas saturations of three different ranges of 24%–28%, 39%–49%, 77%–81% were used. All these data were also used in the study of the initiation and extinction processes and the relationship of initiation to erosion. The partial pressure of oxygen (PO_2) in air was used as our metric for gas saturation and the PO_2 level was measured with YSI dissolved oxygen instruments (YSI 5000, Yellow Springs, OH).

TABLE II. Number of recorded initiation and erosion events.

I_{SPPA} (W/cm^2)	Gas saturation	Number of treatments	Initiation and erosion	No initiation and no erosion	Initiation but no erosion	No initiation but erosion
1000	39%–49%	12	0	12	0	0
2000	39%–49%	12	0	11	1	0
3000	39%–49%	12	4	7	0	0
3500	39%–49%	8	5	3	0	0
4000	39%–49%	12	12	0	0	0
5000	24%–28%	8	8	0	0	0
5000	39%–49%	8	8	0	0	0
5000	77%–81%	8	8	0	0	0
7000	39%–49%	8	8	0	0	0
9000	39%–49%	8	8	0	0	0
Treatment number		95	61	33	1	0
Success prediction rate					98.9%	

TABLE III. Number of recorded extinction (excluding perforation) events.

I_{SPPA} (W/cm^2)	Gas saturation	Number of treatments	Number of treatments with extinction (excluding perforation)	Number of extinction events
1000	39%–49%	12	0	0
2000	39%–49%	12	1	1
3000	39%–49%	11	4	5
3500	39%–49%	8	4	13
4000	39%–49%	12	6	8
5000	24%–28%	8	1	1
5000	39%–49%	8	0	0
5000	77%–81%	8	0	0
7000	39%–49%	8	1	1
9000	39%–49%	8	0	0
Treatment number		95	17	29

III. RESULTS

A total of 95 ultrasound treatments was applied to 33 pieces of 1–3-mm-thick porcine atrial wall. The acoustic backscattering signals recorded and tissue effects produced by the corresponding ultrasound treatments are included in the following analysis. The initiation phenomenon was observed in 62 of 95 treatments (Table II). The extinction (excluding perforation) phenomenon was observed in 17 of 95 treatments (Table III).

A. Relationship between initiation and erosion

Results show that the initiation and the erosion processes are highly correlated. As shown in Table II, no erosion was observed in any of the 33 treatments where initiation was not detected. Among 61 of 62 treatments where initiation was detected, visible erosion was also observed in the tissue. Therefore, initiation predicted erosion, or lack or erosion, successfully at a rate of 98.9% (94 out of 95 treatments).

Figure 5 graphically depicts the correlation between initiation and erosion. All three tissue samples were treated for 8 min by ultrasound pulses with a PD of 3 cycles, a PRF of 20 kHz, an I_{SPPA} of 3500 kHz, and gas saturation of 40%–45%. The first three rows show the backscatter in fast time and slow time mode, backscatter energy versus time, and the moving SD of backscatter energy versus time, respectively. The pictures in the last row show the tissue effects generated by the corresponding ultrasound treatments. In panel (A), a nearly flat backscatter energy moving SD trace indicates that no initiation occurred, and there was no erosion in the tissue. In panels (B) and (C), the backscatter energy moving SD increased significantly and remained high for a period of time. Correspondingly, erosion appeared in both tissue samples.

B. Variability of initiation and extinction

Initiation was highly stochastic in nature, particularly at intermediate intensities (~3000 W/cm^2). For example, at I_{SPPA}'s of 3000 and 3500 W/cm^2, initiation occurred in an unpredictable manner (Table IV). The same 8-min ultrasound exposure (3-cycle PD, 20-kHz PRF and 39%–49% gas saturation) was applied to all the treatments reported in Table IV. In 10 of 19 treatments, however, neither initiation nor erosion was observed (Table IV). Both initiation and erosion were observed in the other nine cases.

After initiation, extinction also occurred in a random manner at intermediate intensities. An 8-min ultrasound exposure (I_{SPPA}'s of 3000–4000 W/cm^2, 3-cycle PD, 20-kHz PRF, 39%–49% gas saturation) was applied to all the treatments in Table V. But, out of 21 treatments during which initiation was observed, extinction was detected in 14 cases (Table V).

Furthermore, in some treatments, extinction events and subsequent reinitiation of the highly backscattering environment occurred in an unpredictable manner (Table V). In 3 out of the 14 treatments where extinction was detected, no subsequent reinitiation occurred. Erosion was observed; none of these tissue samples was perforated. In two treatments, multiple extinction and reinitiation events occurred, and erosion without perforation was observed. In the remaining nine treatments, multiple extinction and reinitiation events were observed, and tissue was eventually perforated.

Figure 5 demonstrates the variability of initiation and extinction resulting in different tissue effects even when the same acoustic parameters were applied. In panel (A), neither initiation nor erosion was seen. In panel (B), both initiation and extinction were detected, and the tissue was eroded, although no perforation occurred within the 8-min exposure. In panel (C), initiation without extinction was observed, and the tissue was perforated.

C. Initiation delay time versus intensity

Multiple pulses at a PD of 3 cycles, a PRF of 20 kHz, and I_{SPPA}'s of 1000, 2000, 3000, 4000, 5000, 7000, and 9000 W/cm^2 were applied. Gas saturation was kept at 39%–49%. At I_{SPPA}'s of 1000 and 2000 W/cm^2, initiation was almost never observed within the 8-min ultrasound exposure (Fig.

TABLE IV. Number of recorded initiation, erosion, and perforation events at I_{SPPA}'s of 3000–3500 W/cm^2.

I_{SPPA} (W/cm^2)	Number of treatments	No initiation and no erosion	Initiation and erosion	Initiation and erosion, no perforation	Initiation and perforation	No initiation, but erosion	Initiation, but no erosion
3000	11	7	4	4	0	0	0
3500	8	3	5	1	4	0	0
Treatment number	**19**[a]	**10**[a]	**9**[a]	5	4	0	0

[a]Marks the columns referred to in the text.

TABLE V. Number of recorded extinction and reinitation events at I_{SPPA}'s of 3000–4000 W/cm^2.

I_{SPPA} (W/cm^2)	Number of treatments	Initiation and erosion	Extinction	Extinction, no reinitiation, and no perforation	Extinction, reinitiation, but no perforation	Extinction, reinitiation, and perforation
3000	11	4	4	3	1	0
3500	8	5	4	0	1	3
4000	12	12	6	0	0	6
Treatment number	31	**21**[a]	**14**[a]	**3**[a]	**2**[a]	**9**[a]

[a]Marks the columns referred to in the text.

7). At an I_{SPPA} of 3000 W/cm^2, initiation sometimes occurred after one or more minutes (Fig. 7). At an I_{SPPA} at or above 4000 W/cm^2, initiation always occurred (Fig. 7).

The initiation delay time is dependent upon intensity. It was shorter with higher intensity [Fig. 7, Fig. 8(A)]. The sample size for each I_{SPPA} is listed in Table VI. For example, the mean initiation delay time was 66.9 s at an I_{SPPA} of 4000 W/cm^2 and 3.6 ms at an I_{SPPA} of 9000 W/cm^2, a 4-order-of-magnitude difference ($p<0.0001$; *T*-test) (Table VI). Variances in the initiation delay times were also lower with higher intensity. For example, the standard deviation in initiation delay time was 33.3 s at an I_{SPPA} of 4000 W/cm^2 and 1.9 ms at an I_{SPPA} of 9000 W/cm^2, a 4-order-of-magnitude difference (Table VI).

D. Initiation delay time versus gas saturation

Multiple pulses at a PD of 3 cycles, a PRF of 20 kHz, and an I_{SPPA} of 5000 W/cm^2 were applied. Gas saturation in the ranges of 24%–28%, 39%–49%, and 77%–81% were used to study the effects of gas saturation on initiation delay time. The sample size was eight for each gas saturation range. Results show that the initiation delay time was shorter with higher gas saturation [Fig. 8(B)]. For example, the mean initiation delay times were 133.1, 48.0, and 24.7 ms at gas saturation ranges of 21%–24%, 39%–49%, and 77%–81%, respectively (Table VI). The variances of initiation delay times were lower with higher gas saturation. For example, the standard deviations of initiation delay time were 78.3, 46.4, and 25.0 ms at gas saturation ranges of 21%–24%, 39%–49%, and 77%–81%, respectively (Table VI).

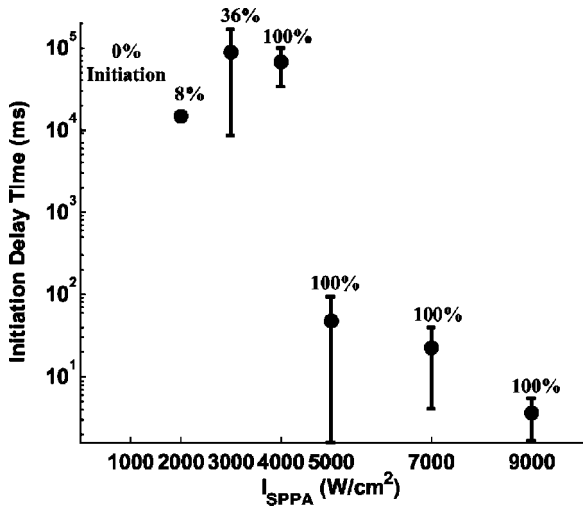


FIG. 7. Initiation delay time as a function of I_{SPPA}. I_{SPPA}'s of 1000, 2000, 3000, 4000, 5000, 7000, and 9000 W/cm^2 were tested. A PD of 3 cycles, a PRF of 20 kHz, and gas saturation range of 39%–49% were used for all the ultrasound exposures. Initiation delay time was plotted as mean and standard deviation values. The sample size is listed as number of treatments in Table VI. The number above each data point is the percentage of trial in which initiation was detected for this same set of parameters. Initiation delay time was shorter with high intensity. For example, the mean initiation delay time was 66.9 s at an I_{SPPA} of 4000 W/cm^2 and 3.6 ms at an I_{SPPA} of 9000 W/cm^2, a 4-order-of-magnitude difference ($p<0.0001$; t-test).

E. High-intensity initiating sequences

Our results also show that, once initiated, the highly backscattering environment can be maintained at significantly lower pulse intensities than that required to initiate the process. Figure 9 demonstrates that, after initiation with 100-ms 9000 W/cm^2 pulses, a highly backscattering environment associated with observed erosion could be continued with 2000-W/cm^2 pulses. Among 12 treatments at 2000 W/cm^2, without an initiation sequence at 9000 W/cm^2, initiation occurred only once and no erosion was observed in any of the 12 cases (Table II). A PD of 3 cycles and a PRF of 20 kHz were used in both the high-intensity initiating pulses and low-intensity sustaining pulses.

IV. DISCUSSION

In this paper, we show that without initiation of a highly backscattering environment, there is no erosion. The initiation predicted erosion or lack of erosion successfully at a rate of 98.9% (Table II). Initiation and extinction phenomena were also observed in the *in vivo* experiments performed by our colleague Tran (Tran *et al.*, 2003). Tran *et al.* show that without initiation, no lesion was created in kidney tissues. We believe that initiation is the key to decreasing the variability of cavitational bioeffects often reported in the literature (Fry *et al.*, 1995).

We hypothesize that initiation of a highly backscattering environment is due to the generation of a dynamic ensemble of microbubbles (''bubble cloud'') which are maintained over time by an appropriately designed series of acoustic pulses. A bubble cloud likely contains microbubbles of various sizes and temporal persistence, many of which are effective cavitation nuclei. Whenever subsequent pulses fail to maintain a ''predisposing'' bubble cloud, extinction occurs. This bubble cloud is easily detected through observation of an enhanced rapidly changing acoustic backscattering environment.

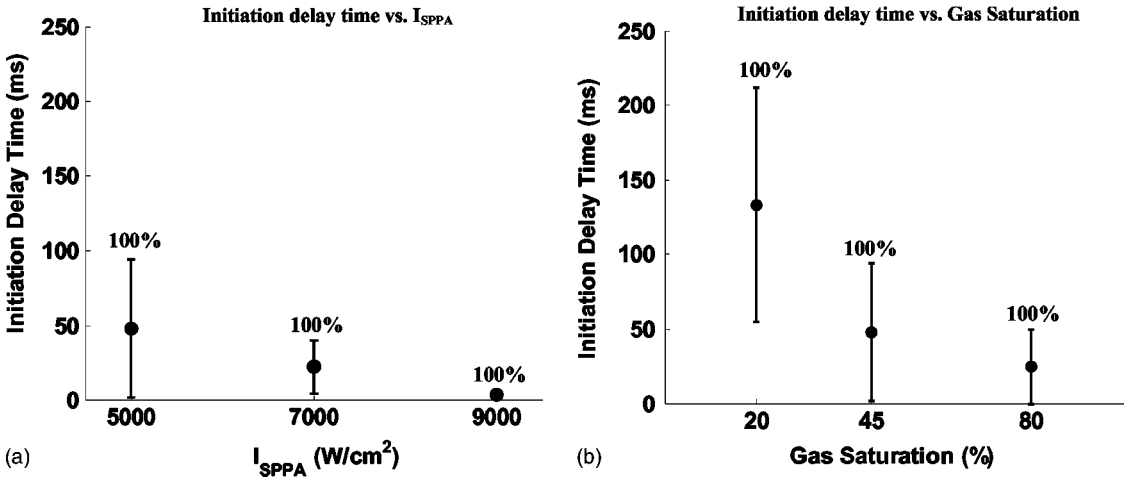


FIG. 8. Panel (A) shows the initiation delay time as a function of intensity. I_{SPPA}'s of 5000, 7000, and 9000 W/cm^2 were tested. A PD of 3 cycles, a PRF of 20 kHz, and gas saturation of 39%–49% were applied to all the exposures in panel (A). Initiation delay times and variances in initiation delay times were shorter with higher intensity. For example, the mean initiation delay time was 48.0 ms at an I_{SPPA} of 5000 W/cm^2 and 3.6 ms at an I_{SPPA} of 9000 W/cm^2. Panel (B) shows the initiation delay time as a function of gas saturation. Gas saturation ranges of 24%–28%, 39%–49%, 77%–81% were tested and plotted as gas saturations of 25%, 45%, and 80% for convenience of display. A PD of 3 cycles, a PRF of 20 kHz, and an I_{SPPA} of 5000 W/cm^2 were applied to all the exposures in panel (B). Initiation delay times and variances in initiation delay times were shorter with higher gas saturation. For example, the mean initiation delay time was 133.1, 48.0, and 24.7 ms at gas saturation ranges of 21%–24%, 39%–49%, and 77%–81%, respectively. Initiation delay time was plotted as mean and standard deviation values ($N=8$) in both panels. The number above each data point is the percentage of trial in which initiation was detected for the same set of parameters.

Our hypothesis is consistent with the results that initiation delay times and variances in initiation delay times are shorter with higher gas saturation. With higher gas content, more nuclei are available to initiate the ensemble of microbubbles. Therefore, it takes less time to initiate the bubble cloud, and the probability of initiation is high. A similar explanation can be applied to results showing that initiation delay times and variances in initiation delay times are shorter with higher I_{SPPA}. With higher I_{SPPA}, the volume where the pulse intensities are greater than the cavitation threshold is larger. The size range of microbubbles that can be utilized as effective cavitation nuclei also increases as intensity increases (Apfel and Holland, 1991). As a result, with higher I_{SPPA}, more cavitation nuclei are available to initiate the bubble cloud.

Furthermore, a cooperative environment between the acoustic pulses and the bubble cloud is also required to achieve the desired tissue outcome. This is supported by our results showing that the extinction was observed and perforation was not generated at some intermediate intensities (e.g., an I_{SPPA} of 3000 W/cm^2). The failure to achieve perforation is possibly due to the inability of acoustic pulses to maintain the bubble cloud.

We hypothesize that, with appropriately adjusted acoustic pulse parameters, each pulse arrives to find appropriate cavitation nuclei within the bubble cloud to produce some desired tissue effect after the bubble cloud is initiated. Other bubbles in the bubble cloud, along with ''fragments'' of microbubbles generated from previous energetic collapses, undergo dynamic changes (e.g., grow, decay, or spatially move) to form more cavitation nuclei at the focus, therefore predisposing the cloud for subquent pulses. Our goal is to design a set of acoustic parameters, perhaps temporally changing based on real-time feedback, in order to create an optimum cooperative condition where each pulse can both generate desired tissue effects and provide sufficient cavitation nuclei for the next pulse.

Tissue erosion by high-intensity ultrasound is a very complex process. We postulate that tissue erosion can be

TABLE VI. Initiation delay time.

I_{SPPA} (W/cm^2)	Gas saturation	Treatment	Initiation	Initiation percentage	Initiation delay time (ms) Mean	Initiation delay time (ms) Standard deviation
1000	39%–49%	12	0	0	⋯	⋯
2000	39%–49%	12	1	8.3%	14 459.2	0
3000	39%–49%	11	4	36.4%	88 429.9	79 879.6
4000	39%–49%	12	12	100%	66 865.3	33 287.4
5000	24%–28%	8	8	100%	133.1	78.3
5000	39%–49%	8	8	100%	48.0	46.4
5000	77%–81%	8	8	100%	24.7	25.0
7000	39%–49%	8	8	100%	22.1	18.0
9000	39%–49%	8	8	100%	3.6	1.9

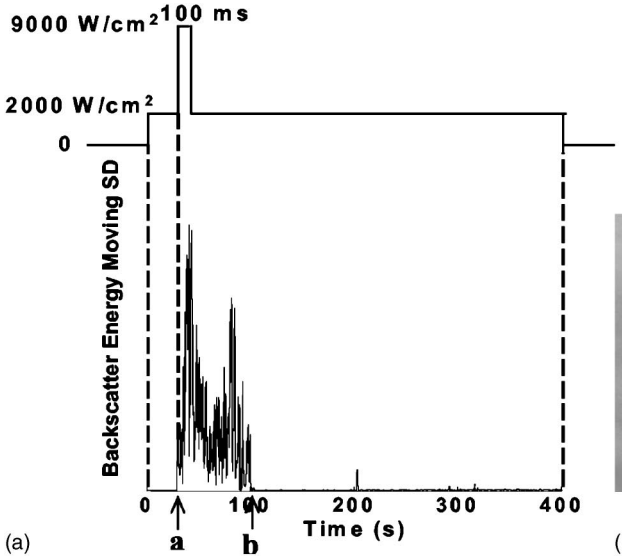


FIG. 9. In panel (A), acoustic pulses with an I_{SPPA} of 2000 W/cm^2 were applied from time 0 to 400 s. Panel (A) shows the moving SD of backscatter energy versus time. No initiation was detected until 100-ms 9000 W/cm^2 pulses were applied at "a" (initiation). The highly backscattering environment was continued by 2000-W/cm^2 pulses and extinguished at "b" (extinction). Panel (B) shows the erosion observed in the porcine wall tissues produced by the corresponding acoustic pulses. Please note that no erosion was produced at 2000 W/cm^2 alone among 12 treatments tested (Table II). (All the pulses applied were at a PD of 3 cycles, a PRF of 20 kHz, and gas saturation of 42%–43%).

achieved through several mechanisms, including cavitational and noncavitational mechanisms. The high correlation between the backscatter behavior and erosion seen in these studies suggests that under the conditions used here, the mechanism is likely associated with cavitation including fluid effects such as water jets and microstreaming. However, this does not preclude a contribution from other noncavitational effects. At this point, we do not have a detailed understanding of the mechanism for the tissue erosion and how the activity of the bubble cloud changes temporally and spatially with acoustic parameters. In order to understand the mechanisms responsible for this erosion process, we are planning experiments using optical attenuation and scattering methods and high-speed camera to monitor the activity of bubble cloud generated by ultrasound pulses and interaction between the bubble cloud and the tissue.

Our results also show that once the highly backscattering environment (and erosion) is initiated by high-intensity pulses, it can be maintained, and erosion can be continued by significantly lower intensity pulses than that required for initiation. This suggests a strategy where short, high-intensity pulses are used to initiate the highly backscattering environment, low-intensity pulses are used to maintain it, and high-intensity pulses are used again (when necessary) to reinitiate it when extinction is detected. We believe this strategy will allow us to operate the erosion process at a lower intensity, thus minimizing propagated energy and perhaps preventing thermal damage to the overlying tissue and the transducer.

An attempt has been made to develop a statistical approach to detect initiation and extinction based on backscatter variability. It should not be surprising that a violently cavitating environment associated with tissue erosion is highly correlated with high backscatter variability. High variability, particularly related to cavitational bioeffects, has been an impediment to the development of therapeutic applications (Fry *et al.*, 1995). Understanding and controlling the variability, particularly the "initiation" and "extinction" phenomena, may be the key to making cavitational therapy predictable and controllable.

It is also worth noting that extinction may provide a useful means for perforation detection. If tissue is not perforated, reinitiation can be easily achieved by high-intensity pulses. Therefore, a reinitiation test after extinction may serve as a way to detect perforation. Color-flow Doppler imaging would also be used for this purpose. Perforation of atrial septum can be determined by the detection of a blood flow between two atria in the *in vivo* situation (Xu *et al.*, 2004).

In summary, we believe that a thorough understanding of initiation and extinction are the key to controlling the erosion process as well as other therapeutic bioeffects produced by cavitation. Proper maintenance of a synergistic interaction between insonation and the dynamically changing ensemble of microbubbles can be controlled by adjustment of acoustic parameters (e.g., pulse intensity, PD, and PRF). Our goal is to make the maintenance of the synergistic environment controllable and predictable with a minimum of propagated energy. Furthermore, it may be possible to achieve this control using easily obtained variations in acoustic pulse parameters under real-time feedback condition.

V. CONCLUSIONS

There is a high correlation between the enhanced, rapidly changing backscatter and the erosion process. The presence and absence of initiation of the highly backscatter environment successfully predict erosion and lack of erosion at a rate of 98.9%. Initiation delay time, presumably the formation time of the bubble cloud, decreases significantly with higher intensity. Further, once initiated, erosion can be maintained at a lower intensity than that required for initiation. We believe that understanding of the underlying mechanism of initiation (and possible extinction) of an ensemble of predisposing microbubbles is the key to controlling the erosion process as well as many other therapeutic effects due to cavitation.

ACKNOWLEDGMENTS

We would like thank Dr. Achi Ludomirsky for his consultation in HLHS and other congenital cardiac diseases. We also thank Dr. Oliver Kripfgans and Jessica Parsons for their help with the calibration procedure. This research has been funded by grants from the National Institutes of Health RR14450.

AIUM (**1998**). Acoustic Output Measurement Standard for Diagnostic Ultrasound Equipment, AIUM/NEMA.

Apfel, R. E., and Holland, C. K. (**1991**). ''Gauging the likelihood of cavitation from short-pulse, low-duty cycle diagnostic ultrasound,'' Ultrasound Med. Biol. **17**, 179–185.

Atchley, A. A. (**1988**). ''Thresholds for cavitation produced in water by pulsed ultrasound,'' Ultrasonics **26**, xxx–xxx.

Belahadji, B., Franc, J. P., and Michel, J. M. (**1991**). ''A statistical analysis of cavitation erosion pits,'' J. Fluids Eng. **113**, 700–706.

Bouakaz, A., Frinking, P. J. A., Jong, N. D., and Bom, N. (**1999**). ''Noninvasive measurement of the hydrostatic pressure in a fluid-filled cavity based on the disappearance time of micrometer-sized free gas bubbles,'' Ultrasound Med. Biol. **25**, 1407–1415.

Calabrese, A. M. (**1996**). Ph.D. thesis: ''Threshold measurements and production rates for inertial cavitation due to pulsed, megahertz-frequency ultrasound,'' Chaps. 5 and 6. The University of Mississippi.

Chakravarty, A., and Walton, A. J. (**2001**). ''Light emission from collapsing superheated steam bubbles in water,'' J. Lumin. **92**, 27–33.

Chen, W. S., Matula, T. J., and Crum, L. A. (**2002**). ''The disappearance of ultrasound contrast bubbles: Observations of bubble dissolution and cavitation nucleation,'' Ultrasound Med. Biol. **28**, 793–803.

Child, S. Z., Hartman, C. L., Schery, L. A., and Carstensen, E. L. (**1990**). ''Lung damage from exposure to pulsed ultrasound,'' Ultrasound Med. Biol. **16**, 817–825.

Dalecki, D., Raeman, C. H., Child, S. Z., and Carstensen, E. L. (**1995**). ''Intestinal hemorrhage from exposure to pulsed ultrasound,'' Ultrasound Med. Biol. **21**, 1067–1072.

Dunn, F., and Fry, F. J. (**1971**). ''Ultrasonic threshold dosages for the mammalian central nervous system,'' IEEE Trans. Biomed. Eng. **18**, 253–256.

Everbach, E. C., Makin, I., Azadniv, M., and Meltzer, R. S. (**1997**). ''Correlation of ultrasound-induced hemolysis with cavitation detector output *in vitro*,'' Ultrasound Med. Biol. **23**, 619–624.

Fairbank, W. M., and Scully, M. O. (**1977**). ''A new non-invasive technique for cardiac pressure measurement resonant scattering of ultrasound from bubble,'' IEEE Trans. Biomed. Eng. **BME-24**, 107–110.

Frizzell, L. A., Chen, E., and Lee, C. (**1994**). ''Effects of pulsed ultrasound on the mouse neonate: Hind limb paralysis and lung hemorrhage [Comment],'' Ultrasound Med. Biol. **20**, 53–63.

Fry, F. J., Sanghvi, N. T., Foster, R. S., Bihrle, R., and Hennige, C. (**1995**). ''Ultrasound and microbubbles: Their generation, detection, and potential utilization in tissue and organ therapy—experimental,'' Ultrasound Med. Biol. **21**, 1227–1237.

Holland, C. K., and Apfel, R. E. (**1990**). ''Thresholds for transient cavitation produced by pulsed ultrasound in a controlled nuclei environment,'' J. Acoust. Soc. Am. **88**, 2059–2069.

Jarman, P. D., and Taylor, K. J. (**1965**). ''Light flashes and shocks from a cavitation flow,'' Br. J. Appl. Phys. **16**, 675–682.

Jochle, K., Debus, J., Lorenz, W. J., and Huber, P. (**1996**). ''A new method of quantitative cavitation assessment in the field of a lithotripter,'' Ultrasound Med. Biol. **22**, 329–338.

Kripfgans, O. D., Fowlkes, J. B., Miller, D. L., Eldevik, O. P., and Carson, P. L. (**2000**). ''Acoustic droplet vaporization for therapeutic and diagnostic applications,'' Ultrasound Med. Biol. **26**, 1177–1189.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic, London).

Lush, P. A., and Angell, B. (**1984**). ''Correlation of cavitation erosion and sound pressure level,'' J. Fluids Eng. **106**, 347–351.

Madanshetty, S. I., Roy, R. A., and Apfel, R. E. (**1991**). ''Acoustic microcavitation: Its active and passive acoustic detection,'' J. Acoust. Soc. Am. **90**, 1515–1526.

Matsumoto, Y., Yoshizawa, S., and Teiichiro, I. (**2002**). ''Dynamics of bubble cloud in focused ultrasound,'' International Symposium on Therapeutic Ultrasound, Seattle, WA, pp. 290–299.

Messino, C. D., Sette, D., and Wanderlingh, F. (**1963**). ''Statistical approach to ultrasonic cavitation,'' J. Acoust. Soc. Am. **35**, 1575–1583.

Ohl, C. D. (**2000**). ''Luminescence from acoustic-driven laser-induced cavitation bubbles,'' Phys. Rev. E **61**, 1497–1500.

Ohl, C. D., Lindau, O., and Lauterborn, W. (**1998**). ''Luminescence from spherically and aspherically collapsing laser induced bubbles,'' Phys. Rev. Lett. **80**, 393–396.

Phillip, A., and Lauterborn, W. (**1998**). ''Cavitation erosion by single laser-produced bubbles,'' J. Fluid Mech. **361**, 75–116.

Phillip, A., and Ohl, C. D. (**1995**). ''Single bubble erosion on a solid surface,'' Interational Symposium on Cavitation, CAV'95, Deuville, France, 297–303.

Poliachik, S. L. (**1999**). ''Effect of high-intensity focused ultrasound on whole blood with and without microbubble contrast agent,'' Ultrasound Med. Biol. **25**, 991–998.

Roy, R. A., Church, C. C., and Calabrese, A. (**1990a**). ''Cavitation produced by short pulses of ultrasound,'' in *Frontiers of Nonlinear Acoustics*, 12 ISNA, edited by D. T. Blackstock and M. F. Hamilton (Elsevier, Amsterdam), pp. 476–491.

Roy, R. A., Madanshetty, S. I., and Apfel, R. E. (**1990b**). ''An acoustic backscattering technique for the detection of transient cavitation produced by microsecond pulses of ultrasound,'' J. Acoust. Soc. Am. **87**, 2451–2458.

Roy, R. A., Atchley, A. A., Crum, L. A., Fowlkes, J. B., and Reidy, J. J. (**1985**). ''A precise technique for the measurement of acoustic cavitation thresholds and some preliminary results,'' J. Acoust. Soc. Am. **78**, 1799–1805.

Shi, W. T., Forsberg, F., Tornes, A., Ostensen, J., and Goldberg, B. B. (**2000**). ''Destruction of contrast microbubbles and the association with inertial cavitation,'' Ultrasound Med. Biol. **26**, 1009–1019.

Smith, N. B., and Hynynen, K. (**1998**). ''The feasibility of using focused ultrasound for transmyocardial revascularization,'' Ultrasound Med. Biol. **24**, 1045–1054.

Tomlinson, W. J., and Matthews, S. J. (**1994**). ''Cavitation erosion of aluminium alloys,'' J. Mater. Sci. **29**, 1101–1108.

Tran, B. C. (**2003**). ''*In Vivo* Comparison of Multiple Pulse and CW strategies for Microbubble-Enhanced Ultrasound Therapy,'' IEEE Ultrasonics Symposium, Hawaii, pp. 1J–4.

Wetheril, G. B., and Brown, D. W. (**1991**). *Statistical Process Control Theory and Practice* (Chapman and Hall, London).

Xu, Z. (**2004**). ''Controlled ultrasound tissue erosion,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 726–736.

Sensitivity of a tucuxi (*Sotalia fluviatilis guianensis*) to airborne sound

Alexander Liebschner, Wolf Hanke, Lars Miersch, and Guido Dehnhardt[a)]
Allgemeine Zoologie und Neurobiologie, Ruhr-Universität Bochum, ND 6/33, 44780 Bochum, Germany

Matthias Sauerland
Universität Münster, Institut für Neuro- und Verhaltensbiologie, Abteilung Verhaltensbiologie, Badestrasse 13, 48149 Münster, Germany

(Received 2 March 2004; revised 11 October 2004; accepted 18 October 2004)

Auditory systems of cetaceans are considered highly specialized for underwater sound processing, whereas the extent of their hearing capacity in air is still a point of issue. In this study, the sensitivity to airborne sound in a male tucuxi (*Sotalia fluviatilis guianensis*) was tested by means of a go/no go response paradigm. Auditory thresholds were obtained from 2 to 31.5 kHz. Compared to the hearing thresholds of other dolphins as well as of amphibian mammals, the sensitivity to airborne sound of the test subject is low from 2 to 8 kHz, with the highest threshold at 4 kHz. Thresholds at 16 and 31.5 kHz reveal a sharp increase in hearing sensitivity. Thus, although not obtained in this study, the upper aerial hearing limit is in the ultrasonic range. A comparison of the present data with the underwater audiogram of the same test subject referred to sound intensity indicates that the sensitivity of Sotalia to underwater sound is generally better than to airborne sound. [DOI: 10.1121/1.1828851]

PACS numbers: 43.80.Lb [JAS] Pages: 436–441

I. INTRODUCTION

The evolution of dolphins from land-living carnivores resulted in extensive anatomical and physiological transformations, not least concerning their sensory abilities. A complex sonar system for underwater orientation and prey detection is one of the most impressive features proving dolphins to be highly adapted to their aquatic environment. However, the ability to obtain sensory information in air might be still useful. For instance, ''spy hopping'' whales are suggested to raise themselves vertically out of the water to gain visual information about the environment (Layne, 1958; Pike, 1962; MacAskie, 1966; Cummings *et al.*, 1971; Norris, 1974; Herman and Forestell, 1977). This corresponds with results from studies on vision in bottlenose dolphins (*Tursiops truncatus*), revealing that in bright light aerial visual acuity compares well to that under water (Herman *et al.*, 1975).

Studies on the hearing sensitivity to airborne sound are only available for the bottlenose dolphin (Babushina, 1979) and the harbor porpoise (*Phocoena phocoena*) (Kastelein *et al.*, 1997). Comparing aerial with underwater hearing thresholds in these two species by means of perceived minima of sound intensities yields a higher sensitivity to underwater sound for all frequencies tested (Babushina, 1979; Kastelein *et al.*, 1997, 2002). Following the argumentation by Au (1993) and Yost (1994), mammalian ears are primarily sound intensity detectors, or energy detectors if the signal duration is short enough to influence the hearing threshold. Johnson (1968) showed that the bottlenose dolphin hearing threshold varies with the duration of the signal according to the equation $I/I_\infty = 1 + \tau/t$, where I_∞ is the threshold intensity for a very long signal and τ is a constant in the order of tenths of a second, depending on frequency. For technical reasons, received intensity, measured in W/m^2, is difficult to determine. Therefore it has become standard practice to compare audiograms obtained in the same medium by the directly measured sound pressure. However, the sound pressure in water must be about 60 times that required in air to produce the same intensity. As most authors assume that the marine mammal ear is a sound intensity detector (Ketten, 2000), the comparison of audiograms taken in different media is usually done by sound intensities. Data that were recorded as sound pressures are transformed according to the equation $I = p^2/(c^*\rho)$ (I = intensity; p = pressure; c = speed of sound; ρ = density of the medium). However, the parameter applicable for comparisons of hearing threshold data in air with those obtained in water is still a matter of discussion (Ketten, 2000).

In this study the authors psychophysically determined aerial auditory thresholds of *Sotalia fluviatilis guianensis*, a small delphinid species from the east coast of South America, and compared them on the basis of sound intensity with thresholds previously obtained for the same animal under water (Sauerland and Dehnhardt, 1998).

II. METHOD

A. Subject

The test subject ''Paco,'' a 28-year-old *Sotalia fluviatilis guianensis*, has been kept at the Münster Dolphinarium (Germany) since 1991. Paco measures 1.7 m in length and weighs about 60 kg. In addition to our tests the animal took part in three shows daily. Of the 3 to 4 kg of sprats and

a)Author to whom correspondence should be addressed. Electronic mail: dehnhardt@neurobiologie.ruhr-uni-bochum.de

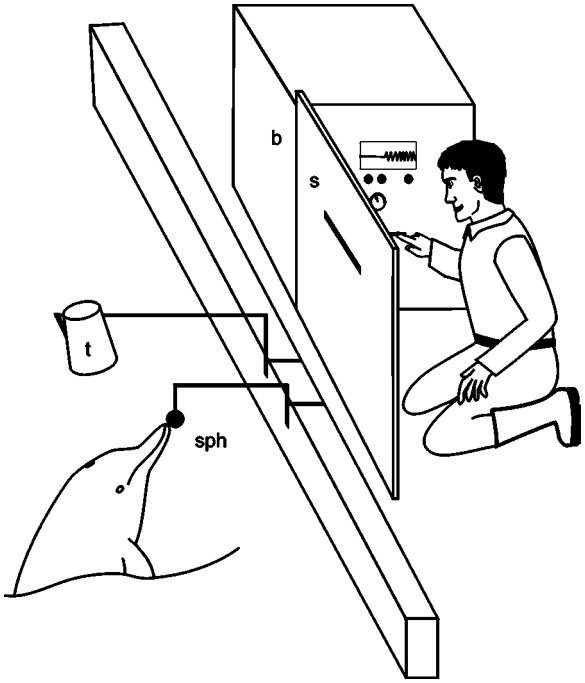


FIG. 1. Experimental setup. The electronic instruments were housed in a transportable box (“b”). An opaque screen (“s,” 100×50 cm^2) with small slits was attached to the rear of the box. It allowed the experimenter to observe the animal during tests without unintentional cueing. A U-shaped stationing device was mounted on the concrete edge of the pool with a little sphere (“sph,” diameter 8 cm) hanging 45 cm above the water surface at a distance of 50 cm from the pool wall. When in station, the dolphin touched the sphere with the tip of its rostrum. The transducer for the test signals (“t”) was mounted above the animal station so that the distance between transducer and the dolphin’s auditory meatus was about 50 cm.

herring fed per day, about 1.5 kg could be used for reinforcement during hearing tests. Tests were carried out once per day, on 5 to 7 days per week.

B. Apparatus

Threshold tests were conducted in a rectangular concrete pool (20×10 m^2, depth 4 m) of the Münster Dolphinarium. The experimental setup is shown in Fig. 1. The electronic instruments were housed in a transportable box. An opaque screen (100×50 cm^2) with small slits was attached to the rear of the box. It allowed the experimenter to observe the animal during tests without unintentional cueing. A U-shaped stationing device was mounted on the concrete edge of the pool with a little sphere (diameter 8 cm) hanging 45 cm above the water surface at a distance of 50 cm from the pool wall. When in station, the dolphin touched the sphere with the tip of its rostrum. This way the animal’s body was out of the water up to the flippers. The transducer for the test signals was mounted above the animal station so that the distance between the transducer and the dolphin’ s external auditory meatus was about 50 cm, depending on the frequency tested (see Sec. II C). To exclude the transduction of stimuli to the rostrum via the sphere, part of the sphere holder consisted of rubber and the sound transducer was suspended elastically. In addition, a laser vibrometer (Polytec OFV 3001 with sensor head OFV 303) was used to detect possible oscillations of the sphere caused by the loudspeaker at 31.5 kHz. Velocity of the sphere at 31.5 kHz could not be measured and was thus clearly below 0.1 mm/s rms for a stimulus 63.5 dB above threshold, i.e., displacement amplitude was lower than 3.4×10^{-13} m at threshold. Displacement of water particles at the underwater hearing threshold of Sotalia at 32 kHz [64 dB re 1 μPa (Sauerland and Dehnhardt, 1998)] is 4×10^{-14} m. That means that in the worst case the (not measurable) vibrations of our stationing sphere might have been ten times larger in displacement than the underwater sound wave at hearing threshold. But the pressure delivered by the elastically suspended sphere is several orders of magnitude lower than the pressure in an underwater sound wave of same displacement. Bullock *et al.* (1968) found that the sensitivity to a 30-kHz stimulus delivered with a hydrophone pressed against the skin of *Stenella coeruleo-alba* was 40 to 50 times lower at the tip of the lower jaw than at its side. Moreover, a pressure wave that reaches a large area of the head should be more effective than a locally applied stimulus of equal amplitude. Hence, the reactions of our Sotalia were due to airborne sound.

C. Stimuli

The sinusoidal test signal was produced with a programmable Kontron function generator (model 8022). Rise-fall times of the test signal (150 ms) as well as the test signal length of 4 s were timed with a control generator connected to the AM-input of the function generator.

Before going to the transducer, the test signal passed an attenuator (1-dB steps possible, custom designed) and a Wintec Stereo Amplifier (model 6002). Different loudspeakers were used as transducers: a Europa CL 200 speaker for threshold measurements at 2 and 4 kHz, a Tonsil Hi-Fi-Calotte (GTC 8/10/1) at 8 and 16 kHz, and a Conrad Electronic model (No. 33 5835-66) at 31.5 kHz.

The received sound pressure level of the test signal at the animal’s station was measured at 0 dB of attenuation each day before tests took place and served as a reference for the data obtained that day. These measurements were carried out with a Norsonic microphone (model 1201/15463) and a Norsonic sound analyzer (model 110) for test frequencies up to 16 kHz. At 31.5 kHz the same microphone and a Norsonic sound measuring system (model 823) were used. Using the same sound analyzing systems, fluctuations of the received level of the test signal as well as the ambient noise at the animal’s station were measured. Fluctuations of the test signal were presumably caused by reflections from the water surface and the walls of the pool. By varying the distance and the angle of the speakers related to the animal’s station signal fluctuations were reduced to $\leqslant5$ dB. Ambient noise at the animal’s station was measured in $\frac{1}{3}$ octave bands from 0.8 kHz up to 20 kHz (Norsonic microphone 1201/15463 and Norsonic sound analyzer 110) and at 31.5 kHz (Norsonic sound measuring system 823). Noise decreased from 37 dB (re 20 μPa) at 0.8 kHz to 13 dB at 31.5 kHz, with fluctuations of less than 4 dB.

D. Procedure

Tests were carried out according to a go/no go response paradigm. A trial started with the dolphin stationing at the experimenter's hand. Following a hand cue, the dolphin immediately swam to the station. After a 3-s delay the test signal was projected for 4 s. When the dolphin detected the signal it immediately left the station (go response) and swam to the experimenter to receive a fish reward. In signal-absent trials the animal remained in station for 15 s (no go response) until the experimenter signaled the end of the trial with a release tone by means of a training whistle. The dolphin's responses were reinforced differentially to bias the animal towards the go response. A correct response to signal-present trials was rewarded with one sprat and one herring, whereas the animal received two sprats for a correct response to signal-absent trials. If the animal remained in station during a signal-present trial (miss), the experimenter immediately terminated the test cycle. Missing a signal as well as leaving the station in signal-absent trials during the 15-s interval (false alarm) were designated as wrong responses and were not reinforced.

The sequence of signal-present and signal-absent trials was based on a modified Gellerman random series (Gellerman 1933) with half of the trials being signal-absent. Thresholds were determined by using the ''up/down'' or ''staircase'' method, a variation of the method of limits, which results in a 50% correct detection threshold (Levitt, 1971; Gescheider, 1985). The signal amplitude was attenuated in 2-dB steps on each signal-present trial, until the animal failed to respond to a test signal. The signal level was then increased in 2-dB steps, until the dolphin responded correctly again, thereby indicating that it could detect the signal. Intensities at which the animal reversed its response behavior were taken as data points. The occurrence of false alarms did not alter the sequence of attenuator settings. A day's data collection consisted of 24–36 trials, of which 2–6 trials were reversals. The average of the signal at ten consecutive reversals yielded the estimate of the mean threshold at a given frequency.

Four consecutive blocks of ten transition points were taken for each frequency. When the estimated threshold mean of the last two blocks were within 3 dB, the overall threshold was calculated for that frequency from consecutive blocks with mean thresholds within that 3 dB range. If the thresholds of the last two blocks differed by more than 3 dB, sessions were continued until two consecutive blocks with estimated thresholds within a 3-dB range were obtained.

After having completed the collection of threshold data at all test frequencies, mean thresholds were determined from another ten reversals at 2, 4, 8, and 16 kHz in order to estimate the consistency of threshold data over a longer period of time. The time interval between sessions that had been conducted in the first place and the repetition of data collection at a given frequency was about 2 months.

III. RESULTS

The aerial hearing sensitivity of the test subject is presented in Fig. 2. The overall thresholds and the range of transition points for each test frequency are shown in Table I.

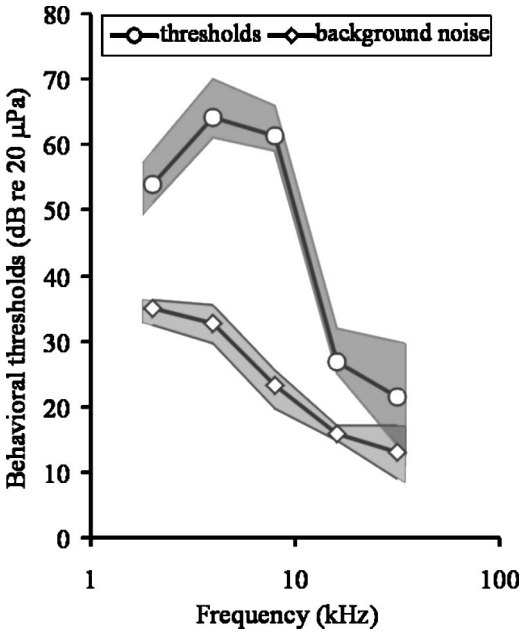


FIG. 2. Aerial hearing in the Sotalia dolphin Paco at five tested frequencies (2.0–31.5 kHz) and the range of transition points (each data point represents at least two consecutive blocks with mean thresholds within a 3-dB range) as well as the average background noise and its fluctuations.

Comparatively high threshold data were obtained from 2 to 8 kHz, with a decreasing performance from 2 to 4 kHz by 10 dB. Mean thresholds at 4 and 8 kHz differ by only 3 dB, while the range of transition points is almost the same (Table I).

Threshold data indicate an increasing hearing sensitivity for frequencies above 8 kHz. While the mean threshold at 16 kHz is already 34 dB lower than that determined for 8 kHz, the lowest threshold was measured at the highest frequency tested (22 dB re 20 μPa at 31.5 kHz). At this frequency the ambient noise level and the range of transition points overlap, indicating that the threshold at 31.5 kHz was affected by masking and that the absolute threshold could be even lower. The overall false alarm rates for Paco at different test frequencies ranged from 12% to 30% (Table I).

Replicated measurements about 2 months later confirmed the threshold data; only at 16 kHz the estimated mean threshold was more than 3 dB higher, but still within the range of transition points obtained before (Table I).

VI. DISCUSSION

The results of this study demonstrate that *Sotalia fluviatilis guianensis* is sensitive to airborne sound as our test subject responded well to all test signal frequencies and auditory thresholds could be obtained. The high threshold at 8 kHz is in accordance with threshold data obtained at this frequency for the same animal under water (Sauerland and Dehnhardt, 1998), thereby indicating that the threshold shift in the lower frequency range represents a characteristic of the hearing system of this dolphin both under water and in air. Sauerland and Dehnhardt (1998) hypothesize that the poor performance of Paco at 8 kHz might be attributed to hearing loss caused by long-term exposure to low-frequency pump noise.

TABLE I. Absolute hearing thresholds of *Sotalia fluviatilis guianensis*.

Test-frequency (kHz)	Mean thresholds (dB *re* 20 μPa)	Range of the transition points (dB *re* 20 μPa)	False alarm rate % of signal absent trials	Mean threshold in repetition tests (dB *re* 20 μPa)
2	54	51–59	20	54
4	64	61–70	12	67
8	61	59–66	12	62
16	27	25–32	30	31
31,5	22	14–30	15	

Only for two other delphinid species, the harbor porpoise and the bottlenose dolphin, are data describing the sensitivity to airborne sound available. The comparison of our data with those obtained for a harbor porpoise (Kastelein *et al.*, 1997) is possible at 2 and 4 kHz where thresholds of the harbor porpoise are 5 and 18 dB lower, respectively. Babushina (1979) tested aerial hearing in the bottlenose dolphin with frequencies up to 110 kHz. As thresholds in the bottlenose dolphin were determined at frequencies differing from those used in the present study, a direct comparison of thresholds is not possible and only general statements can be made. Both audiograms show similarities for frequencies up to 10 kHz (see Fig. 3). The threshold data of the bottlenose dolphin increased from 1 to 4.5 kHz by 13 dB while in the Sotalia the increase amounts to 10 dB from 2 to 4 kHz. Above these frequencies both audiograms show a decrease of thresholds. Thresholds of the bottlenose dolphin decrease from 4.5 to 10 kHz by 24 dB, those of Sotalia by 34 dB from 8 to 16 kHz. The absolute values indicate that the bottlenose dolphin was more sensitive at lower frequencies. In contrast, above 16 kHz the relation is reversed. While Babushina (1979) found a constant decrease of hearing sensitivity for frequencies up to 110 kHz, sensitivity in the Sotalia increased up to the highest frequency tested (31.5 kHz).

To compare the present results with amphibian mammals, the aerial audiograms of a California sea lion (*Zalophus californianus*) (Moore and Schusterman, 1987), a northern fur seal (*Callorhinus ursinus*) (Moore and Schusterman, 1987), and a harbor seal (*Phoca vitulina*) (Møhl, 1968) are included in Fig. 3. The absolute values for frequencies up to 8 kHz show a comparatively low auditory sensitivity of the Sotalia dolphin. The largest differences over all these frequencies exist to the northern fur seal and range from 45 dB at 2 kHz, over 42 dB at 4 kHz to the maximum of 48 dB at 8 kHz. The differences to the California sea lion and the harbor seal are only little less with a maximum of 42 dB for the harbor seal and 45 dB for the sea lion at 8 kHz. At 16 kHz Paco's hearing ability improves and is now in between the sensitivity of the California sea lion and the harbor seal, differing by only 1 dB. The difference from the northern fur seal with 20.5 dB is quite high, but while all the amphibian mammals show a sharp decrease in sensitivity at 32 kHz (the harbor seal already at 22.5 kHz), which indicates a conventional high-frequency limit, the Sotalia improves further.

All audiograms shown in Fig. 3 reveal a notable hearing loss at 4 kHz. Because previously reported audiograms for a California sea lion (Schusterman, 1974) and harp seal (Terhune and Ronald, 1971) did not exhibit this hearing loss, Moore and Schusterman (1987) argued that the differences were most likely due to noise in the earlier studies. However, in our study noise level at 4 kHz was not significantly increased (see Sec. III). Moore and Schusterman (1987) also discuss the possibility that the hearing loss at 4 kHz in their California sea lion and northern fur seal as well as in Møhl's (1968) harbor seal may be caused by differences in the sound route for air versus water media. But the same hearing loss for aerial sound occurs in the Sotalia and the bottlenose dolphin (Fig. 3). The absolute hearing losses from 2 to 4 kHz are comparatively high (10 dB in the Sotalia, 7 dB in the California sea lion and the harbor seal, and 13 dB in the northern fur seal; the bottlenose dolphin showed a hearing loss of 13 dB from 1 to 4.5 kHz). In contrast to pinnipeds,

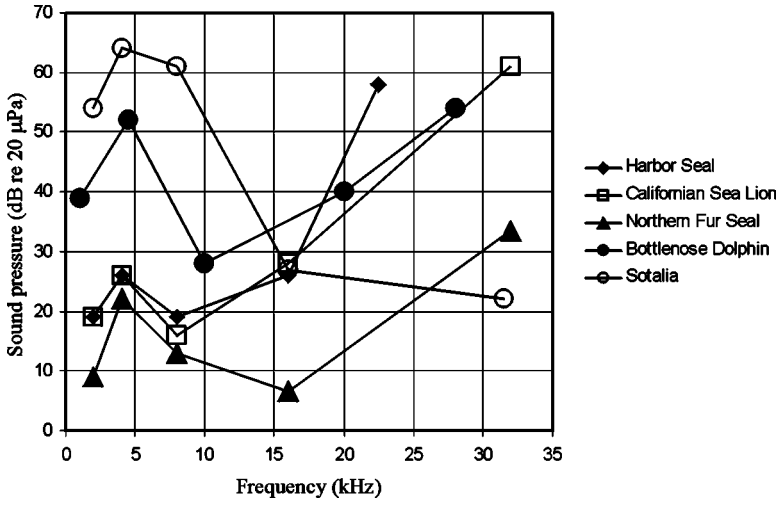


FIG. 3. Thresholds of the Sotalia dolphin to airborne sound (present study) in comparison to those obtained for the bottlenose dolphin (Babushina, 1979) in a comparable frequency range, a California sea lion (Moore and Schusterman, 1987), a northern fur seal (Moore and Schusterman, 1987), and a harbor seal (Møhl, 1968).

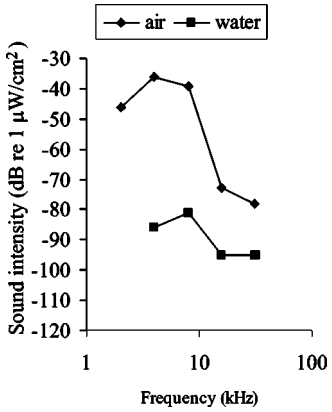


FIG. 4. Thresholds of the Sotalia dolphin to airborne (present study) and underwater sound (Sauerland and Dehnhardt, 1998) compared by sound intensity.

dolphins probably hear by sound transmission through the skin and fatty layers to the inner ear in air as well as in water (Ketten *et al.*, 2000; Dehnhardt, 2002). The fact that the hearing loss at 4 kHz could not be explained by different transmission routes for aerial and underwater sound in dolphins gives reason to question this explanation for pinnipeds as well.

The comparison of our results with the underwater audiogram of the same animal via sound intensities is represented in Fig. 4. Both audiograms decrease sharply at frequencies above 8 kHz. In air this decrease is even more striking with 34 dB/oct compared to 14 dB/oct under water. The sensitivity to airborne sound over all tested frequencies is lower than to underwater sound. The highest difference exists at 4 kHz with 48 dB ($re\ 1\mu W/cm^2$). With increasing frequencies the difference decreases to 20 dB at 16 kHz and 17 dB at 31.5 or 32 kHz, respectively.

If mammalian ears are considered primarily as sound intensity detectors and hearing abilities under water and in air are compared by sound intensities, we have to deal with physical problems. Each reflection boundary like the water surface and the wall of the pool influences the relationship between sound pressure and sound intensity so that the equation $I=p^2/(c*\rho)$ does not apply any more. The amount of the deviation is unknown and different in each experiment.

If we assume, on the other hand, that hearing abilities under water and in air should be compared by sound pressure (cf. Kastak and Schusterman, 1998), we come to somewhat surprising results. Following our understanding of dolphins possessing an excellent auditory system adapted to underwater hearing we would expect the dolphin to be more sensitive to underwater sound, which is not the case for all frequencies in our Sotalia. Threshold curves cross at approximately 10 kHz (graphs not shown). For frequencies up to 8 kHz the test subject still appears to be more sensitive to underwater sound. But the differences are only 14 dB ($re\ 20\ \mu Pa$) at 4 kHz and 6 dB ($re\ 20\ \mu Pa$) at 8 kHz, clearly less than in the comparison by sound intensities. Above 8 kHz the relation reverses. At 16 and 32 kHz the test subject is more sensitive to airborne sound. The difference amounts to 14 dB ($re\ 20\ \mu Pa$) or 19 dB ($re\ 20\ \mu Pa$), respectively, when thresholds are compared by sound pressure. The comparison of aerial and underwater sound detection thresholds via sound pressure for the bottlenose dolphin (Babushina, 1979) shows better hearing abilities under water than in air for most frequencies with two exceptions at 1 and 10 kHz. The comparison of pressure-related hearing thresholds in water and air in the harbor porpoise (Kastelein *et al.*, 1997, 2002) shows only small differences clearly below 10 dB for frequencies up to 4 kHz, while higher frequencies remain to be tested. In summary, comparing the underwater and airborne sound sensitivity of the three delphinid species tested so far using sound pressure as the relevant stimulus parameter shows no general trend for this family, while using sound intensity makes underwater hearing appear superior to aerial hearing.

Our results show that the evolution of auditory and sonar capabilities of *S. fluviatilis guianensis* in water did not result in an inability to perceive sound in air. According to the threshold data, *Sotalia fluviatilis guianensis* is not only sensitive to aerial sound but might also be able to use aerial sound as a means to gain information about its environment, especially in the ultrasonic range. It is remarkable in this respect that *Sotalia fluviatilis guianensis* produces echolocation pulses with one peak at 80–95 kHz and a second peak at 30 kHz (Kamminga *et al.*, 1993). An extension of the data base on aerial auditory and sonar capabilities in dolphins is required to draw a clear picture of the importance of audition in air in a dolphin's life.

ACKNOWLEDGMENTS

We thank the Dolphinarium Münster GmbH for their cooperation. The work reported here was supported by a grant of the ''Volkswagenstiftung'' to GD.

Au, W. W. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).
Babushina, Ye. S. (**1979**). ''Localisation by the dolphin of the source of tonal and pulse signals in water and air,'' Vestr. Leningr. Univ. **3**, 119–121.
Bullock, T. H., Grinnell, A. D., Ikezono, E., Kameda, K., Katsuki, Y., Nomoto, M., Sato, O., Suga, N., and Yanagisawa, K. (**1968**). ''Electrophysiological Studies of Central Auditory Mechanisms in Cetaceans,'' Z. Vergl. Physiol. **59**, 117–156.
Cummings, W. C., Fish, J. F., and Thompson, P. O. (**1971**). ''Bioacoustics of marine mammals of Argentina: R/V Herd Cruise 71-3,'' Antarct. J. U.S. **6**, 266–268.
Dehnhardt, G. (**2002**). ''Sensory systems,'' in *Marine Mammal Biology*, edited by A. R. Hoelzel (Blackwell Science, Malden), pp. 116–141.
Gellerman, L. W. (**1933**). ''Chance orders of alternating stimuli in visual discrimination experiments,'' J. Genet. Psychol. **42**, 206–208.
Gescheider, G. A. (**1985**). *Psychophysics. Method and Theory* (Erlbaum, Hillsdale, NJ).
Herman, L. M., and Forestell, P. H. (**1977**). ''The Hawaiian humpback whale: Behavior,'' Proc. 2nd Conf. Biol. Marine Mammals, San Diego, CA., December 1977, p. 29 (abstract).
Herman, L. M., Peakcock, M. F., Yunker, M. P., and Madsen, C. J. (**1975**). ''Bottlenose Dolphin: Double-split pupil yields equivalent aerial and underwater diurnal acuity,'' Science **189**, 650–652.
Johnson, C. S. (**1968**). ''Relation between absolute threshold and duration of tone pulse in the bottlenosed porpoise,'' J. Acoust. Soc. Am. **43**, 757–763.
Kamminga, C., van Hove, M. T., Engelsma, F. J., and Terry, R. P. (**1993**). ''Investigations on Cetacean Sonar X: A comparative analysis of underwater echolocation clicks of *Inia* spp. and *Sotalia* spp.,'' Aquat. Mammals **19**(1), 31–43.
Kastak, D., and und Schusterman, R. J. (**1998**). ''Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology,'' J. Acoust. Soc. Am. **103**, 2216–2228.

Kastelein, R. A., Au, W. W., and de Haan, D. (**2002**). ''Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals,'' J. Acoust. Soc. Am. **112**, 334–344.

Kastelein, R. A., Nieuwstraten, S. H., Staal, C., van Ligtenberg, C. L., and Versteegh, D. (**1997**). ''Low-frequency aerial hearing of a harbour porpoise (*Phocoena phocoena*),'' in *The Biology of the Harbour Porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall, (De Spil, Woerden, The Netherlands), pp. 295–312.

Ketten, D. R. (**2000**). ''Cetacean Ears,'' in *Hearing by Whales and Dolphins*, edited by W. W. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 43–108.

Layne, J. N. (**1958**). ''Obervations on fresh water dolphins in the upper Amazon,'' J. Mammal. **39**, 1–22.

Levitt, H. (**1971**). ''Transformed Up-Down Methods in Psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

MacAskie, I. V. (**1966**). ''Unusual example of group behavior by killer whales (*Orcinus rectipinna*),'' Murrelet **47**, 38.

Møhl, B. (**1968**). ''Auditory sensitivity of the common seal in air and water,'' J. Aud Res. **8**, 27–30.

Moore, P. W. B., and Schusterman, R. J. (**1987**). ''Audiometric Assessment of northern fur seals (*Callorhinus ursinus*),'' Marine Mammal Sci. **3**(1), 31–53.

Norris, K. S. (**1974**). *The Porpoise Watcher* (Norton, New York).

Pike, G. (**1962**). ''Migration and feeding of gray whale (Eschrichtius gibbosus),'' J. Fish. Res. Board Can. **19**, 815–838.

Sauerland, M., and Dehnhardt, G. (**1998**). ''Underwater audiogram of a tucuxi (*Sotalia fluviatilis guianensis*),'' J. Acoust. Soc. Am. **103**, 1199–1204.

Schusterman, R. J. (**1974**). ''Auditory sensitivity of a California sea lion to airborne sound,'' J. Acoust. Soc. Am. **56**, 1248–1251.

Terhune, J. M., and Ronald, K. (**1971**). ''The harp seal, *Pagophilus groenlandicus* (Erxleben, 1777), X. The air audiogram,'' Can. J. Zool. **49**, 385–390.

Yost, W. A. (**1994**). *Fundamentals of Hearing: An Introduction* (Academic, New York).

Are high perches in the blackcap *Sylvia atricapilla* song or listening posts? A sound transmission study

Nicolas Mathevon[a)]
Equipe "Communications Acoustiques," NAMC CNRS UMR 8620, Université Paris-Sud-Orsay and LBA Université Jean Monnet-Saint-Etienne, 23 rue Michelon, 42023 Saint-Etienne cedex 2, France

Torben Dabelsteen and Sandra H. Blumenrath
Department of Animal Behaviour and Centre for Sound Communication, Zoological Institute, University of Copenhagen, Tagensvej 16, DK-2200 Copenhagen N, Denmark

(Received 18 May 2004; revised 4 October 2004; accepted 18 October 2004)

Birds often sing from high perches referred to as song posts. However, birds also listen and keep a lookout from these perches. We used a sound transmission experiment to investigate the changes for receiving and sending conditions that a territorial songbird may experience by moving upwards in the vegetation. Representative song elements of the blackcap *Sylvia atricapilla* were transmitted in a forest habitat in spring using a complete factorial design with natural transmission distances and speaker and microphone heights. Four aspects of sound degradation were quantified: signal-to-noise ratio, excess attenuation, distortion within the sounds determined as a blur ratio, and prolongation of the sounds with ''tails'' of echoes determined as a tail-to-signal ratio. All four measures indicated that degradation decreased with speaker and microphone height. However, the decrease was considerably higher for the microphone than for the speaker. This suggests that choosing high perches in a forest at spring results in more benefits to blackcaps in terms of improved communication conditions when they act as receivers than as senders. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1828805]

PACS numbers: 43.80.Lb, 43.80.Ev [JAS] Pages: 442–449

I. INTRODUCTION

In songbirds loud advertising songs typically function in establishment of territories and repellence of competitors as well as in female attraction and stimulation (Kroodsma and Byers, 1991; Catchpole and Slater, 1995). Advertising songs may travel over long distances and hence may constitute an energy saving substitute for movements across and within territories and for visual behavior that may be screened by obstacles. However, advertising songs are also subject to attenuation and modification during propagation, especially through forests where the vegetation constrains sound propagation (Wiley and Richards, 1978, 1982; Blumenrath and Dabelsteen, 2004). Spherical spreading and absorption associated with multiple scattering by the vegetation attenuate sounds and hence reduce their signal-to-noise ratio (Michelsen, 1983). The duration of silences between sound elements is modified as a result of sound reverberation that elongates sounds with tails of echoes (Holland *et al.*, 2001a). Furthermore, selective frequency filtering, reverberation and atmospheric turbulences distort amplitude and frequency patterns over time, giving recorded sounds a blurred appearance (Wiley and Richards, 1982; Dabelsteen *et al.*, 1993).

Songbirds seem to have evolved different strategies to counteract degradation. For instance, sounds used for long-range communication may be fitted to the properties of the transmission channel by their frequencies (Morton, 1975; Dabelsteen *et al.*, 1993; Nemeth *et al.*, 2001; Mathevon *et al.*, 2004). It is also believed that active choice of sender and receiver positions may indicate an adjustment to the heterogeneous forest environment made of various superposed layers of vegetation with different conditions for sound propagation. For instance, the wren *Troglodytes troglodytes* perches in response to playback of propagation-degraded song, probably to improve ranging conditions (Mathevon and Aubin, 1997; Holland *et al.*, 2001b).

Sound propagation experiments have shown that the birds' choices of high perches for song posts may counteract sound degradation and thereby increase the transmission range of their songs (Mathevon *et al.*, 1996). However, perched birds also observe and listen to conspecifics, and previous transmission studies suggest that perching in a forest habitat might sometimes improve signal reception more than signal transmission (Dabelsteen *et al.*, 1993; Holland *et al.*, 1998). Both of these studies were made in late autumn (mid-November and early December) to simulate conditions in very early spring long time before leaf burst where only a few species sing. However, the environmental conditions are different in April–May when the singing activity of most species culminates, for instance by having higher temperatures, quickly developing ground covering and leaf burst. These differences make it difficult to assess the external validity of the results from late autumn in terms of how well they predict sound transmission and reception during spring. Besides this, in a recent study based on sound propagation experiments in two different Australian forests, Padgham (2004) shows first that attenuation is reduced more by increases in receiver height than in source height, and, second, that the height of the source seems to be consistently more effective at reducing the impact of reverberation than any

a)Electronic mail: mathevon@univ-st-etienne.fr

change in receiver height. Thus, the question of the importance of perching appears more complex than it was originally thought and more investigations are needed to clarify the question.

The objective of the present study is to test the relative influence of perching for the sender and the receiver roles at the seasonal peak time for singing in the forest, using migratory blackcap, *Sylvia atricapilla*, a warbler known to use high perches during singing. We made a sound transmission experiment at the start of trees' leaf burst when the species returns from migration and sings intensively. Loudspeaker and microphone height were varied to simulate different sender and receiver heights, respectively, and sound elements from the advertising song of the species were used as test sounds. To obtain representative knowledge of the changes that occur during transmission, we measured four aspects of sound degradation: signal-to-noise ratio, attenuation in excess to that predicted by geometric spreading, tail-to-signal ratio which is a measure of the reverberations filling the inter note silences, and blur ratio which expresses the distortions of the patterns of amplitude and frequency modulations over time.

II. METHODS

The experiment was made 17–18 and 20–22 April 1998 under good weather conditions (temperature around 10 °C; relative humidity around 70%; almost no wind) in a typical blackcap habitat, a mixed primeval deciduous forest at Strødam Biological Field Station (Denmark). This period corresponds to the time where male blackcaps start singing, set their territories and attract their sexual partner. The test site, which was dominated by 20-m-high trees (diameter 0.3–0.4 m at 1 m above the ground), had patches of undergrowth of bushes and young trees of various heights and species, and a ground covering of herbs, dead branches and a few fallen trees. The canopy started 6–9 m above the ground. Most of the undergrowth had already come into full leaf whereas leaf burst had just started in the canopy of the high trees.

The blackcap advertising song starts with relatively high-pitched twitter sounds (T-sounds) and ends with louder, more low frequency, species characteristic motif sounds (M-sounds) (Cramp, 1992). We selected the same five representative twitter (T1–T5) and motif (M1–M5) sound elements used in a previous transmission experiment [see Fig. 1 in Dabelsteen and Mathevon (2002)]. The test sounds, which came from high quality recordings of advertising song made at a distance of 3 m from singing males with a Nagra IIIB tape recorder and an omnidirectional Sennheiser MD211N microphone, were digitized through an OROS acquisition board equipped with an anti-aliasing filter (120 dB/oct) and isolated using the SYNTANA signal analysis package (Aubin, 1994). The DAT test tape contained five repetitions of a series of the five M-sounds followed by five repetitions of a series of the five T-sounds, with 1-s pauses between successive sounds. At the start and end of the tape, we inserted 1.5 s of silence.

The test tape was played back from a SONY TCD-D10PRO DAT recorder connected to a DENON DCA-600 power amplifier, a high-pass filter (1.2 kHz, $f_{-3\mathrm{dB}}$) and a VIFA 1-in. Neodymium tweeter (Larsen and Dabelsteen, 1997). T- and M-sounds were played back at a natural sound pressure levels (SPL) of respectively 78 and 85 dB(A) at 1 m. The calibration of the SPL was made with a Brüel & Kjær sound level meter (type 2236) equipped with a $\frac{1}{2}$-in. condenser microphone (type 4188) covered with a windscreen (type UA 0459). The test sounds were rerecorded through the same sound level meter connected to a HHB PORTADAT PDR 1000 DAT recorder. Loudspeaker and microphone were mounted on telescopic masts (Clark, type QT 15N/HP) pointing towards each other.

The ten sounds were transmitted along three representative transects over 12.5, 25, and 50 m considering that the diameter of a blackcap territory is about 50 m (Cramp, 1992). The loudspeaker (4 and 9 m) and the microphone (2, 4, and 9 m) heights were chosen to represent natural perches. The experiment followed a complete factorial design with five replications of 54 transmission pathways: 3 transects $\times 3$ distances $\times 2$ speaker heights $\times 3$ microphone heights.

As playback equipment modifies test sounds, we rerecorded the sounds in an open area at a distance of 1.5 m from the loudspeaker with both loudspeaker and microphone 9 m above the ground. This recording provided the *model sounds*, one for each sound type. The sounds transmitted over 12.5–50 m constituted the *observation sounds*. The degradation of the observation sounds was quantified by comparing them with their respective model sounds, which had only suffered negligible degradation. Following Dabelsteen and Mathevon (2002) we selected for analysis the first two replications of each observation sound and transmission pathway that were neither corrupted by wind noise, nor masked by transient external noise such as vocalizations of other birds. For each pathway we also measured the stationary background noise in a 1-s segment taken from one of the 1.5-s pauses at the start or the end of the recording.

Models, observation sounds and segments of background noise were band-pass filtered at 1–9 kHz (Standford Research System, type SR650, 115 dB/oct) and digitized through an acquisition board (Signal Data DSP SPB2 signal processor board, $\Delta f = 22\,050$ Hz). The digitized sounds were then band-pass filtered using sound-specific filter bandwidths that matched the frequency range of each of the ten test sounds (sound, frequency range in kHz; T1: 3.7–8.7; T2: 2.1–6.6; T3: 2.3–4.7; T4: 2.2–8.0; T5: 3.9–7.7; M1: 1.7–3.3; M2: 2.1–3.7; M3: 2.5–3.5; M4: 2.1–3.3; M5: 2.4–6.2). The filtering and the subsequent quantification of four aspects of degradation were made with SIGPRO ver. 1.4 (Pedersen, 1998).

The *signal-to-noise ratio (SNR)* was calculated from the energies of the observation sound (Ey) and the background noise (En), both over the duration of the model sound, as $\mathrm{SNR} = 10 \log ((\mathrm{Ey}-\mathrm{En})/\mathrm{En})$. The *tail-to-signal ratio (TSR)*, which expresses the amount of elongation with tails of echoes, was calculated from Ey and the energy of the tail of echoes (Et) as $\mathrm{TSR} = 10 \log (\mathrm{Et}/\mathrm{Ey})$. The *excess attenuation (EA)*, which is the attenuation of the observation sound in excess to the 6 dB per doubling of distance predicted by spherical spreading, and the *blur ratio (BR)*, which expresses

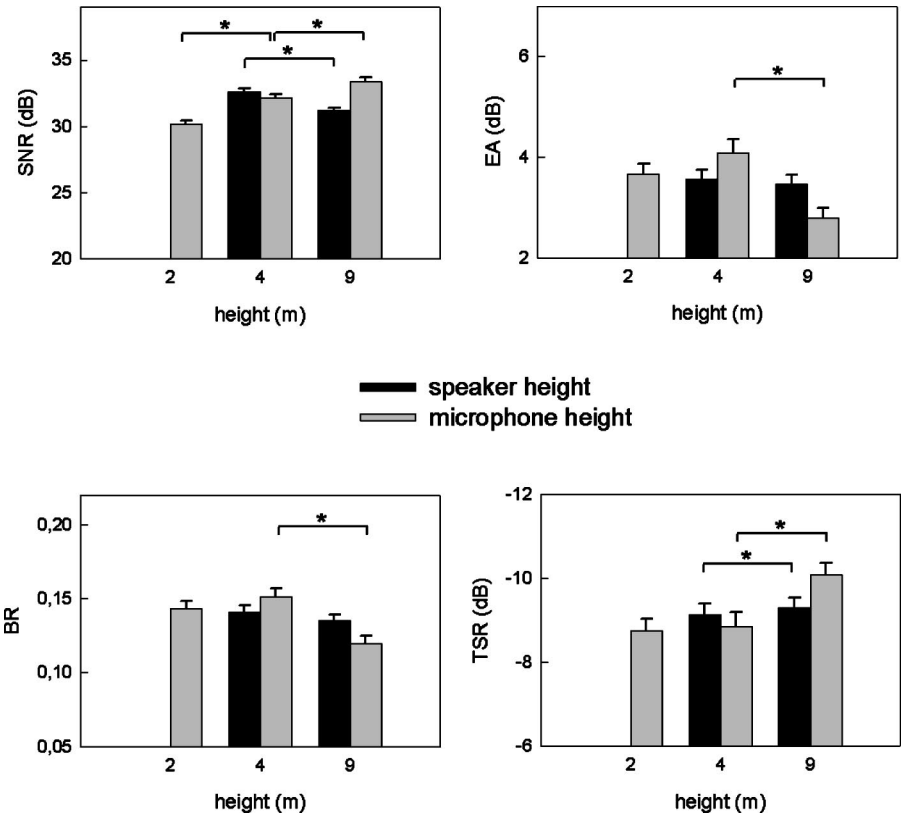


FIG. 1. Main effects (mean ±s.e.) of speaker and microphone heights on signal-to-noise ratio (SNR), excess attenuation (EA), blur ratio (BR) and tail-to-signal ratio (TSR) of M-sounds. Asterisks indicate significant differences, see text.

the distortion in amplitude and frequency patterns over time, were derived from the relation of the amplitude functions (AF) of model and observation sounds. The energy of the observation AF (Ey_{AF}) relative to the energy of the model AF determines the factor k_{AF} that is used to attenuate the model sound to the same energy as the observation sound. EA was then calculated as $EA = -20 \log k_{AF}\text{-}A$, where A is the attenuation caused by spherical spreading. The energy of the difference signal (Ex_{AF}) between the observation AF and the k_{AF}-attenuated model AF expresses the blurring of the observation sound. BR was calculated as $BR = Ex_{AF}/Ey_{AF}$. The detailed protocol of the signal analysis is reported elsewhere (Dabelsteen *et al.*, 1993; Holland *et al.*, 1998, 2001b). Statistics were made with the Statgraphics package, STSC Inc.

The data obtained for M- and T-sounds were analyzed separately and subjected to a multifactor ANOVA as follows: 3 transects $\times$ 5 sounds $\times$ 3 distances$_{(12.5\text{ m},25\text{ m},50\text{ m})}$ $\times$ 2 speaker heights$_{(4\text{ and }9\text{ m})}$ $\times$ 3 microphone heights$_{(2,4\text{ and }9\text{ m})}$ with 2 replications. This should give a sample size of 540. However, the data set was lower because the two envelope-based measures, EA and BR, could not be quantified for one of the T-sounds that is rich in harmonics (Dabelsteen and Mathevon, 2002), and because background noise corrupted especially the tails of echoes contributing to the TSR measure. For M-sounds, this resulted in data sets of 534 (EA, BR and SNR) and 506 (TSR); for T-sounds data sets of 420 (EA and BR), 526 (SNR) and 432 (TSR). Prior to the statistical analysis, the data for TSR, EA and BR were respectively $\log 10$, square root $(+3)$ and $\log 10$ transformed to cope with the requirements for parametric tests. Main and two-factor interaction effects were considered, and 95% LSD confidence limits were used for *post hoc* comparisons.

III. RESULTS

The effect of the propagation distance explains most (31%–62.6%) of the total variation for all measured parameters. The signal-to-noise ratio (SNR) decreases, and excess attenuation (EA), blur ratio (BR) and tail-to-signal ratio (TSR) increase with distance. The effects of transect and test sound explain a large part of the variation for some of the measures, 5.7%–26.0% and 0.9%–21.6%, respectively. Transect, distance and test sound have a significant effect on all of the degradation measures (F-ratio = 4.3–1356.8, df = 1–4, $p < 0.02$–$0.000\,01$).

The microphone and the speaker height explain less of the total variation than the other factors. However, there is a striking difference between their effects: the microphone height explains 1.7%–3.2% (M-sounds) and 1.3%–4.1% (T-sounds) of the total variation, the speaker height only 0.01%–0.9% (M-sounds) and 0.4%–1.5% (T-sounds). Whereas the effect of microphone height is significant for all degradation measures (F-ratio = 13.5–62.7, df = 2, $p < 0.000\,01$), this only applies to some of the measures for speaker height (F-ratio = 5.74–36.6, df = 2, $p < 0.017$–$0.000\,01$). Increasing the microphone height, especially from 4 to 9 m, decreases degradation in all aspects for both sound types (Figs. 1 and 2): SNR increases, BR, EA and TSR decrease. Increasing the speaker height has a somewhat different and smaller effect, especially for M-sounds where neither EA (F-ratio = 0.54, df = 1, $p = 0.46$) nor BR (F-ratio = 3.71, df = 1, $p = 0.055$) is affected (Figs. 1 and 2).

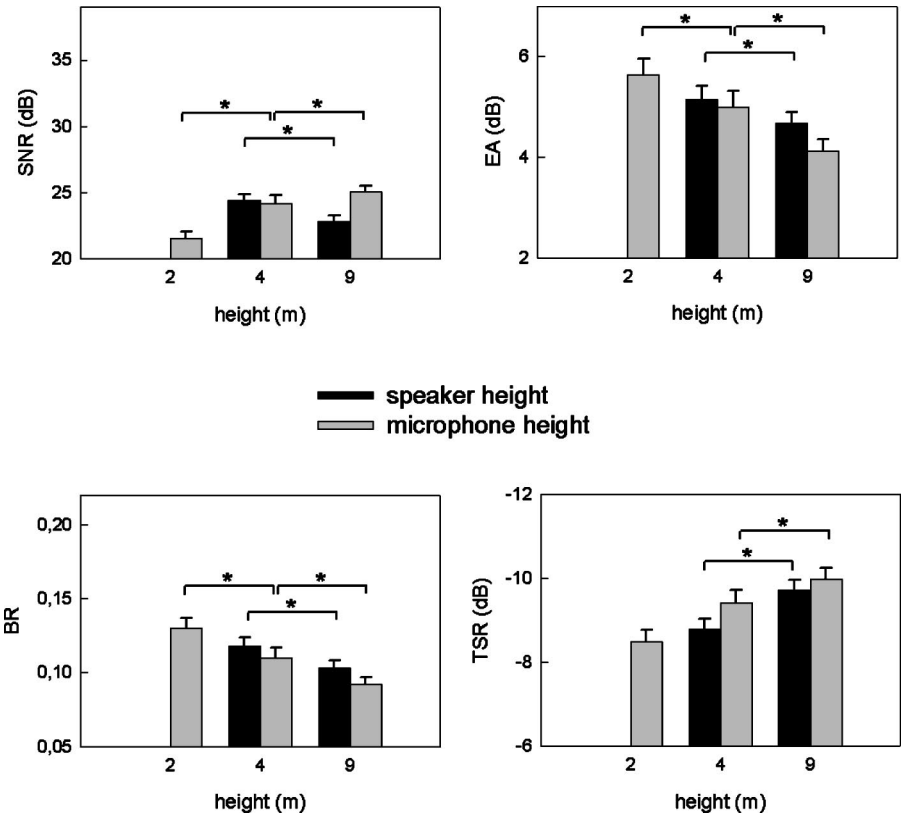


FIG. 2. Main effects (mean±s.e.) of speaker and microphone heights on SNR, EA, BR and TSR of T-sounds. Terminology as in Fig. 1.

Moreover, an increase in speaker height increases degradation with respect to SNR for both M- and T-sounds.

Overall, the two-factor interactions explain only a small part of the variation, and as expected from previous studies, distance by transect is the most important two-factor interaction (F-ratio=2.52–42.5, df=4, $p<0.04$–$0.000\,01$). Those where distance and microphone height are one of the two factors are also of some importance, for instance microphone by distance where the interaction effect comes from the variation in which of the microphone heights 2 and 4 m causes most degradation (Figs. 3 and 4). Unlike this, the influence of speaker by distance is negligible (Figs. 3 and 4).

To investigate what the effects of perching correspond to in terms of virtual horizontal approaches that would give the same effects, we followed the method described in Holland *et al.* (1998) which allows us to relate the measured difference in the degradation values to the horizontal distance that would have produced a comparable change. We thus looked at the average effects of increasing sender (speaker) and receiver (microphone) height from 4 to 9 m when the distance between sender and receiver is 50 m, i.e., one territory diameter. We first estimated regression lines with log propagation distance for each of the four degradation measures for both M- and T-sounds ($degradation = a + b \times log\ distance$), and calculated what the degradation for each measure would be at 50 m according to these regressions. The eight regressions were then used to translate the measured average effects of perching into the virtual horizontal approach distances that would have produced the same effects (*degradation at 50 m−measured average effect of* $perching = a + b \times log$ *new distance*; *50 m−new* $distance = virtual$ *horizontal approach*). The virtual horizontal approach distances, which are shown in Fig. 5, accentuate the clear differences between the effects of speaker and microphone elevation. All but one of the elevations of loudspeaker and microphone from 4 to 9 m is equivalent to a horizontal approach to respectively another receiver or sender. The one exception is for SNR of the T-sounds where an elevation of the speaker height from 4 to 9 m is equivalent to a horizontal retreat. Moreover, elevating the microphone from 4 to 9 m corresponds to a much larger approach distance than the same elevation of the loudspeaker, especially for M-sounds (Fig. 5). The difference is less pronounced for T-sounds, and for one measure, TSR, elevation of the speaker has a larger effect than elevation of the microphone. For the M-sounds, the virtual horizontal approach derived from the effect of elevating the speaker 5 m is comparable to this elevation height, whereas the virtual approach derived from the elevation of the microphone is considerably larger corresponding to half a territory diameter. For the T-sounds, elevation of the speaker has a larger effect than for the M-sounds, but this effect does not consistently correspond to a virtual approach, and the elevation of the microphone has a smaller effect than for the M-sounds (Fig. 5).

IV. DISCUSSION

The present study uses experimental transmission of blackcap song to test whether high perching in songbirds is equally profitable for receivers and singers. For the first time this is done at the natural time in spring when the singing activity of the actual test species peaks, and by quantifying aspects of all of the main types of transmission-induced sound modifications. In terms of improvements in signal

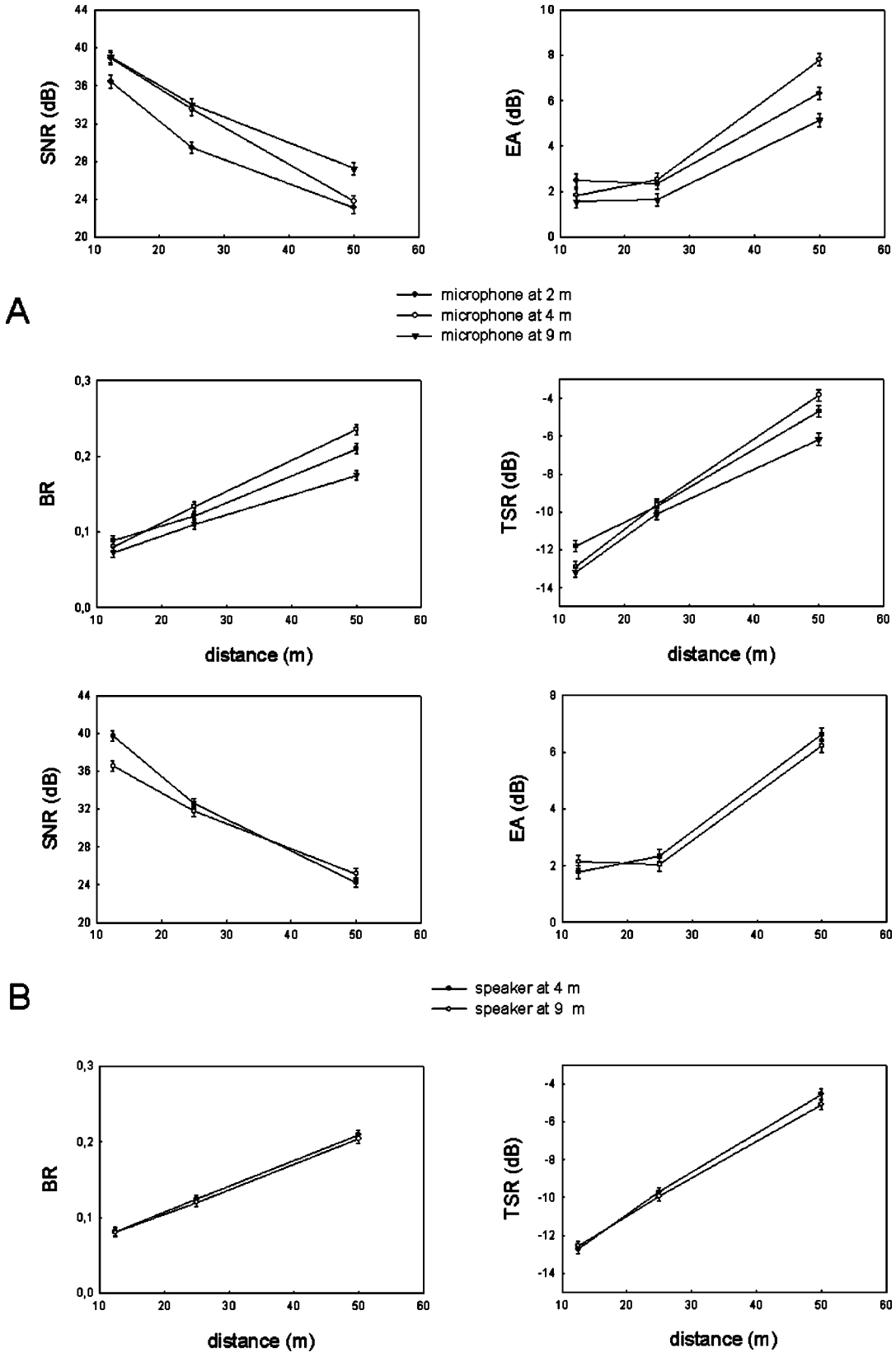


FIG. 3. Interaction effects between distance and microphone (A) or speaker (B) heights on SNR, EA, BR and TSR of M-sounds. Terminology as in Fig. 1.

transmission that can be obtained by moving upwards in the vegetation at the time of territory establishment, the results show that high perching above the undergrowth is more important to a blackcap that acts as a receiver than to one that acts as a sender. Although previous studies made at a more unnatural time of the year suggested that this might be the case (Dabelsteen *et al.*, 1993; Holland *et al.*, 1998), this is the first study that clearly demonstrates this difference in benefits for the sender and the receiver roles.

Elevating the microphone from 4 to 9 m above ground

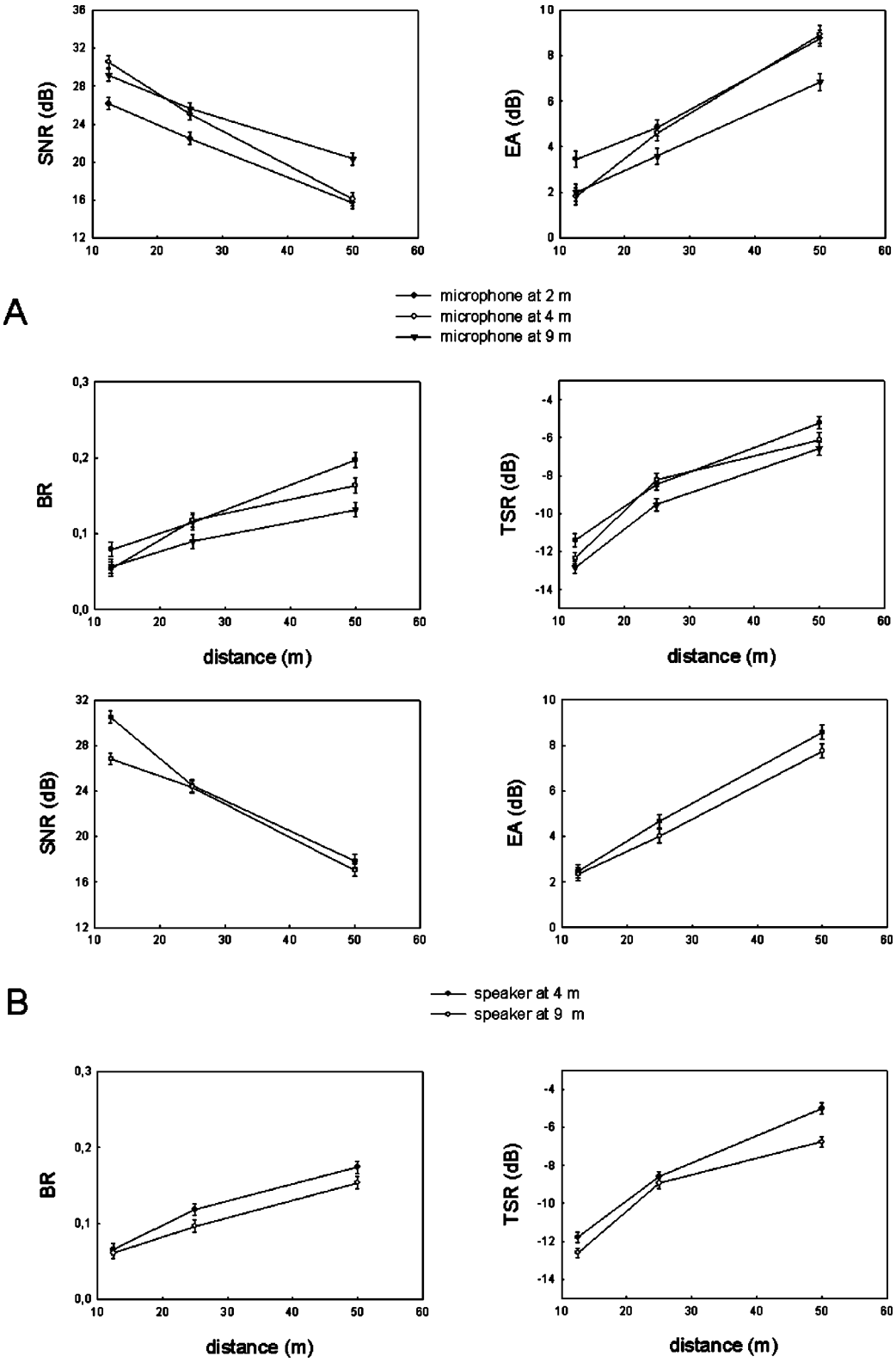


FIG. 4. Interaction effects between distance and microphone (A) or speaker (B) heights on SNR, EA, BR and TSR of T-sounds. Terminology as in Fig. 1.

level improves all of the degradation measures for both motif (M) and twitter (T) sounds. Elevating the microphone from 2 to 4 m gives a more varied effect, probably because 4 m is still inside the undergrowth, most of which had already developed leaves. The excess attenuation (EA) and the blur ratio (BR) of the received sounds are considerably reduced by perching at 9 m. A low EA is always important to ensure long range communication in a noisy forest environment during spring, and a low BR may be of primary importance for different communication processes, for instance in the blackcap where the species-specific decoding relies heavily on aspects of frequency modulation, cues for which BR can be a good indicator of degradation (Mathevon and Aubin, 2001). The reduction in degradation obtained by elevating the mi-

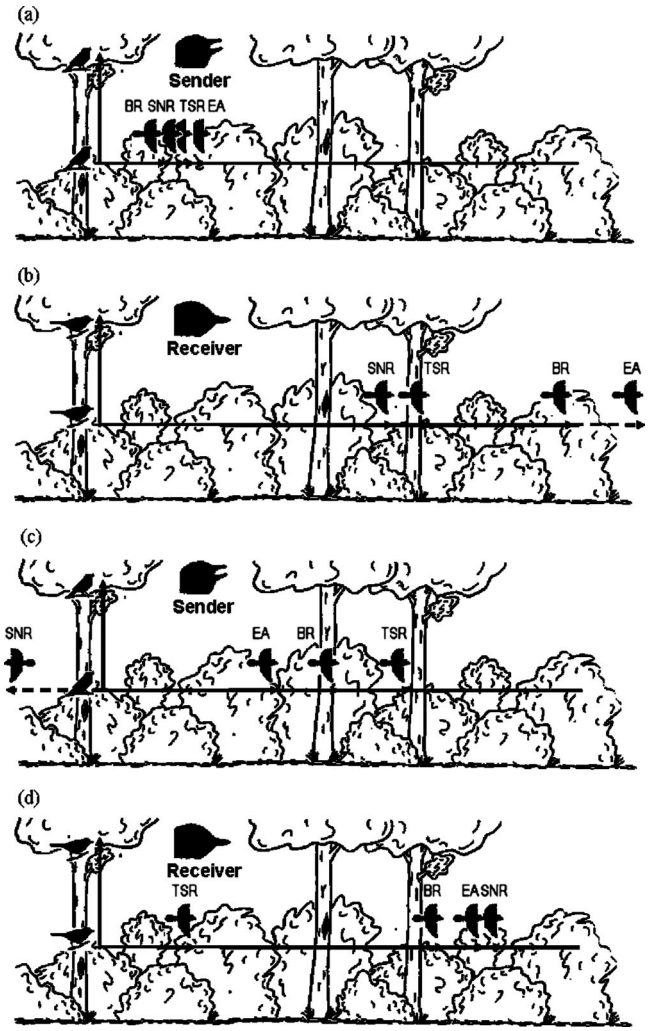


FIG. 5. Effect of increasing the perch height from 4 to 9 m on the communication range of blackcap song: (a) and (b) consider M-sounds, (c) and (d) T-sounds. The changes in sound degradation (usually a reduction), which a sender or receiver obtains by increasing (vertical arrows) its perch height from 4 to 9 m, are translated into virtual horizontal distances (horizontal arrows) that it would have to approach respectively a receiver or a sender to obtain equivalent changes in degradation. Each horizontal arrow (indicated by a "flying" bird) denotes one aspect of degradation: SNR = signal-to-noise ratio; EA = excess attenuation; BR = blur ratio; TSR = tail-to-signal ratio. The initial distance between sender and receiver is assumed to be 50 m. Horizontal lines indicate half a territory diameter (i.e., 25 m); dashed lines represent an approach longer than 25 m (b) or a retreat (c). All movements are to scale. (See text for further details.)

crophone to 9 m also applies to the tail-to-signal ratio (TSR), at least as regards M-sounds. A low TSR is important because the tails of echoes may interfere with the song elements by forward masking and also fill in interelement pauses in the songs (e.g., Holland *et al.*, 2001b). Altogether, the present results support that a songbird receiver should choose high perches to optimize sound signal reception.

The improvements obtained by a high speaker position seem far less pronounced. Indeed, the influence of speaker height was apparent only for T-sounds. For T-sounds, the tail-to-signal ratio (TSR), the excess attenuation (EA) and the blur ratio (BR) significantly decreased with speaker height. As M-sounds constitute the most powerful sounds of the blackcap song and hence those presenting the largest capacity for long-range propagation, these results indicate that the choice of a high song post may allow a blackcap male to only slightly increase the active space of its songs. Conversely, T-sounds are often sung alone in the undergrowth for long periods (Cramp, 1992, personal observations). The increased degradation of T-sounds resulting from the choice of a low song post may allow males to impair the decoding of this kind of signal by an undesirable audience, for example potential eavesdroppers in the context of courtship (e.g., Dabelsteen *et al.*, 1998; Balsby and Dabelsteen, 2003).

Although the physical mechanisms behind the differential effects of emitter and receiver positions on sound transmission are unclear, these effects are likely to be explained by the propagation path of the emitted sounds. Down at 4 m the density of the vegetation is higher and the tree trunks have larger diameters than at 9 m, which is well above the undergrowth and at the underside of or just inside the canopy. At mid-spring when this experiment was made the process of bursting into leaf has just started in the canopy whereas it is almost completed in the undergrowth. Altogether, this should cause more filtering, scattering and reverberation at 4 m than at 9 m (e.g., Wiley and Richards, 1978, 1982).

ACKNOWLEDGMENTS

We thank Claude Chappuis who has kindly given us the high quality recordings used in this study, and Thorsten Balsby for commenting on earlier drafts of the manuscript. Center for Sound Communication and The Danish National Research Foundation have funded the study. The experiments comply with the current laws of Denmark.

Aubin, T. (**1994**). "Syntana: a software for the synthesis and analysis of animal sounds," Bioacoustics **6**, 80–81.

Balsby, T. J. S., and Dabelsteen, T. (**2003**). "Male calling between courtship sequences in whitethroats: a way to counter intrusions from neighboring rivals," Behav. Processes **63**, 149–157.

Blumenrath, S. H., and Dabelsteen, T. (**2004**). "Degradation of great tit *Parus major* song before and after foliation: Implications for vocal communication in deciduous forests," Behavior 8, 935–958.

Catchpole, C. K., and Slater, P. J. B. (**1995**). *Bird Song. Biological Themes and Variations* (Cambridge U. P., Cambridge, UK).

Cramp, S. (**1992**). *The Birds of the Western Palearctic Vol. VI* (Oxford U. P., Oxford).

Dabelsteen, T. (**1984**). "An analysis of the full song of the Blackbird *Turdus merula* with respect to message coding and adaptations for acoustic communication," Ornis. Scand. **15**, 227–239.

Dabelsteen, T., and Mathevon, N. (**2002**). "Why do songbirds sing intensively at dawn? A test of the acoustic transmission hypothesis," Acta Ethologica **4**, 65–72.

Dabelsteen, T., Pedersen, S. B., and Larsen, O. N. (**1993**). "Habitat-induced degradation of sound signals: quantifying the effects of communication sounds and bird location on blur ratio, excess attenuation and signal-to-noise ratio," J. Acoust. Soc. Am. **93**, 2206–2220.

Dabelsteen, T., McGregor, P. K., Lampe, H. M., Langmore, N. E., and Holland, J. (**1998**). "Quiet song in songbirds: an overlooked phenomenon," Bioacoustics **9**, 89–105.

Holland, J., Dabelsteen, T., Pedersen, S. B., and Larsen, O. N. (**1998**). "Degradation of song in the wren *Troglodytes troglodytes*: Implications for information transfer and ranging," J. Acoust. Soc. Am. **103**, 2154–2166.

Holland, J., Dabelsteen, T., Bjørn, C. P., and Pedersen, S. B. (**2001a**). "The location of ranging cues in wren song: evidence from calibrated interactive playback experiments," Behaviour **138**, 189–206.

Holland, J., Dabelsteen, T., Pedersen, S. B., and Paris, A. L. (**2001b**). ''Potential ranging cues contained within the energetic pauses of transmitted wren song,'' Bioacoustics **12**, 3–20.
Kroodsma, D. E., and Byers, B. E. (**1991**). ''The function(s) of bird song,'' Am. Zool. **31**, 318–328.
Larsen, O. N., and Dabelsteen, T. (**1997**). ''The 1 Vifa Neodymium Tweeter: a versatile speaker for playback experiments,'' Bioacoustics **8**, 323–326.
Mathevon, N., and Aubin, T. (**1997**). ''Reaction to conspecific degraded song by the wren *Troglodytes troglodytes*: Territorial response and choice of song post,'' Behav. Processes **39**, 77–84.
Mathevon, N., and Aubin, T. (**2001**). ''Sound-based species-specific recognition in the blackcap *Sylvia atricapilla* shows high tolerance to signal modifications,'' Behaviour **138**, 511–524.
Mathevon, N., Aubin, T., and Dabelsteen, T. (**1996**). ''Song degradation during propagation: importance of song post for the wren Troglodytes troglodytes,'' Ethology **102**, 397–412.
Mathevon, N., Aubin, T., Dabelsteen, T., and Vielliard, J. M. E. (**2004**). ''Are communication activities shaped by environmental constraints in reverberating and sound-absorbing forest habitats ?'' in *Advances in Bioacoustics, vol. I,* edited by J. M. E. Vielliard, M. L. Da Silva, and R. A. Suthers (Academia Brasileira de Ciencias, Rio de Janeiro), pp. 259–263.
McGregor, P. K. (**1994**). ''Sound cues to distance: the perception of range,'' in *Perception and Motor Control in Birds*, edited by P. R. Green and M. Davis (Springer, Berlin), pp. 74–94.
Michelsen, A. (**1983**). ''Biophysical basis of sound communication,'' in *Bioacoustics, a comparative approach*, edited by B. Lewis (Academic, New York), pp. 3–38.
Morton, E. S. (**1975**). ''Ecological sources of selection of avian sounds,'' Am. Nat. **109**, 17–34.
Nemeth, E., Winkler, H., and Dabelsteen, T. (**2001**). ''Differential degradation of antbird songs in a Neotropical rainforest: Adaptation to perch height?'' J. Acoust. Soc. Am. **110**, 3263–3274.
Padgham, M. (**2004**). ''Reverberation and frequency attenuation in forests—implications for acoustic communication in animals,'' J. Acoust. Soc. Am. **115**, 402–410.
Pedersen, S. B. (**1998**). ''Preliminary operation manual for signal processor SIGPRO, ver.1.3.,'' Center for Sound Communication, Odense University.
Wiley, R. H., and Richards, D. G. (**1978**). ''Physical constraints on acoustic communication and the atmosphere: Implications for the evolution of animal vocalizations,'' Behav. Ecol. Sociobiol. **3**, 69–94.
Wiley, R. H., and Richards, D. G. (**1982**). ''Adaptations for acoustic communication in birds: Sound transmission and signal detection,'' in *Acoustic Communication in Birds, Vol. 1*., edited by E. H. Kroodsma and D. E. Miller (Academic, New York), pp. 131–181.

Iteration of transmit-beam aberration correction in medical ultrasound imaging

Svein-Erik Måsøy
Department of Circulation and Imaging, Norwegian University of Science and Technology, Trondheim, 7489 Norway

Trond Varslot
Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, 7489 Norway

Bjørn Angelsen
Department of Circulation and Imaging, Norwegian University of Science and Technology, Trondheim, 7489 Norway

(Received 14 April 2004; revised 26 September 2004; accepted 30 September 2004)

Simulations of iterative transmit-beam aberration correction using a time-delay and amplitude filter have been performed to study the convergence of such a process. Aberration in medical ultrasonic imaging is usually modeled by arrival-time and amplitude fluctuations concentrated on the transducer array. This is an approximation of the physical aberration process, and may be applied to correct the transmitted signal using a time-delay and amplitude filter. Estimation of such a filter has proven difficult in the presence of severe aberration. Presented here is an iterative approach, whereby a filter estimate is applied to correct the transmit-beam. This beam induces acoustic backscatter better suited for arrival-time and amplitude estimation, thus facilitating an improved filter estimate. Two correlation-based methods for estimating arrival-time and amplitude fluctuations in received echoes from random scatterers were employed. Aberration was introduced using eight models emulating aberration produced by the human abdominal wall. Results show that only a few iterations are needed to obtain corrected transmit-beam profiles comparable to those of an ideal aberration correction filter. Furthermore, a previously developed focusing criterion is found to quantify the convergence accurately. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1823213]

PACS numbers: 43.80.Vj, 43.58.Ry, 43.20.Fn [FD] Pages: 450–461

I. INTRODUCTION

Aberration in medical ultrasound imaging is observed as reduced resolution in the images. It is mainly produced by spatial variation of acoustic parameters (mass density and bulk compressibility) in the human body wall. The loss of resolution may, in many situations, render a reliable diagnosis based on these images difficult to obtain. Extensive research has therefore been carried out in order to solve this problem.

Iteration of transmit-beam aberration correction is defined as a process where a set of aberration parameters is estimated; the estimated parameters are used for correcting the transmitted ultrasound beam; and a new estimate of the same parameters is calculated. This process is then repeated. The parameters are typically arrival-time or arrival-time and amplitude fluctuations. Iteration of transmit-beam aberration correction is sometimes referred to as adaptive imaging or autofocusing, but these terms are also used to denote aberration correction in general.

In the presented work, iteration of transmit-beam aberration correction is studied. This process is abbreviated *transmit-beam iteration* in the remainder of the article.

Flax and O'Donnell[1,2] studied transmit-beam iteration using estimated arrival-time differences between neighbor elements on the receiving array. They considered aberration from a thin phase-screen just in front of the array. Using such aberration, a transmit-beam iteration process—for correcting the phase of the transmit signal using time-delays estimated from neighbor correlation—was argued to be a process that inherently converges to an ideal transmit focus.

In Refs. 3 and 4, the morphology of the abdominal wall was studied. It was found that a single time-delay or phase-screen is not adequate for modeling aberration of the ultrasound wave. This is due to the fact that aberration consists of both phase and amplitude aberration, and that these effects occur throughout the whole thickness of the body wall.[4] In this situation, the arguments of Flax and O'Donnell[2] are not sufficient. In Ref. 5, it was shown that an appropriate time-delay and amplitude filter can produce close-to-ideal correction. It has yet to be shown that iterative transmit-beam aberration correction based on estimating a time-delay and amplitude filter from random scatterers will yield a similar correction.

In Refs. 6–8, transmit-beam iteration was performed using different methods for aberration correction, but no consistent measure of convergence was introduced. In Ref. 6 several iterations were performed in order to estimate phase aberrations only. The efficiency of the correction was evaluated using the root-mean-square (rms) difference between the estimated phase and a reference phase, where the refer-

ence was obtained from a beacon signal (point source). In many practical situations such a beacon signal is not available. This metric is therefore not useful for evaluating convergence of transmit-beam iteration in most imaging situations.

Rigby *et al.*[9] performed *in vivo* transmit-beam iteration using time-delays with a 1.75D array. They used a beamsum-channel correlation method for estimating arrival-time fluctuations and found the algorithm to converge after three or four iterations. The results obtained showed improved image quality, but it is not certain to what the algorithm converged as no reference values could be obtained from the subjects investigated.

Other authors have also described transmit-beam iteration,[10–14] but only performed aberration correction on either the received signal, or on the transmitted and the received signal. No further iterations were carried out.

In order to obtain qualitative data concerning the convergence of a transmit-beam iteration process, two aberration estimation methods are compared in this article. Both methods estimate arrival-time and amplitude fluctuations using signals from random scatterers. The estimated arrival-time and amplitude fluctuations are then used as a matched filter for time-delay and amplitude aberration correction.

The first estimation method correlates each element signal with a reference signal. The reference signal is a weighted and modified beamformer output of the received signal.[15] The second method uses an eigenfunction decomposition of the cross spectrum to maximize the expected energy in the received signal.[16]

In order to evaluate the quality of an aberration correction method, Mallart and Fink developed a focusing criterion based on the van Cittert–Zernike theorem.[11] An analogous criterion was developed by Liu and Waag.[10] Lacefield and Waag[14] discuss the utility of this focusing criterion since the van Cittert–Zernike theorem is only valid for propagation in a homogeneous medium. The width of the average receive coherence function at different levels was suggested as an alternative measure to evaluate an aberration correction method. A monotonic relation between the coherence widths and the effective widths of point spread functions was observed in single-transmit images.

Both of these measures are used in this article, in order to evaluate the convergence of the transmit-beam iteration process. The process is also evaluated by comparing arrival-time and amplitude fluctuation estimates to those obtained from point source simulations. A simulation with a point source in the focus of the array provides an optimal situation for observing aberration of the ultrasound wave, and serves as a good reference.

Absorption effects, electronic noise, and acoustic reverberation noise were not included in the simulations.

II. THEORY

A. Signal and aberration correction modeling

Following Angelsen[17] (Ch. 11), the aberration is modeled by relating the Green's function for the wave equation with constant coefficients to the Green's function for the wave equation with spatially variable coefficients using a filter denoted the *generalized frequency-dependent screen*. The frequency response of this filter describes the aberration introduced to each frequency component of the signal.

If the generalized frequency-dependent screen is independent of the position in space at which the backscatter was created, the signal received at array coordinate $\mathbf{r}_a$ can be written as[15]

$$y(\mathbf{r}_a;\omega)=s(\mathbf{r}_a;\omega)f(\mathbf{r}_a;\omega). \tag{1}$$

The function $f(\mathbf{r}_a;\omega)$ is an integral over a volume containing scatterers distributed in space, and represents the unaberrated acoustic backscatter signal. It does, however, depend on the transmitted beam, and is thus a function of the transmit aberration.

The situation where aberration on an array element satisfies the assumption of being independent of the spatial position of the scatterer, is denoted *scatterer-independent aberration*.[15] This can be viewed as concentrating all aberration of the inhomogeneous medium to a layer at the array surface. For an extended aberrator of varying thickness, this assumption is generally not satisfied, but may be a good approximation inside a region surrounding the focal point; *the isoplanatic patch*. A received signal according to Eq. (1) is thus obtained by focusing the transmitted beam to the inside of the isoplanatic patch.

The two aberration estimation methods employed in this article perform aberration correction using a time-delay and amplitude correction filter with transfer function

$$h(\mathbf{r}_a;\omega)=a(\mathbf{r}_a)e^{i\omega\tau(\mathbf{r}_a)}. \tag{2}$$

The time-delay τ, and amplitude a, are functions of the array coordinate $\mathbf{r}_a$, but do not depend on frequency. This approximation of the correction filter is valid for band-limited signals assuming scatterer-independent aberration.

It has been shown that a time-delay and amplitude filter produces close-to-ideal correction (no aberration), if correct estimates for the arrival-time and amplitude fluctuations are obtained, even in the case of severe aberration.[5]

B. Scatterer-independent aberration and the van Cittert–Zernike theorem

When the scatterers are randomly distributed in space, the backscatter signal is a stochastic variable. Assuming scatterer-independent aberration, the cross-spectrum between the received signal at location $\mathbf{r}_p$ and $\mathbf{r}_n$ on an array may be expressed using Eq. (1)

$$R(\mathbf{r}_p,\mathbf{r}_n)=s(\mathbf{r}_p)s^*(\mathbf{r}_n)F(\mathbf{r}_p,\mathbf{r}_n). \tag{3}$$

Here, $F(\mathbf{r}_p,\mathbf{r}_n)$ is the cross-spectrum of the backscatter signal without aberration. Dependence on frequency has been omitted for notational convenience.

In Ref. 18, the van Cittert–Zernike theorem was developed for incoherent acoustic backscatter and propagation through a homogeneous medium. If the aberration is scatterer-independent, the van Cittert–Zernike theorem may be applied. In this case, $F(\mathbf{r}_p,\mathbf{r}_n)=F(\mathbf{r}_p-\mathbf{r}_n)\equiv F(\boldsymbol{\xi})$ is computed as (Ref. 17, 11.55)

$$F(\boldsymbol{\xi}) = \frac{\sigma_\nu^2}{4\pi^2} \int_{S_a} s(\mathbf{r}+\boldsymbol{\xi}) s^*(\mathbf{r}) o(\mathbf{r}+\boldsymbol{\xi}) o^*(\mathbf{r}) \mathrm{d}\mathbf{r}. \tag{4}$$

Here σ_ν^2 is the scattering intensity, the integration is performed over the array surface S_a, and $o(\mathbf{r})$ denotes the array apodization function.

Equation (4) shows that the coherence in the received signal is limited by the aberration as well as the apodization function. This has been experimentally observed,[14] although it was not compared to an explicit theoretical prediction.

III. ESTIMATORS

In this article, two previously developed estimators[15,16] are employed to study transmit-beam iteration. For the convenience of the reader and to introduce notation, the rationale behind both estimators is briefly reviewed. Then, the two methods are compared, and new insight into the similarities and differences between them is provided.

Both estimators are based on the cross-spectrum of the received acoustic backscatter. For the purpose of this study, the received signal is assumed to be a Gaussian stochastic process with zero mean value. This implies that all statistical information is contained in the covariance function, or equivalently, the cross-spectrum. For a time-delay and amplitude correction filter as in Eq. (2), it is sufficient to consider the cross-spectrum at a single frequency.

For a given frequency ω, the cross-spectrum between the *element signals* $y_p(\omega)$ and $y_n(\omega)$, received at element p and n respectively, is defined as

$$R_{pn} = \mathrm{E}[y_p \ y_n^*]. \tag{5}$$

Frequency dependence has been dropped for notational convenience.

For the comparison to be useful, both estimation methods use the same estimate of the cross-spectrum. In order to obtain a proper estimate of the cross-spectrum with low variance, an average over statistically independent backscatter signals is used. In a practical situation, statistically independent signals can be obtained by imaging scatterers which are replaced between each consecutive transmit-beam, *e.g.*, blood or contrast agents. Alternatively, nonoverlapping regions of the scatterer distribution may be utilized by combining beams in a linear/sector scan.[6,19]

The estimate of the cross-spectrum is a cross-periodogram given as

$$\tilde{R}_{pn} = \frac{1}{K} \sum_{k=1}^{K} y_{kp} y_{kn}^*, \tag{6}$$

where k denotes received backscatter signals from different random scatterer realizations, and K is the total number of such realizations. To further lower the variance of the estimate in Eq. (6), additional averaging over a small band of frequencies is performed which results in a smoothed cross-periodogram.

Reference 20 (p. 703) shows that the variance of the cross-spectrum estimate in Eq. (6) may be found as

$$\mathrm{Var}[|\tilde{R}_{pn}|] \sim \frac{1}{2K} |R_{pn}|^2 \left(\frac{1}{|w_{pn}|^2} + 1 \right)$$

$$\mathrm{Var}[\angle \tilde{R}_{pn}] \sim \frac{1}{2K} \left(\frac{1}{|w_{pn}|^2} - 1 \right),$$

where the coherence w_{pn} is defined as

$$w_{pn} = \frac{R_{pn}}{\sqrt{R_{pp} R_{nn}}}. \tag{7}$$

This implies that the variance of the cross-spectrum is high when the coherence is low, and vice versa.

A. Modified beamformer output

This section offers a brief description of the *modified beamformer output* (MBFO) estimator, which was presented in Ref. 15. The basic premise for this method is that the received signal can be written as in Eq. (1), that is, scatterer-independent aberration is assumed. Applying the same discrete notation as in Eq. (5) to denote elements p and n in Eq. (3), and solving for s_p leads to

$$s_p = \frac{R_{pn}}{F_{pn}} \frac{1}{s_n^*}. \tag{8}$$

In order to use all possible correlation information to estimate the phase and amplitude of s_p, a weighted average $\hat{s}_p$ is defined

$$\hat{s}_p = \sum_{n=1}^{N} W_{pn} \frac{\tilde{R}_{pn}}{F_{pn}} \frac{1}{\hat{s}_n^*}, \tag{9}$$

where N is the total number of elements on the array. Here, W_{pn} is a set of weights and $\tilde{R}_{pn}$ is the estimate for R_{pn}.

In this article, the weights are chosen as

$$W_{pn} = |\tilde{w}_{pn}|^2 \Big/ \sum_{n=1}^{N} |\tilde{w}_{pn}|^2, \tag{10}$$

where $\tilde{w}_{pn}$ is an estimate of the coherence w_{pn} [Eq. (7)] based on $\tilde{R}_{pn}$. Thus, the estimates of R_{pn} with low variance are emphasized. Furthermore, the phase of F_{pn} is not known, and F_{pn} is therefore replaced by its absolute value. The MBFO estimator $\tilde{s}_p$ is then

$$\tilde{s}_p = \sum_{n=1}^{N} W_{pn} \frac{\tilde{R}_{pn}}{|F_{pn}|} \frac{1}{\tilde{s}_n^*}. \tag{11}$$

An estimate for $|F_{pn}|$ can be found from the van Cittert–Zernike theorem as formulated in Eq. (4).[15]

The estimator in Eq. (11) is a set of N coupled nonlinear equations which has to be solved, that is, for $p \in \{1,\cdots,N\}$. An iterative solution method as described in Ref. 15 was utilized for this purpose. The initial estimate for s_p in the iterative solution method was chosen as zero phase and unity amplitude across the array.

B. Eigenfunction estimator

The *eigenfunction estimator* (EFE) was presented in Ref. 16. Thus, only a short description of the method is provided here. Consider the stochastic vector of receive signals at a particular frequency ω and transducer elements indexed from 1 to N

$$\mathbf{y}=[y_1 \;\; y_2 \;\; \cdots \;\; y_N]^T.$$

Given a vector $\mathbf{h}=[h_1\, h_2 \cdots h_N]^T$, a stochastic linear functional $\mathcal{L}_h$ may be defined on $\mathbf{y}$ as

$$\mathcal{L}_h\mathbf{y}=\mathbf{h}^{\mathrm{H}}\mathbf{y}=\sum_{p=1}^{N} y_p h_p^*, \tag{12}$$

where H denotes the Hermitian of the vector.

The quantity $\mathcal{L}_h\mathbf{y}$ from Eq. (12) is the temporal frequency result when a filter with transfer function $h_p(\omega)$ is applied to the signal received at transducer element p before the standard beamforming procedure is executed. It is a stochastic variable with associated variance expressed as

$$\|\mathcal{L}_h\mathbf{y}\|^2 \equiv \mathrm{E}[\mathcal{L}_h\mathbf{y}(\mathcal{L}_h\mathbf{y})^{\mathrm{H}}]=\mathbf{h}^{\mathrm{H}}E[\mathbf{y}\mathbf{y}^{\mathrm{H}}]\mathbf{h}=\mathbf{h}^{\mathrm{H}}R\mathbf{h}. \tag{13}$$

Here, R is the cross-spectrum matrix of the receive signal at the frequency ω.

The variance is the *expected energy* for $\mathcal{L}_h\mathbf{y}$ at this frequency. Since the matrix R is Hermitian the expected energy, subject to the constraint $\mathbf{h}^{\mathrm{H}}\mathbf{h}=1$, is maximized when $\mathbf{h}$ is an eigenvector associated with the largest eigenvalue of R (see Ref. 21, Ch. 6.5).

Through finding the eigenvector which maximizes the expression in Eq. (13), a match filter which maximizes the *speckle brightness*,[22] is constructed. The normalization $\mathbf{h}^{\mathrm{H}}\mathbf{h}=1$ for each frequency ensures that the correction filter does not alter the frequency distribution for the energy of the aberration-corrected transmit-beam.

The eigenvector $\tilde{\mathbf{h}}$ associated with the largest eigenvalue of $\tilde{R}$ is calculated and used as an estimate of the filter $\mathbf{h}$.

C. Comparison of the estimators

Both methods estimate aberration from the cross-spectrum of stochastic backscatter. The MBFO estimator also assumes the aberration on a receive element to be independent of the spatial position of the scatterers, *i.e.*, a signal model according to Eq. (3). The EFE estimator makes no such assumption.

The MBFO estimator has been shown to be equivalent to correlating the received signal with a correlation reference;[15] a modified beamformer output

$$\begin{aligned}\tilde{s}_p &= \frac{1}{K}\sum_{k=1}^{K} y_{kp} b_{kp}^* \\ b_{kp} &= \sum_n W_{pn}\frac{1}{|F_{pn}|\tilde{s}_n} y_{kn}.\end{aligned} \tag{14}$$

The modified beamformer output, b_{kp}, is formed by using a weight term W_{pn} and a correction term $1/|F_{pn}|s_n$ for each element signal y_{kn}.

The same interpretation is possible for the EFE

$$\begin{aligned}\tilde{h}_p &= \frac{1}{K}\sum_{k=1}^{K} y_{kp}\beta_k^* \\ \beta_k &= \sum_n \frac{1}{\lambda}\tilde{h}_n^* y_{kn}.\end{aligned}$$

Equal weight, $1/\lambda$, is placed on all element signals when forming the modified beamformer output β_k. The correction term in this case is h_n^*.

A major difference between the two estimators is that the correction term for the MBFO estimate is obtained by applying the aberration correction filter as an inverse filter, while the correction for the EFE is obtained by matched filtering. In addition, the MBFO estimator utilizes a different set of weights for each transducer channel p, thus obtaining a different correlation reference signal for each channel. The EFE estimator makes use of the same correlation reference for all channels.

To compare the estimators further, it is instructive to consider the case where $F_{pn}=|F_{pn}|$. This will be the case when, for example, the scattering medium is incoherent and all phase aberration of the transmitted beam has been corrected.

Let S be the diagonal matrix

$$S=\begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_N \end{pmatrix}.$$

Equation (13) is then reformulated as

$$\|\mathcal{L}_h\mathbf{y}\|^2=\mathbf{h}^{\mathrm{H}}R\mathbf{h}=\mathbf{h}^{\mathrm{H}}SFS^{\mathrm{H}}\mathbf{h}, \tag{15}$$

where F is the cross-spectrum matrix for the unaberrated acoustic backscatter. Therefore, $S^{\mathrm{H}}\mathbf{h}$ must be an eigenvector of F. Now, since F is real, then the eigenvector $S^{\mathrm{H}}\mathbf{h}$ is real as well. In this case, the phase of h_p is equal to that of s_p; the phase estimated by the EFE will be an unbiased estimate for the phase of the screen.

Furthermore, it is easy to see that if $\mathbf{h}$ is an eigenvector of R with eigenvalue λ, then

$$h_p=\frac{1}{\lambda}\sum_n R_{pn}h_n=\sum_n \frac{|F_{pn}||h_n|^2}{\lambda}\frac{R_{pn}}{|F_{pn}|}\frac{1}{h_n^*}. \tag{16}$$

The EFE therefore satisfies an equation of the same type as Eq. (11) for the MBFO, with weights $W_{pn}=|F_{pn}||h_n|^2/\lambda$.

If the weights W_{pn} in Eq. (11) are required to satisfy $\Sigma_n W_{pn}=1$, then the MBFO will by construction be an unbiased estimate for the screen.[15] However, for the EFE this requirement is not necessarily fulfilled. The result is a biased estimate of the amplitude.

Express the amplitude bias in a multiplicative fashion

$$h_p=\alpha_p s_p,$$

where α_p is real and positive, and s_p, as previously, denotes the screen. Inserting this into Eq. (16) yields

$$s_p=\sum_n \frac{|F_{pn}||s_n|^2\alpha_n}{\lambda\alpha_p}\frac{R_{pn}}{|F_{pn}|}\frac{1}{s_n^*}.$$

The fact that a normalized set of weights will obtain an unbiased estimate for the screen implies that the amplitude bias may be expressed as a solution to

$$\alpha_p = \sum_{n=1}^{N} \frac{|F_{pn}||s_n|^2}{\lambda} \alpha_n . \tag{17}$$

Because of the Toeplitz structure of F, and the fact that $|F_{pn}|$ decreases off the main diagonal, any solution α_p of Eq. (17) will decrease as a function of p when p moves toward the edges of the array. If $|F_{pn}|$ decreases monotonically, then α_p will also decrease monotonically from a maximum in the central region of the array. The filter amplitude is therefore an estimate for an apodized version of the screen amplitude. This apodization has previously been discussed,[16] but the expression for the apodization is new.

In general, when F is not real, the relationship between h_p and s_p is more complicated. It is, however, possible to show that an iterative transmit-beam aberration correction procedure will converge to a h_p which has a phase that concurs with the screen.[16] Applying the correct phase for aberration correction will result in a F which satisfies $F_{pn} = |F_{pn}|$. The preceding argument may then be used to assert that an apodized amplitude estimate is also obtained.

By omitting the phase of F_{pn}, an error is introduced in the MBFO estimate of the screen. Equation (9) can be written as

$$\hat{s}_p = \sum_{n=1}^{N} W_{pn} \frac{\tilde{R}_{pn}}{|F_{pn}|} \frac{e^{-i\Delta_{pn}}}{\hat{s}_n^*}, \tag{18}$$

where Δ_{pn} is the phase of F. Neglecting this phase will therefore, in general, contribute both to a phase and an amplitude error in the estimation of $\tilde{s}_p$ from Eq. (9). Assuming the transmit-beam iteration process converges to the true phase of s_p, as discussed above, F will be real valued and $\hat{s}_p$ becomes by definition equal to $\tilde{s}_p$.

D. Arrival-time and amplitude estimates

After obtaining an estimate at the center angular frequency, ω_0, for the scatterer-independent screen s_p and the energy maximizing filter h_p, using the MBFO and the EFE estimator, respectively, arrival-time and amplitude fluctuations were calculated in a standard way.[15]

Note that although in the presented work only arrival-time and amplitude fluctuations were used, both the MBFO and the EFE may be employed to estimate a phase and amplitude aberration correction filter for all frequency components in the signal.

IV. SIMULATIONS

The simulations presented in this article were performed using the two-dimensional (2D) simulation setup shown in Fig. 1. An angular spectrum operator was used for homogeneous propagation of the simulated signals.[5] A beam was propagated from the transducer through a body wall model to the scattering region. There it was scattered according to the Born approximation and propagated back to the transducer.

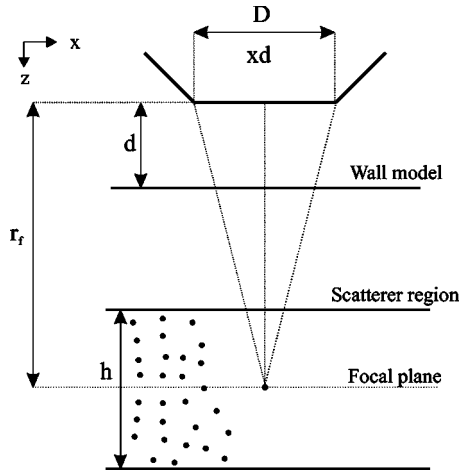


FIG. 1. An ultrasound pulse was propagated from the transducer (xd), through a body wall to a scattering region. Scattering was computed according to the Born approximation and propagated back through the body wall to the transducer.

Eight body wall models were generated using equally spaced time-delay screens, filtered and tuned to obtain characteristics according to abdominal wall measurements.[23] The body wall models were also used in Ref. 5. A thorough description of the body wall models is offered in this reference.

The point source simulations were of a one-way nature: A point source was situated at the position of maximum energy of the transmitted beam in the focal plane for each of the aberrators. An emitted pulse from the source, identical to the transmit pulse from the array, was propagated to the array and processed to obtain a reference for the arrival-time and amplitude fluctuations.

A. Simulation parameters and data processing

The simulation domain was 10.24 cm in the lateral direction (x direction in Fig. 1) with a resolution of 0.2 mm. To avoid reflections at the edges of the spatial domain due to the fast Fourier transform being periodic, the signal was tapered to zero with a raised cosine window over a 2.54 cm-wide band. The sampling frequency was 35.1 MHz providing a time window of 58.3 μs. The transmitted pulse had a center frequency of 2.5 MHz and a -6 dB bandwidth of 1.5 MHz. An array aperture size of 20 mm with pointlike elements was chosen. The focal depth of the array was set to 60 mm. The medium through which the signals were propagated had a speed of sound equal to that of water; 1523 m/s. Geometric focusing was removed from all received signals prior to further processing of the results.

To generate a realistic speckle signal, an area of 30.5 mm (time window of 20 μs), centered with 15.25 mm to each side of the focal plane was used as a scattering region (see Fig. 1). The scatterer density was approximately 1600 scatterers per square centimeter. The scatterers were uniformly distributed in space, and had a Gaussian distributed reflection strength.

For each transmit-beam iteration, scattering from 20 independent realizations of the scattering region were simulated for the purpose of cross-spectrum estimation.

Estimation of arrival-time fluctuations for the point

source simulation was performed with a phasefront tracking algorithm.[5] The method has proved to yield accurate estimates of the wavefront, and is not sensitive to waveform deformation which occurs behind the wavefront.

For all arrival-time fluctuation estimates presented in this article, a linear fit was subtracted in order to remove refraction steering of the beam.

Amplitude fluctuations from the point source simulations were determined by taking the Fourier transform of the received signal on each element as a function of time. The amplitude on each element of the array was calculated as the arithmetic mean of the amplitudes of the now frequency-dependent signal, over a band of frequencies ranging from 2–3 MHz. This band was chosen empirically.

In order to evaluate the accuracy of the estimation methods, the relative L^2 distances between arrival-time and amplitude estimates and their respective references obtained from the point source simulations were calculated. The L^2 distance was normalized with respect to the L^2 norm of the reference, and was thus calculated as

$$d(x,x_{\mathrm{ref}})=\sqrt{\sum_{i=1}^{N}|x_i-x_{i,\mathrm{ref}}|^2}\Big/\sqrt{\sum_{i=1}^{N}|x_{i,\mathrm{ref}}|^2}.$$

Here x is the arrival-time or amplitude estimate, and x_{ref} is the reference value obtained from the point source simulations. The mean value was subtracted from all estimates prior to the calculation of the L^2 distance.

For the comparison to be useful, the amplitude estimates and the point source reference need to have equal power. Assuming the estimated values for the amplitude fluctuations are proportional to the reference, a gain factor α may be defined as

$$\hat{\mathbf{a}}=\alpha\mathbf{a}.$$

Here, $\hat{\mathbf{a}}$ is the estimate and $\mathbf{a}$ the point source reference. The gain factor was determined by minimizing the error between the reference and the estimate

$$\alpha=\frac{\hat{\mathbf{a}}^{\mathrm{T}}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}.$$

To ensure equal power, the estimated arrival amplitudes were then scaled using the gain factor, prior to the calculation of the relative L^2 distance.

The focusing criterion was calculated according to the derivations by Mallart and Fink.[11] They defined a focusing criterion as

$$C=\frac{\int_{-\infty}^{+\infty}(\Sigma_{p=1}^{N}y_p(t-\tau_p))^2\mathrm{d}t}{N\cdot\Sigma_{p=1}^{N}\int_{-\infty}^{+\infty}y_p^2(t)\mathrm{d}t},$$

where N is the number of elements on the receiving array. Liu and Waag[10] independently proposed a similar criterion denoted the *waveform similarity factor*.

For a point source, the value of C lies between 0 and 1. For an incoherent medium Mallart and Fink showed that the maximum value of C is 2/3. Note that C can only attain its maximum value if τ_p is properly estimated.

The focusing criterion was, as described earlier, used to evaluate the convergence of the iterative aberration correction procedure, along with the widths of the average receive coherence function introduced by Lacefield and Waag.[14]

The average coherence function for the received signal, from now on denoted *coherence function*, was calculated as

$$\overline{w}_{p-n}=\frac{1}{N_\Delta}\sum_{N_\Delta}\tilde{w}_{pn}=\frac{1}{N_\Delta}\sum_{N_\Delta}\frac{\tilde{R}_{pn}}{\sqrt{\tilde{R}_{pp}\tilde{R}_{nn}}},$$

where N_Δ denotes the number of element pairs with separation $p-n$. Note that the coherence function was only calculated for the center frequency of the signal. The magnitude of the coherence function was interpolated to a resolution of 0.05 mm sampling, before the width of the magnitude of the coherence function was calculated at levels 0.6 and 0.4.

All received data were corrected using arrival-time fluctuation estimates, obtained by the estimators, prior to the calculation of the focusing criterion C and the coherence function. For the focusing criterion, the linear fit of the arrival time estimates was not subtracted prior to receive correction.

Beam profiles in the focal plane of the array were acquired as the rms value of the temporal signal at each spatial position. These profiles were used for the visual evaluation of the effect of the different aberration correction methods.

V. RESULTS

Simulations were performed using eight different aberrators. To limit the amount of presented data, detailed results are only offered for two of the aberrators; w6 and s6. The w6 and s6 aberrator represent weak and strong aberration, respectively, and are representative for overall performance of the iterative transmit-beam aberration correction. Only the parameter C is presented for all aberrators, as this proved to be the best criterion by which to quantify the transmit-beam iteration results.

In all the results presented, MBFO and EFE denote results obtained using the corresponding method for estimating arrival-time and amplitude fluctuations.

All results in this section are labeled with an iteration number. The iteration number is defined according to the transmit-beam. One transmit-beam iteration is defined as consisting of an estimation of arrival-time and amplitude fluctuations; an application of these estimates to a transmit-beam in order to obtain a corrected transmit-beam profile; and finally receiving scattering generated by the corrected transmit-beam. In this labeling scheme, *iteration 0* refers to the initial transmit-beam, where no aberration correction is applied. The arrival-time and amplitude fluctuations estimated using scattering created by the transmit-beam from iteration 0 are used to form the first truly corrected transmit-beam. These arrival times, amplitudes, and the resulting beam profiles are thus labeled *iteration 1*, and so on.

Figure 2 shows beam profiles in the focal plane of the array. For the w6 aberrator, the corrected beam profiles appear to converge after two iterations for both methods. The resulting beam profiles are very well corrected and close to the unaberrated profile. In the case of the s6 aberrator, one

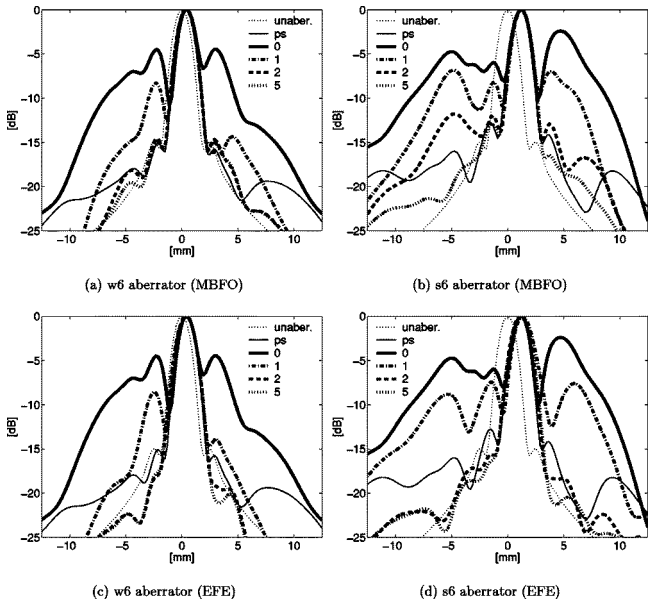


FIG. 2. Beam profiles in the focal plane for the w6 and s6 aberrators. Zero on the horizontal axis represents the center axis of the array. All profiles are normalized to their maximum value. The reference profile represents the situation with no aberration, and the profile denoted ps shows correction using the point source reference. The numbers in the legend refer to the iteration number of the correction procedure, where the 0-iteration zero is alabel. Thus the digit is perfered profile is the uncorrected transmitted profile.

additional iteration is required for MBFO to obtain the same results. Since only minor changes occur from iteration 2 to iteration 5 for the w6 aberrator, and from iteration 3 to iteration 5 for the s6 aberrator, only results from iterations 0, 1, 2, and 5 are presented.

The estimated time-delay and amplitude fluctuations used to produce the corrected beam profiles in Fig. 2 are shown in Figs. 3 and 4. The visual impression of convergence for the estimates is the same as for the beam profiles. It is worth noting, however, that the time-delay estimate is also very accurate after two iterations for MBFO applied to the s6 aberrator. The amplitude improves significantly at the third iteration. The improvement between iterations 2 and 3 for the beam profile is thus mainly explained by an improved amplitude estimate.

The relative L^2 distance between the estimated arrival-time/amplitude and the respective references was computed for each iteration. Figure 5 shows how the distance decreases for the first two iterations. In the case of the w6 aberrator, the distances level out after the second iteration. For the s6 aberrator, the distance for the amplitude levels out in the same manner as for the w6 aberrator. The distance for the arrival-time, however, increases after the second iteration. This is related to the discontinuities in the arrival-time estimates observed in Fig. 3. The value at which the L^2 distance levels out for the EFE amplitude estimate is significantly higher than for the MBFO estimate. This is explained by the fact that the EFE amplitude is apodized relative to the screen, while the MBFO is not.

The magnitude of the coherence functions, and the coherence widths for the received scattering, are presented in Figs. 6 and 7. The coherence widths at different levels increase gradually with iteration, demonstrating an increased degree of spatial coherence in the receive signal.

The focus quality parameter for all aberrators is presented in Fig. 8 as a mean value and a standard deviation. These were calculated using the 20 independent receive signals for each iteration. In concurrence with the theoretical foundation for the parameter,[11] the strong increase in the focus quality parameter C corresponds to the improved focus apparent in the beam profiles in Fig. 2. With the exception of s8, convergence was obtained after one to three iterations using either algorithm.

VI. DISCUSSION

As shown in this article, both the MBFO and EFE algorithm use an average of element signals as the reference value in a correlation process, in other words, a beamforming correlation process. This is conceptually similar to the speckle brightness method,[22] speckle look-back,[12] the beamsum-channel correlation method,[9] and the scaled covariance matrix algorithm[24] for phase estimation. The principal difference is that both methods presented here can estimate both phase and amplitude aberration at all frequency components, and thus represent a generalization of the above described methods.

In order to obtain a correlation-based estimate, a stable reference signal is needed. The variance of the estimate will be as low as possible when the reference signal is coherent with the backscatter signal. In the case of the MBFO estimator, a separate reference b_{kp} is used for each element p. The EFE, on the other hand, uses the same reference, β_k, for all elements.

The MBFO estimate utilizes the signal model to create signals of equal strength at each element by factoring out the effect of the aberration amplitude. It then forms a reference which is coherent with the signal at element p by explicitly applying the coherence function as weights in a weighted average.

The weight function works as a sliding window which efficiently implements a subaperture processing, automatically selecting an appropriate subaperture for the beamformer output from a variance perspective [confer Eq. (14)]. The weighting also ensures that the beamformer output is highly correlated with the element signal where the estimation occurs. Since the subaperture slides across the array, it is desirable with an inverse amplitude filtering in order for the reference signal to attain the same average energy level for each subaperture.

The EFE constructs one signal which is utilized as a common correlation reference for all element signals across the aperture. To this end, no signal model is employed directly. Instead, the reference signal is formed as a weighted coherent sum of the element signals. Assuming no amplitude damping due to absorption, a backscatter signal of large amplitude is the result of constructive interference. A low-amplitude backscatter signal, on the other hand, is the result of destructive interference. As a result, high-amplitude signals will resemble each other more closely than low-amplitude signals. The accuracy with which the aberration correction filter may be estimated, is directly connected to

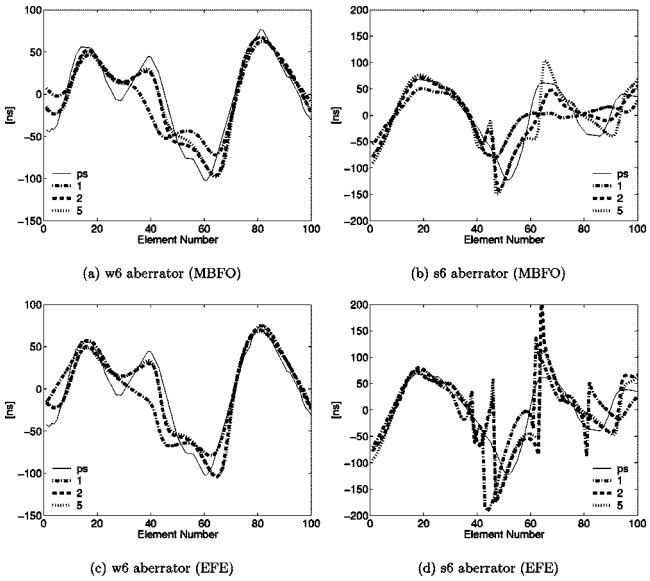


FIG. 3. Arrival-time fluctuations for the w6 and s6 aberrators. The horizontal axis is given in array elements. The curves denoted ps are the references obtained from the point source simulations. The numbering of the arrival-time curves in the legend corresponds to the iteration number of the corrected transmitted beam in Fig. 2. Arrival-time curve number 1 was thus used to obtain beam profile number 1 in Fig. 2.

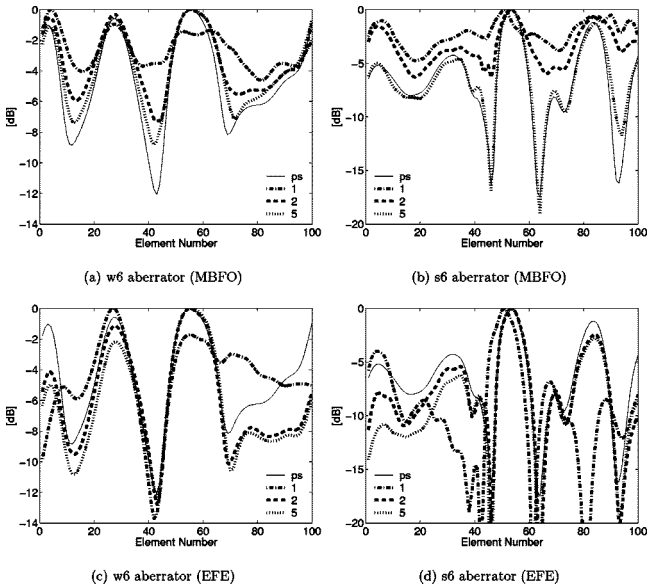


FIG. 4. Amplitude fluctuations for the w6 and s6 aberrators. The horizontal axis is given in array elements. The curves denoted ps are the references obtained from the point source simulations. All amplitude fluctuation curves are normalized to their maximum value. The numbering in the legend is the same as in Fig. 3.

the degree of coherence between an element signal and the reference. In order to form a good estimate, it is therefore of importance to form a reference signal which is highly correlated with the element signals. Furthermore, because signals of high amplitude contribute more to the overall focus quality than low-amplitude signals, it is most important to obtain an accurate estimate for the correction of high-amplitude signals. The weighted coherent sum of element signals should emphasize element signals of large amplitude in order to achieve this. To what degree high amplitudes should be emphasized over low amplitudes is determined by the L^2 norm used when maximizing the expected energy of $\boldsymbol{\mathcal{L}}_h\mathbf{y}$ in Eq. (12).

As described in Sec. IV A, for each transmit-beam iteration, scattering from 20 new realizations of the scattering region were simulated. The objective of this article was to study iteration of transmit-beam aberration correction. For this purpose, a proper estimate of the cross-spectrum [Eq. (6)] was desired. For practical purposes, using 20 transmit-beams for each estimate is unfeasible in a real-time scanning environment. However, if only a time-delay and amplitude aberration correction filter is sought, appropriate model-based averaging of the cross-spectrum over a wider frequency band will yield a similar effect to acquiring independent realizations. In this situation, the number of independent realizations may therefore be greatly reduced. An important issue for implementation will be to determine the number of receive-signals necessary to obtain an adequate estimate.

The transmit-focus of an ultrasound beam may be quantified by measuring the width of the transmit-beam profile. The beam profiles shown in Fig. 2 display a significant improvement in focus quality as a result of the iterative transmit-beam correction process. Furthermore, the convergence toward an almost-ideal beam profile width is rapid. There is, however, a slight offset in the peak of the beam profiles.

This shift, particularly noticeable for the s6 aberrator, is produced by a refraction of the transmitted beam. In this two-way imaging system, where reciprocity implies that the backscattered beam will experience the same refraction as the transmitted beam, the associated shift of the beam profile is not observed from the transducer array. Dealing with refraction of the beam due to aberration through the body wall remains an issue for further research.

For the w6 aberrator, both estimation methods yield transmit-beams with the same degree of focus after two transmit-beam iterations as those obtained using the point source reference for aberration correction. Beyond two iterations, no significant improvement of the beam profiles is achieved. In the case of the s6 aberrator, three iterations are required for the MBFO estimate to achieve as good a correction as the point source reference. Beyond this point, no significant improvement can be found in the transmit-beam profiles. The EFE estimate does not improve the beam profile significantly after the second iteration.

Convergence for the beam profiles is accompanied by an apparent convergence also for the time-delays and amplitudes shown in Figs. 3 and 4. After approximately two iterations, no essential change occurs in the phase estimates for either method. Amplitude estimates are improved with further iteration for the MBFO algorithm, but for the EFE estimate more than two iterations are not required.

In the weakly aberrated case, the time-delays obtained using both estimation methods are very close to the point source reference. For the s6 aberrator, the estimated arrival times exhibit discontinuities, while the point source reference does not. The discontinuities are linked to waveform deformation in the received signals for the s6 aberrator.[15] Waveform deformation results in low signal amplitudes, and

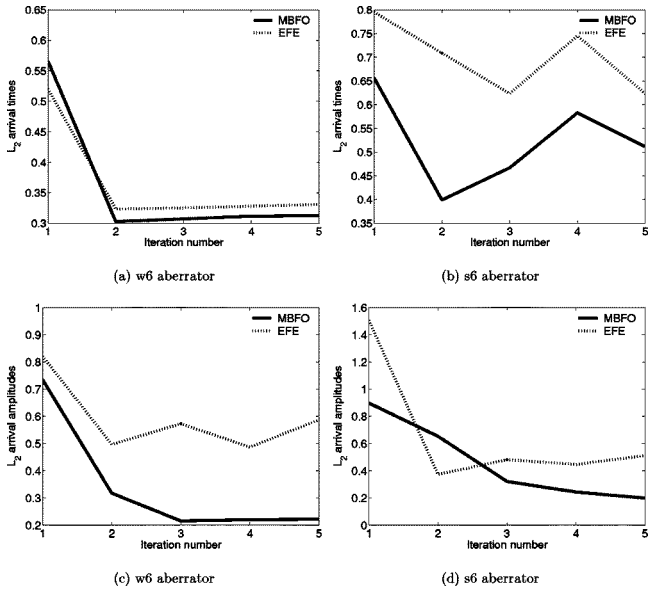


FIG. 5. Relative L^2 distance between point source reference and estimate for w6 and s6 aberrators. The top row shows the L^2 distance for arrival-time fluctuations, and the bottom row for amplitude fluctuations. The horizontal axis indicates the iteration number according to Figs. 3 and 4.

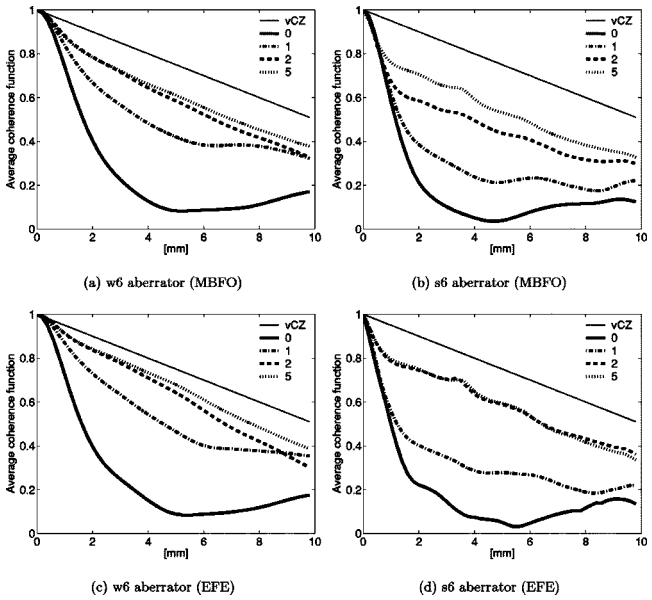


FIG. 6. Absolute value of the average coherence function for the w6 and s6 aberrators. The numbering of the curves corresponds to the iteration number, *i.e.*, coherence functions for the received signals obtained using a transmit-beam with the same number in Fig. 2. The curve denoted vCZ indicates the theoretical upper bound for the coherence based on the van Cittert–Zernike theorem for a homogeneous medium. All received signals were corrected using the estimated arrival-time fluctuations prior to calculating the coherence function.

causes discontinuities in the arrival-times between adjacent element signals.[4,5]

The MBFO algorithm produces amplitude estimates which are close to the point source reference both for the w6 and s6 aberrators. For the s6 aberrator, the corrected beam profile is improved significantly for the third iteration using MBFO, even if the time-delay estimate does not change much from the second to the third. This improvement must therefore be the result of the improved amplitude estimate.

This observation is in accordance with the predicted amplitude error as a result of omitting the phase of F_{pn} in Eq. (11). Transmitting with a correct phase filter will, according to Eq. (4), produce an F_{pn} which is real. The amplitude error will therefore not be introduced in the MBFO estimate based on the corrected backscatter, resulting in a good amplitude estimate.

Due to the discontinuities of the arrival-time estimates, and the apodization of the EFE amplitude relative to the screen, the L^2 distance remains large also after iteration (see Fig. 5). Therefore, the L^2 distance does not adequately reflect the aberration correction capabilities of the associated correction filter.

Rigby *et al.*[9] obtained convergence after three or four iterations, but the convergence quality of the estimates is not certain as no reference could be provided for the subjects used in the study.

The number of independent signals used for estimating covariance or cross-spectra will influence the accuracy with which arrival-times and amplitudes are estimated. This will affect the convergence rate. In Ref. 5, it was shown that introducing amplitude correction in addition to time-delay correction was of vital importance for the side-lobe level of the beam profiles. Proper amplitude correction will thus increase the spatial coherence in the received signal, and reduce the number of iterations required for the aberration correction algorithm to converge. Both of the aforementioned issues could be the catalyst for the overall improved convergence rates in this work compared to the results obtained in Ref. 9.

Based on the van Cittert–Zernike theorem, the width of the coherence function for the receive signal may be used to determine the width of the associated focus. Figure 6 clearly shows an increasing overall width of the coherence functions as a result of iteration for both aberrators and both estimation methods. The visual impression is that performing two iterations is sufficient to obtain almost maximum coherence widths. Beyond this, only minor improvements occur. Thus, the coherence functions do give the correct impression of the converging beam profiles for the iteration process. However, the difficulty herein is determining at which level the width of the coherence function should be measured. As seen in Fig. 7, the choice impacts on the width curves dramatically. This renders the coherence width less attractive for determining the point of convergence.

The parameter C was originally introduced as a measure to quantify the degree of focusing for a given transmit-beam based on the backscatter signal. When applied to the iterative transmit-beam correction procedure, the value for C is observed to increase initially, and then level off at the point beyond which no practical improvement of the beam profile is achieved. (See Figs. 2 and 8.) The point of convergence can therefore be found by identifying the point where this curve levels off. Furthermore, the theoretical upper bound for this parameter may be used as an indication of how close the corrected beam profile is to an ideal transmit-beam profile. The iteration may therefore be set to terminate when a given threshold for C has been reached.

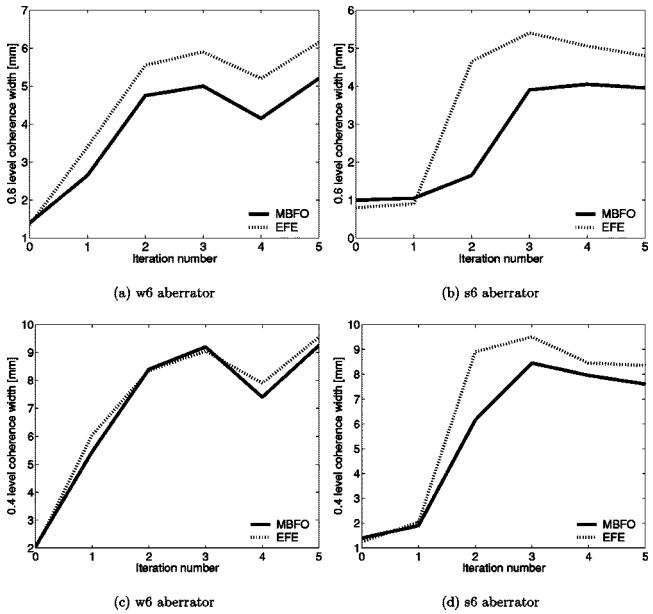


FIG. 7. Coherence widths for the w6 and s6 aberrators at levels 0.6 and 0.4. The horizontal axis shows the iteration number as explained in Fig. 6.

Based on the beam profiles presented for the w6 and s6 aberrator, it is clear, from the focus quality parameter, that the beam profiles for the rest of the aberrators were very well corrected. This was confirmed by visual inspection of the profiles. Further, the results indicate that for $C>0.4$, aberration correction of the transmit beam is close to the correction obtained with the point source reference for these aberrators.

An added advantage of using the parameter C to determine convergence, is the relatively low computational complexity involved. This, combined with the fact that it is computed from information readily available in the backscatter signal, without requiring knowledge of a point source/scatterer, may therefore enable the parameter C to be implemented in an aberration correction procedure without much additional effort.

For the s8 wall, the MBFO algorithm requires five iterations for proper convergence. For this aberrator, the uncorrected beam profile had very high side-lobe levels, where one side lobe was higher than the main lobe. This caused the MBFO algorithm, initially, to focus on this side lobe. Through transmit-beam iteration, the beam profile was improved and a linear slope was detected at iteration 3. Removing the linear component of the arrival time estimate then focused the beam correctly (observed from the point source simulations). For the EFE, this effect is avoided by choosing the eigenvector associated with the second largest eigenvalue because this had a smaller linear component in the phase.[16]

In the presented results, the acoustic scatterers are δ-correlated. This ensures that the acoustic backscatter is a Gaussian stochastic process, at least asymptotically, and is a natural assumption in many imaging situations. Furthermore, the scattering intensity was the same everywhere. This is a good approximation when imaging homogeneous organs, such as liver and spleen. The derivation of the MBFO shows that a spatially variable scattering intensity will affect the estimate in the form of a different cross-spectrum for the unaberrated backscatter, F_{pn}. In this situation, an estimate for F_{pn} may not be obtained using the van Cittert–Zernike theorem unless the scattering intensity is known. An alternative approach would be to utilize the fact that, according to the signal model, the magnitude of the coherence is proportional to $|F_{pn}|$. The EFE will focus the corrected beam also in the situation with spatially variable scattering intensity. However, the focus will be determined by a product of the scattering intensity and the intensity of the transmit-beam, and not the transmit-beam alone. A combination of selecting

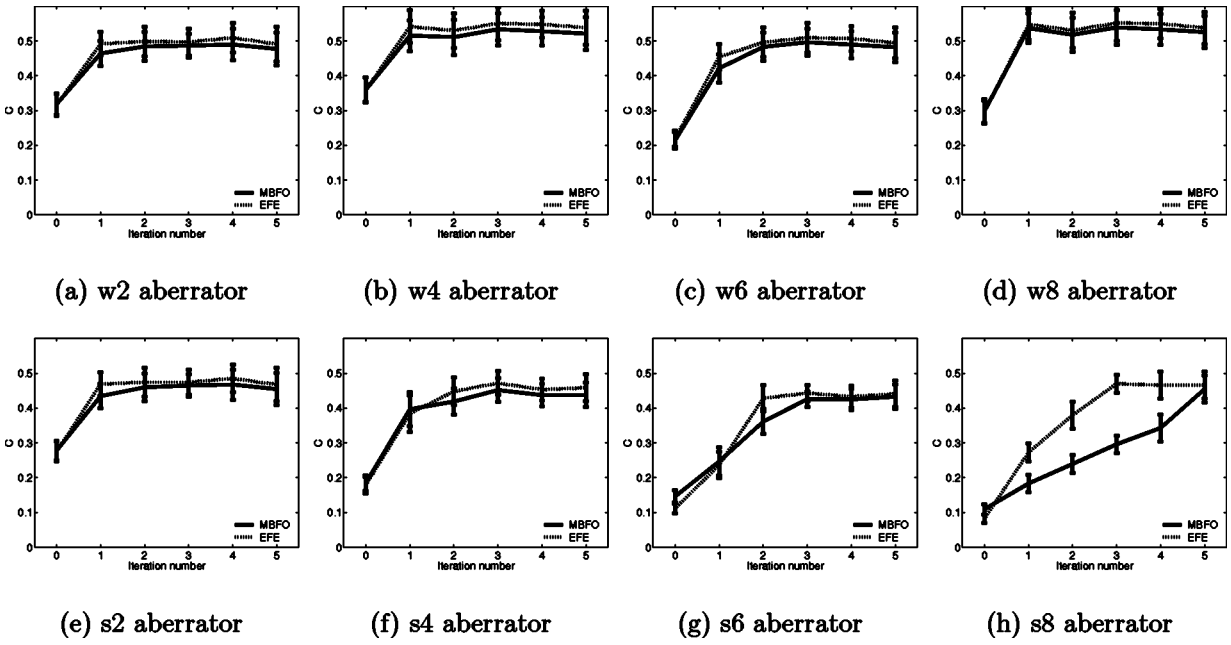


FIG. 8. Mean value and standard deviation of the focus quality parameter C computed for all wall models. The horizontal axis shows the iteration number as explained in Fig. 6. The error bars display the standard deviation. All received signals were corrected using estimated arrival-time fluctuations prior to calculation of the focusing criterion.

the eigenvector with lowest linear component and removing the remaining linear slope will focus the aberration-corrected transmit-beam at the right location. An aberration correction filter may therefore be obtained by either estimation method also in the case of variable scattering intensity. As a consequence of iterative transmit-beam correction, the focal zone will narrow. The assumption of a constant scattering intensity will thus be increasingly better.

In this article, theory and simulations have been presented for ultrasound propagation in a non-absorbing medium. If the medium exhibits absorption which is homogeneous, *i.e.*, the absorption is the same everywhere in the medium, it can still be shown that a match filter is ideal for aberration correction (Ref. 17, Ch. 11). In the presence of heterogeneous absorption, however, the aberration contributed by the absorption should be corrected using an inverse filter.[25,26] In a practical situation, where both heterogeneous absorption and heterogeneous speed of sound are contributing factors, a combination of an inverse filter and a match filter, *e.g.*, a Wiener filter, would probably result in best overall performance.

The results presented here were obtained by simulating sound propagation in two dimensions. This represents a simplification in that out-of-plane aberration/scattering effects are not included. However, combined experimental and simulation studies of wavefront aberration through the abdominal wall have concluded that important aspects of wavefront aberration are observed also in 2D simplified models.[27] Although details may vary, the qualitative aspects of the results are expected to hold, also in a real-world situation.

Although 2D simulations are assumed to give an accurate picture of the aberration correction process, in order to perform aberration correction in a real-world setting certain issues such as element size and directivity must be taken into account.

The array elements must be smaller than the correlation length of the aberration. In the azimuth direction, this requirement will generally be much weaker than standard beamforming requirements. However, the same requirement must also hold in the elevation direction. In practice, this means that either 1.75D or 2D arrays are needed.

For large apertures, element directivity could reduce the signal-to-(electronic)-noise ratio (SNR) along the edges of the array.[14] This will, in turn, lead to a reduced spatial coherence in the measured signal, and thus a less accurate estimate of the correction filter. Since the correlation length of the aberration is generally much larger than the standard beamforming requirement, element signals may be combined in subapertures prior to aberration correction, thereby increasing the SNR. It is also noted that standard apodization will reduce the contribution from the elements along the edges of the array. The reduced accuracy of the filter estimate will therefore only have a limited impact.

Noise was not introduced to the signals used in this study. Both algorithms used here average an aberration corrected signal over a subaperture, or the entire aperture, to create a stable reference for the correlation process. Since electronic noise is uncorrelated between elements, averaging over a subaperture will reduce the noise level in the reference signal.[12] Furthermore, the SNR in standard ultrasound imaging is generally high. Walker and Trahey[28,29] showed that an SNR greater than 15 dB had little effect on the error of correlation-based phase estimates.

The effect of acoustic reverberation noise also represents a challenge, mainly because it is highly correlated both in the temporal and spatial directions. It is beyond the scope of this article to study reverberation noise.

VII. CONCLUSION

Iteration of transmit-beam aberration correction with a time-delay and amplitude filter has been investigated. Two correlation-based algorithms for estimating arrival-time and amplitude fluctuations from random scatterers were employed. The resulting estimates were used to construct a time-delay and amplitude filter for aberration correction.

Results from simulations using eight aberrators, emulating the human abdominal wall, indicate overall convergence for both estimation methods after one to three iterations. Corrected beam profiles obtained after convergence were close to the unaberrated profiles. Transmit-beam iteration thus produced substantial improvements for all investigated aberrators.

In order to quantify the convergence, the focusing criterion C developed by Mallart and Fink,[11] and the width of the average coherence function[14] were calculated for the acoustic backscatter at each iteration.

Both the focusing criterion C and the coherence functions gave the correct impression of convergence for the transmit-beam profiles. However, there is an inherent problem of selecting an appropriate level at which to measure the width of the coherence function.

The focusing criterion C, on the other hand, is not associated with such difficulties. It is shown to determine accurately when convergence of the aberration correction procedure is achieved. The theoretical upper bound for C may be used as a criterion for the termination of the transmit-beam iteration process. It is therefore the most attractive criterion for studying transmit-beam iteration. Furthermore, as it is inexpensive to compute, C may readily be implemented in an aberration correction scheme without much additional overhead.

ACKNOWLEDGMENTS

This work was, in part, supported by the Medicine and Health program of the Research Council of Norway, and partly by the Norwegian University of Science and Technology (NTNU). Professor Harald Krogstad and Dr. Tonni Johansen at NTNU are thanked for useful discussions and comments.

[1] S. W. Flax and M. O'Donnell, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 758–767 (1988).
[2] M. O'Donnell and S. W. Flax, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 768–774 (1988).
[3] L. Hinkelman, T. D. Mast, L. A. Metlay, and R. C. Waag, "The effect of abdominal wall morphology on ultrasonic pulse distortion. Part I. Mea-

surements," J. Acoust. Soc. Am. **104**, 3635–3649 (1998).
[4] T. D. Mast, L. M. Hinkelman, M. J. Orr, and R. C. Waag, "The effect of abdominal wall morphology on ultrasonic pulse distortion. Part II. Simulations," J. Acoust. Soc. Am. **104**, 3651–3664 (1998).
[5] S.-E. Måsøy, T. F. Johansen, and B. Angelsen, "Correction of ultrasonic wave aberration with a time delay and amplitude filter," J. Acoust. Soc. Am. **113**, 2009–2020 (2003).
[6] M. Karaman, A. Atalar, H. Köymen, and M. O'Donnell, "A phase aberration correction method for ultrasound imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 275–282 (1993).
[7] S. Krishnan, K. W. Rigby, and M. O'Donnell, "Improved estimation of phase aberration profiles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 701–713 (1997).
[8] S. Krishnan, K. W. Rigby, and M. O'Donnell, "Efficient parallel adaptive aberration correction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 691–703 (1998).
[9] K. W. Rigby, C. L. Chalek, B. H. Haider, R. S. Lewandowski, L. S. S. M. O'Donnell, and D. G. Wildes, "Improved in vivo abdominal image quality using real-time estimation and correction of wavefront arrival time errors," *IEEE Ultrasonics Symposium* (IEEE, New York, 2000), Vol. 2, pp. 1645–1653.
[10] D.-L. Liu and R. C. Waag, "Correction of ultrasonic wavefront distortion using backpropagation and a reference waveform method for time-shift compensation," J. Acoust. Soc. Am. **96**, 649–660 (1994).
[11] R. Mallart and M. Fink, "Adaptive focusing in scattering media through sound-speed inhomogeneities: The van Cittert–Zernike approach and focusing criterion," J. Acoust. Soc. Am. **96**, 3721–3732 (1994).
[12] G. C. Ng, S. S. Worrel, P. D. Freiburger, and G. E. Trahey, "A comparative evaluation of several algorithms for phase aberration correction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 631–643 (1994).
[13] A. P. Berkhoff and J. Thijssen, "Correction of concentrated and distributed aberrations in medical ultrasound imaging," *1996 IEEE Ultrasonics Symp. Proc.* (IEEE, New York, 1996), Vol. 2, pp. 1405–1410.
[14] J. C. Lacefield and R. C. Waag, "Spatial coherence analysis applied to aberration correction using a two-dimensional array system," J. Acoust. Soc. Am. **112**, 2558–2566 (2002).
[15] S.-E. Måsøy, T. Varslot, and B. Angelsen, "Estimation of ultrasound wave aberration with signals from random scatterers," J. Acoust. Soc. Am. **115**, 2998–3009 (2004).
[16] T. Varslot, E. Mo, H. Krogstad, and B. Angelsen, "Eigenfunction analysis of stochastic backscatter for characterization of acoustic aberration in medical ultrasound imaging," J. Acoust. Soc. Am. **115**, 3068–3076 (2004).
[17] B. Angelsen, *Ultrasound Imaging: Waves, Signals and Signal Processing* (Emantec, Trondheim, Norway, 2000), Vol. II, http://www.ultrasoundbook.com.
[18] R. Mallart and M. Fink, "The van Cittert–Zernike theorem in pulse echo measurements," J. Acoust. Soc. Am. **90**, 2718–2727 (1991).
[19] T. Varslot, B. Angelsen, and R. C. Waag, "Spectral estimation for characterization of acoustic aberration," J. Acoust. Soc. Am. **116**, 97–108 (2004).
[20] M. B. Priestley, *Spectral Analysis and Time Series* (Academic, New York, 1988).
[21] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing* (Prentice–Hall, Upper Saddle River, New Jersey, 2000).
[22] L. F. Nock, G. E. Trahey, and S. W. Smith, "Phase aberration correction in medical ultrasound using speckle brightness as an image quality factor," J. Acoust. Soc. Am. **85**, 1819–1833 (1989).
[23] L. M. Hinkelman, D.-L. Liu, L. A. Metlay, and R. C. Waag, "Measurements of ultrasonic pulse arrival time and energy level variations produced by propagation through abdominal wall," J. Acoust. Soc. Am. **95**, 530–541 (1994).
[24] S. D. Silverstein and D. P. Ceperley, "Autofocusing in medical ultrasound: The scaled covariance matrix algorithm," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 795–804 (2003).
[25] Q. Zhu and B. Steinberg, "Deaberration of incoherent wavefront distortion: An approach toward inverse filtering," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 575–589 (1997).
[26] M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the inverse filter," J. Acoust. Soc. Am. **108**, 223–234 (2000).
[27] T. D. Mast, L. M. Hinkelman, M. J. Orr, V. W. Sparrow, and R. C. Waag, "Simulation of ultrasonic pulse propagation through the abdominal wall," J. Acoust. Soc. Am. **102**, 1177–1190 (1997).
[28] W. F. Walker and G. E. Trahey, "A fundamental limit on the performance of correlation based phase correction and flow estimation techniques," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 644–654 (1994).
[29] W. F. Walker and G. E. Trahey, "A fundamental limit on delay estimation using partially correlated speckle signals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 301–308 (1995).

Erratum: Time reversal for a single spherical scatterer [J. Acoust. Soc. Am. 109(6), 2616–2624 (2001)]

David H. Chambers and A. K. Gautesen

PACS numbers: 43.60.Tj, 43.20.Fn, 43.10.Vx

The original Table I has incorrect values for compressibility contrast α. The corrected table is given below. From the new table we see that compressibility and density contrast contribute equally to the acoustic scattering from spheres composed of hard materials (i.e., metals, glass). Scattering for soft materials (i.e., air, cork) is dominated by the compressibility contrast. Thus, multiple eigenvalues of the time reversal operator would more likely be observed for hard spheres.

TABLE I. Compressibility and density ratios relative to water for common materials.

Material	α	β	β/α
Brass	−0.99	1.3	−1.3
Copper	−0.99	1.3	−1.3
Steel	−0.99	1.2	−1.2
Glass	−0.97	0.70	−0.72
Hard rubber	−0.65	0.25	−0.38
Soft rubber	1.1	−0.052	−0.048
Cork	36	−1.5	−0.042
Oak	−0.81	−0.34	0.42
Pine	−0.60	−0.87	1.4
Air	15 000	−3.0	−0.000 19

 0001-4966/2005/117(1)/462/1/$22.50

Erratum: Phase change measurement, and speed of sound and attenuation determination, from underwater acoustic panel tests [J. Acoust. Soc. Am. 113, 1518–1524 (2003)]

Jean C. Piquette[1]
Naval Undersea Warfare Center, Division Newport, 1176 Howell Street, Newport, Rhode Island 02841

(Received 15 October 2004; accepted for publication 18 October 2004)

[DOI: 10.1121/1.1830669]

PACS numbers: 43.58.Dj, 43.58.Vb, 43.60.Qv, 43.10.Vx

Equation (3) in the subject article[1] should be

$$c_{phase}(\omega)=\frac{2b^{1/4}c_\infty\pi(b\omega^4+\omega_0^4)}{2b^{5/4}\pi\omega^4-c_\infty\pi\omega^2\omega_0\sqrt{b}[\sqrt{2}(a_2+a_0\sqrt{b})+a_1b^{1/4}]+\omega_0^3\left\{\pi[\sqrt{2}c_\infty(a_2-a_0\sqrt{b})+2b^{1/4}\omega_0]-a_1b^{1/4}c_\infty\ell n\left(b\frac{\omega^4}{\omega_0^4}\right)\right\}}. \quad (3)$$

[1] J. C. Piquette and A. E. Paolero, ''Phase change measurement, and speed of sound and attenuation determination, from underwater acoustic panel tests,'' J. Acoust. Soc. Am. **113**, pp. 1518–1524 (2003).

[1]Electronic mail: piquettejc@npt.nuwc.navy.mil.